



1

Chapter 9

2

Sloboda and Parker's recall paradigm for melodic memory: a new, computational perspective

3

4

5

Daniel Müllensiefen and Geraint A. Wiggins

6

Abstract

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

Sloboda and Parker (1985) proposed a new experimental paradigm for research on melodic memory in which participants are asked to listen to novel melodies and to sing back the parts they can recall from memory. In contrast to the many varieties of melodic recognition paradigms frequently used in memory research this sung recall paradigm can answer questions about how mental representations of a melody build up in memory over time, about the nature of memory errors, and about the interplay between different musical dimensions in memory. Although the paradigm has clear advantages with regard to ecological validity, Sloboda and Parker also note a number of difficulties inherent to the paradigm that mostly result from necessity to analyse 'dirty musical data' as sung by mostly untrained participants. This contribution reviews previous research done using the sung recall paradigm and proposes a computational approach for the analysis of dirty melodic data. This approach is applied to data from a new study using Sloboda and Parker's paradigm. This chapter discusses how this new approach not only enables researchers to handle large amounts of data but also make use of concepts from computational music analysis and music information retrieval that introduce a new level of analytic precision and conceptual clarity and thus provide a new interface which connects Sloboda's paradigm to rigorous quantitative data analysis.





1 The seminal study by John Sloboda and David Parker (1985) provided new insights
2 into the construction of mental representations of melodic structures, and into mech-
3 anisms whereby learning of previously unknown melodies is accomplished over
4 repeated attempts. In doing so, it introduced a new experimental paradigm for inves-
5 tigating melodic memory via sung recalls. The work is widely cited in subsequent lit-
6 erature on musical and melodic memory. However, despite the frequent citations and
7 wide dissemination of the results of the study, the use of the recall paradigm has been
8 limited to just a handful of subsequent studies. It has not achieved the level of usage
9 and variation in the music cognition community achieved, for example, by paradigms
10 based on recognition tasks.

11 In this chapter, we review the Sloboda and Parker paradigm from a methodological
12 perspective, and hypothesize that one of the reasons why its uptake in the literature is
13 at odds with the originality of result that it can produce are the relatively 'dirty' data it
14 produces. A lack of standardized methods and the amount of manual work required
15 to deal with these types of data seem to be major impediments to uptake of the para-
16 digm. To ameliorate these drawbacks, we propose use of the computer as a tool for
17 musical data analysis, and also as a means of model building and hypothesis genera-
18 tion. We demonstrate how it can help to make analysis procedures explicit and thus
19 contribute to standardization. We reanalyse the original data from Sloboda and
20 Parker's (1985) study, to demonstrate where computing technology can be applied
21 and what additional value it can bring to the analysis of data from this very rich para-
22 digm, whose full potential, we argue, has yet to be realized.

23 **Paradigms for general memory research**

24 In classic textbooks on experimental psychology (Anderson & Borkowski, 1978,
25 394–396; Kantowitz, Roediger, & Elmes, 1994, pp. 284–285; Kluwe, 1990), we find a
26 well-developed canon of experimental paradigms that have been used to investigate
27 verbal memory. The earliest go back to the very beginnings of psychology as a disci-
28 pline and all of them have evolved and improved with time. For the purposes of the
29 argument here a short description of each paradigm, in summary, will suffice.

30 Serial recall is the oldest experimental paradigm in the psychology of memory. It
31 was first proposed by Hermann Ebbinghaus (1885) and has since been widely deployed
32 in experimental studies, and considerably refined. Its basic task is to remember a list
33 of items in order and recall them subsequently in the same order. Ebbinghaus and
34 other well-known subsequent studies (e.g. Young, 1962) used meaningless syllables as
35 experimental stimuli. Results generated by this paradigm inform us about the amount
36 of repetition necessary to recall a list of items perfectly, the savings in effort and time
37 when lists are relearned, and the decay of items in memory over time.

38 Paired-associate learning is another very early memory paradigm (Calkins, 1894)
39 and can be directly linked to the stimulus–response concept of classical conditioning
40 (Pavlov, 1927). Participants are given a list of pairs of items to memorize. They are
41 then presented with a list containing one item of each pair, and asked to recall its
42 match from the learned list before the match is shown. This process is repeated until
43 the full list of pairs is memorized perfectly. Meaningless syllables and single words are





1 used as items in verbal learning experiments using this paradigm. Central questions
2 investigated concern item characteristics (e.g. the similarity between items, their
3 imageability), the formation of association strengths, and the effects of cognitive
4 mediators on recall performance.

5 Free recall, in contrast with serial recall, requires participants to learn a list of items,
6 but leaves them free to recall the learned items in any order. If the study list is pre-
7 sented several times, then the position of the items is permuted, to avoid order asso-
8 ciations. The most important observable in this paradigm is the effect of the position
9 of an item in the study list on the probability of its recall. Regardless of the length of
10 the study list, the initial items and the final items in the list are usually much better
11 recalled than the items presented in middle positions (e.g. Murdock, 1962). These
12 effects are known as *primacy* and *recency*; they can be reproduced very reliably with the
13 *recency* effect usually being stronger than the *primacy* effect. One interpretation of
14 these effects (Atkinson & Shiffrin, 1968) was that they suggest the existence of distinct
15 memory stores where the first items of a list would have entered a long-term store and
16 the last items would have still been present in a short-term store.

17 Recognition requires participants are to study a list of items, but, in contrast to the
18 recall paradigms, described above, they are subsequently presented with a list contain-
19 ing items from the study list, and also new, previously unseen items. Participants are
20 then asked to indicate which items were encountered previously, or one seen item can
21 be presented along with several unseen items in a multiple-choice selection task.
22 Recognition memory performance is generally much higher than recall memory per-
23 formance (e.g. Shepard, 1967), though there are exceptions (Tulving, 1968). Not only
24 can participants generally recognize high percentages (>80%) of long item lists (e.g.
25 500 items and more) but they are also able to maintain good recognition memories
26 over long time intervals. This long-term performance is in clear contrast to the nega-
27 tive exponential forgetting curve that Ebbinghaus (1885) and others found with serial
28 and other recall paradigms; it holds true for many different types of items (e.g. images,
29 words).

30 **Choosing a paradigm**

31 There are good reasons for the existence and deployment of a wide range of experi-
32 mental paradigms in memory research. First of all, different experimental procedures
33 allow us to study different memory effects and to provide evidence for different types
34 of hypothesis. For example, effects of list length and list position on correct recall from
35 memory are typically studied using serial recall (Young, 1962), while ability to dis-
36 criminate between stimuli has been investigated using the paired-associate learning
37 paradigm (Underwood, Rundquist, & Schulz, 1959), and the effects of different
38 rehearsal strategies can be revealed using the free recall procedure, for example with
39 overt rehearsal (Rundus, 1974). Similarly, results from the various paradigms and
40 their potential to falsify hypotheses of different nature have led to the proposal of cor-
41 respondingly different memory models. For example, a specific class of memory trace
42 model is primarily based on serial recall data (e.g. Nairne, 1990), while models using
43 semantic hierarchies and clustering are often developed in connection with data from





1 free recall studies (Tulving, 1968). Finally, data stemming from different experimental
 2 paradigms is often used to provide complementary evidence about the same model or
 3 hypothesis. This can help refute the claim that effects discovered with a certain para-
 4 digm are merely artefacts of that paradigm. Good examples are primacy and recency
 5 effects which reliably appear in both serial recall (Mueller, 1970) and free recall para-
 6 digms (Murdock, 1962). Complex, modern memory models that claim general appli-
 7 cability are usually based on data generated with diverse and very different paradigms.
 8 One such example is Baddeley's concept of working memory (1986, 2007) which is
 9 able to account for a large number of findings from recall studies using a variety of
 10 different tasks.

11 Paradigms in music memory research

12 Overview

13 Turning to research on memory for music, and specifically memory for *melodies*, we
 14 find a more limited range of basic experimental paradigms. If we naïvely suppose that
 15 the individual notes or pitch intervals of a melody are units comparable to words,¹
 16 syllables, or digits in verbal memory research, then a free recall paradigm in a strict
 17 sense would be, by definition, melodically meaningless (though it might convey har-
 18 monic information): if notes are not recalled in a serial order, but freely, then it is hard
 19 to recognize the stimulus item and therefore to judge whether a participant has actu-
 20 ally recalled it, and, if so, to what degree of accuracy. Some melodic recall studies
 21 include explicit instructions for the participants that the parts (sections, phrases,
 22 motives) of the melody items may be recalled in any order. Especially with longer
 23 melody items, this approach can make the experimental task much easier for the par-
 24 ticipants. Here, the paradigm comes closer to the verbal free recall paradigm, but it
 25 must still be considered a compromise variant of the basic melody recall procedure.
 26 We therefore conclude that there is essentially only one recall paradigm in music
 27 research, which is broadly comparable with verbal serial recall, *regardless of how the*
 28 *recall is actually performed in the experiment*. As we will see below, depending on the
 29 target group of participants, recall response modes can range from singing through
 30 playing on an instrument to using music notation or verbal labels for individual notes.
 31 In all instances though, the serial order of the recalled notes is of primary
 32 importance.

33 This said, it is possible, and not unreasonable, to define very abstract representa-
 34 tions of melody which do not rely on strict note sequence—for example, by pitch class
 35 or interval counting. However, experience suggests that these are not good representa-
 36 tions for musical memory: order does seem to be paramount, as one might expect.
 37 These representations can nevertheless be useful in applications such as music infor-
 38 mation retrieval (for example, harmonic interval statistics are used for this purpose by
 39 Pickens *et al.*, 2003).

¹ In fact, music is extremely context-sensitive, and, statistically, at least, there seems to be some comparison between sequences of three- or four-note chords and words (Mauch, Müllensiefen, Dixon, & Wiggins, 2008), though how deep this comparison runs is far from clear.



1 Nor is paired-associate learning often seen in music research. This may be again due
2 to the sequential nature of music, and melodies in particular. It is hard to imagine a
3 musically meaningful task where musical or melodic elements can be presented in
4 pairs and to be encoded together. At least, one would have to consider passages of
5 music that can be paired according to specific attributes (e.g. timbre, pitch range,
6 tempo, specific rhythms). This paradigm would clearly make no musical sense at all if
7 one took notes or intervals as the atomic elements to be remembered in pairs.

8 Recognition paradigms in music memory research

9 Recognition is the only other memory paradigm available for research on memory for
10 melodies. Recognition has been used extensively over the past 40 years and has diver-
11 sified into a few major subparadigms, the most important of which we now briefly
12 describe.

13 **Pitch comparison** The **Deutsch paradigm** was first used by its eponymous designer
14 in the early 1970s, in a tradition of many studies targeting short-term memory, its
15 capacity, and the conditions for interference effects for verbal memory. Deutsch dedi-
16 cated a series of studies to the limits of the auditory store holding pitch information
17 and to the question of how memory representation for pitch could be eliminated by
18 interference effects (Deutsch, 1970, 1972, 1974, 1975a, 1975b; Deutsch & Feroe, 1975).
19 The paradigm used in these studies was later used by other researchers in follow-up
20 studies. In its basic form, the Deutsch paradigm can be described thus. The participant
21 is presented with a single target pitch which they are asked to keep in memory. Then,
22 they listen to some intervening stimuli in an *interpolation* phase (called *retention* else-
23 where): typically, the stimuli in the interpolation phase are varied as factors of the
24 independent variable. These stimuli can be verbal auditory material or a varied number
25 of pitches which can be more or less harmonically or melodically related to the target
26 stimulus. Finally, a comparison pitch is played and the participant judges whether it is
27 the same as the initial pitch or not.

28 The results generated by experiments using this paradigm seem to indicate that
29 there may be a store for pitch information, separate from any verbal auditory store,
30 and which is influenced only to a minor degree by other characteristics of a tone, such
31 as loudness, timbre, or direction (i.e. the ear it is presented to). However, this pitch
32 memory system, which has been hypothesized to be a separate subsystem in Baddeley's
33 auditory working memory (Pechmann & Mohr, 1992), is fairly strongly influenced by
34 pitches presented in the interpolation phase of the Deutsch paradigm. The interfer-
35 ence with the memory representation of the target pitch is particularly strong if the
36 interpolation pitches are close to the target in continuous pitch space, or if they are
37 identical to the comparison pitch.

38 **AB comparison** is widely used in melodic memory research (e.g. DeWitt & Crowder,
39 1986; Dowling, 1972, 1978; Dowling & Fujitani, 1971; Idson & Massaro, 1978). It is
40 similar to the Deutsch paradigm in that the participant is asked to compare two stimuli
41 and indicate whether they are identical or not, and, sometimes, rate their confidence in
42 their judgement. But with AB comparison, the target and comparison stimuli are
43 longer melodic sequences, parts of real melodies, or excerpts from a polyphonic piece.

1 The retention phase between the presentation of the target stimulus, A, and the com-
2 parison stimulus, B, is less relevant than in the Deutsch paradigm: material presented
3 in this phase often serves only as a distractor.

4 Variation in applications of AB comparison is often in the melodic attributes that
5 are held the same or made different between A and B. The rationale is that, if two
6 melodic sequences differ in a specific melodic attribute (e.g. contour) but participants
7 nonetheless indicate that A and B are identical, then this particular attribute seems not
8 to have been encoded in memory (Idson & Massaro, 1978, p. 554). Therefore, the
9 paradigm is particularly well suited to uncovering which melodic attributes are encoded
10 in memory, and how different melodic structures and contexts favour the encoding of
11 particular attributes or combinations of attributes. The *scale and contour* theory of
12 melodic memory (Dowling, 1978) is mostly based on data from AB comparison, and
13 it has delivered much insight into how melodies of different lengths are encoded
14 (Long, 1977), under which conditions interval versus contour information is encoded
15 (Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Edworthy, 1983, 1985), and how
16 the encoding process of melodic phrases proceeds during the course of a real listening
17 experience (Dowling, Tillmann, & Ayers, 2002).

18 AB comparison is related to testing procedures from psychophysics, where two
19 stimuli are to be compared with regard, for example, with difference in loudness or in
20 pitch. Consequently, bias-free scoring procedures originally applied in psychophysical
21 tests, such as d' (Green & Swets, 1966), are often employed in memory experiments of
22 AB comparison to obtain values of the independent variable.

23 **List-wise recognition** is closer to the recognition procedure usually found in verbal
24 memory experiments. In the study phase, short melodies are presented one after
25 another forming a list. After a short retention interval, there is a test phase, in which
26 another list of melodies is played to the participants. The test list contains melodies
27 used in the study list as well as new melodies. Participants indicate whether each item
28 on the test list was presented in the study phase, and rate their confidence in their
29 judgement. In addition, or as an alternative, participants may be asked rate the pleas-
30 antness of the melodies. Pleasantness ratings haven been used as a measure of implicit
31 memory (Halpern & Müllensiefen, 2008) on the basis that the previous exposure
32 to the same melody increases its aesthetic appreciation (the mere exposure effect:
33 Zajonc, 1968).

34 There are two major differences between list-wise recognition and AB comparison.
35 First, in list-wise recognition, the significance of the retention phase is relatively very
36 limited, and its length and the content are generally not part of the experimental
37 design. Second, list-wise recognition admits greater distance in time and more inter-
38 vening musical material between corresponding items on the study and test lists. If the
39 lists are substantially long, then episodic memory representations of particular melody
40 items are likely to be lost, due to interference or to decay of the original memory trace.
41 This makes list-wise recognition well suited to implicit memory testing.

42 **Dynamic recognition paradigm for familiar melodies** was popularized by Matthew
43 Schulkind and collaborators in a series of publications (Schulkind, 2000, 2004;
44 Schulkind, Posner, & Rubin, 2003) on the recognition of well-known tunes. In the
45 first trial, participants listen to an incipit of a given length from a melody, and then are





1 asked to indicate the title of the tune if they can recognize it. If a participant is unable
 2 to name the tune confidently, the incipit is played again, in repeated trials, with one
 3 more note from the melody included each time; there is a prompt for a recognition
 4 judgement after each trial. For each incipit length, the recognition rate across partici-
 5 pants is recorded and is subsequently compared with melodic and rhythmic events
 6 happening or being completed at that point in the melody.

7 **Recognition versus recall** The greater diversification within the recognition para-
 8 digm corresponds with the far greater number of studies that have been published
 9 using recognition, rather than recall, as the participants' response. We will discuss
 10 some reasons for the dominance of the various forms of the recognition paradigm in
 11 research on melodic memory below. For now, we conclude by remarking that, in con-
 12 trast to the few studies using a melodic recall paradigm, there is a plethora of melodic
 13 recognition studies in the literature and that recognition is significantly more promi-
 14 nent, more diversified and more developed in melodic memory research. Given the
 15 potential benefits of employing different memory paradigms in verbal research (find-
 16 ing different effects, coming up with different types of models, and corroborating the
 17 same effects/models from a different perspective, etc.), we suggest that there is a clear
 18 motivation to develop the recall paradigm for melodic memory. Furthermore, we will
 19 advocate the use of computers in the analysis and modelling of the melodic recalls
 20 generated from this paradigm.

21 **Studies using recall paradigms for melodic memory**

22 Studies using recall paradigms are much less common than those employing one of
 23 the variants of recognition surveyed above. The differences between recall studies lie
 24 mostly in whether novel or familiar melodies are used as experimental stimuli, in
 25 whether participants are musically trained or untrained (or perhaps the study com-
 26 pares performance between these two groups), and in the participants' response mode.
 27 It is worth briefly reviewing some recall studies here to provide a feel for the spectrum
 28 of existing methods, and to position the specific experimental paradigm of Sloboda
 29 and Parker (1985) therein.

30 Musical dictation was used by Deutsch (1980) to provide evidence for her hierarchical-
 31 generative system of mental representations (Deutsch & Feroe, 1981). Musically well-
 32 trained listeners listened to sequences of 12 sine tones, representing different degrees
 33 of structural difficulty. The tone sequences were produced from the Deutsch–Feroe
 34 model, applying different generative rules, to generate easier or more difficult sequences.
 35 Some sequences included leading tones, while others were mainly built around triads,
 36 and others were segmented on the basis of melodic versus temporal gaps. The partici-
 37 pants' task was to write down, in staff notation, as much as they could remember from
 38 the stimuli. The notated responses were evaluated by calculating the relative number
 39 of correct pitches at the correct serial position. Deutsch found more correct recall for
 40 the easier melodies, that could be encoded more efficiently as predicted by the
 41 Deutsch–Feroe model. She therefore concluded that memorizing a melodic sequence
 42 consists of encoding the hierarchical structure of the sequence and its alphabet; a sin-
 43 gle chunk from the sequence (e.g. the incipit) would also have to be encoded, to allow





1 for a subsequent reconstruction of the whole sequence from memory. The Deutsch–
2 Feroe model is specified in a generative form, and producing sequences from it that
3 can be used as experimental stimuli is straightforward. But it is not an analytic device
4 capable of encoding an existing melody; to perform that more complex task, a fully
5 functional parser would be needed. This memory model is therefore of limited use for
6 experiments that use real, existing melodies as stimuli. The music dictation method
7 employed in this experiment requires a good level of musical training among the par-
8 ticipants and is therefore not suitable for exploring memory representation of
9 untrained subjects.

10 To overcome this last limitation, Davies and Yelland (1977) asked their partici-
11 pants to draw from memory a representation of the contours of short, song-like,
12 newly composed sequences, to which they had listened. Contour was represented as
13 rising, falling, or horizontal lines between successive notes. The resulting draw-
14 ings were scored according to the number of lines with correct inclination at the cor-
15 rect serial position. Davies and Yelland were not concerned with effects of melodic
16 structure, but varied the number of repeated listenings and the type of training
17 procedure used to familiarize participants with the contour drawing method as inde-
18 pendent variables. As expected, participants drew increasingly more accurate con-
19 tours over repeated listenings to the same melody. This corroborates the assumption
20 that mental representations of melodies become more accurate with repeated expo-
21 sure. With regard to the training procedure comparison, the best results came
22 from participants who had practised contour drawing with well-known melodies
23 from long-term memory in silence as opposed to another group that practised
24 drawing novel melodies that were presented aurally and whose members received
25 feedback on their drawings. Davies and Yelland interpret this group difference
26 as pointing to the importance of comparing novel stimuli with settled internal repre-
27 sentations. Also, it seems that contour information is not explicitly abstracted during
28 listening, but that melodies are represented as tonal analogies from which abstrac-
29 tions such as contour can be derived in subsequent serial scans through the
30 representation.

31 While Davies and Yelland's experimental procedure is suitable for participants with
32 little or no musical training, it is limited to reflecting the accuracy of melodic contour
33 encoding and it ignores other parameters such as interval size, harmonic implication
34 or rhythmic duration. It also requires a transfer of 'up-and-down' information from
35 the auditory domain into the visual domain, and so it is not possible to separately
36 distinguish if participants with low scores had an inaccurate representation of the
37 melody as an auditory object, or if they were less good at transferring information
38 from the auditory domain to the visual response domain. The different training pro-
39 cedures could potentially have trained these two different processes to different
40 degrees.

41 Contour label recall was used by Williamson, Baddeley, and Hitch (2006) (and
42 similarly Keller, Cowen, & Saults, 1995) to test recall memory for short pitch sequenc-
43 es. It asks for recall of the contour of four-note isochronic pitch sequences from work-
44 ing memory. Each note of a sequence could be either 'high', 'medium', or 'low', in
45 relation to the other notes. In each trial, after hearing the pitch sequence, participants



1 were asked to fill out a matrix in a paper booklet indicating the contour category of
2 each of the four notes. Since participants' responses are expressed in a written nota-
3 tion requiring no musical knowledge, they need no musical training or any singing
4 skills. The resulting data are very clean compared with sung recalls from the para-
5 digms, described below, which require skilled transcription and expert analysis. In
6 principle, the paradigm is not limited to the evaluation of memory representation of
7 melodic contour (i.e. higher-than and lower-than judgements with respect to pitch
8 height): judging duration or rhythmic categories as well as intervallic distance should,
9 in theory, also be possible. However, it is limited in that, with musically untrained
10 listeners, memory load issues can mean that it is possible test only a single musical
11 dimension (contour or rhythm or intervals) at a time, so interactions between dimen-
12 sions cannot be studied.

13 Also, experimental stimuli must be relatively simple and the number of different
14 stimulus class labels is limited; Miller's (1956) rule of thumb of 7 ± 2 different magni-
15 tude values or categories that are simultaneously manageable on the same perceptual
16 dimension may possibly hold as an upper limit (for a more recent affirmation of
17 Miller's original results, see Shiffrin & Nosofsky, 1994). However, many popular
18 melodies make use of more than seven different pitches, and processing and memory
19 constraints are aggravated by the requirement to report the category label of each note
20 of a series in order. Determining the limits of the paradigm before floor effects are
21 reached in terms of the number of different category classes, as well as with respect to
22 number of musical dimensions that can be combined and in terms of the length of the
23 note sequence would give valuable insight into to which degree this recall paradigm
24 can be used with non-artificial melodic stimuli.

25 Vocal recall of known songs was used by Halpern (1989) and Levitin (1994) to
26 determine the accuracy for the absolute pitch of melodies from well-known folk or
27 pop songs. Here, the accuracy of the memory representation of the relative structure
28 of the melodies was of less importance because only the first note (Halpern) or the first
29 three notes (Levitin) sung by the participants were used for analysis; participants that
30 failed to produce a stable pitch were excluded from the sample. The analysis of the
31 sung recalls must have been quite laborious in both cases. In Halpern's study, the
32 starting pitches of the folk songs were transcribed by an expert musician with refer-
33 ence to a keyboard. Levitin used a computer program to estimate fundamental fre-
34 quency using the Fast Fourier Transform. He determined the pitch of each of the three
35 starting notes of the participants' recalls by selecting just the steady state portion of the
36 sung notes with an audio editor and rounding the fundamental frequency to the near-
37 est semitone. This paradigm is, in principle, quite similar to Sloboda and Parker's
38 paradigm in that sung responses from memory are recorded and transcribed, and, in
39 consequence, researchers using vocal recall face the same problems as Sloboda and
40 Parker, which we outline below. A lot of expert work is required to transform the sung
41 audio recordings into score notation and finally into a numerical representation of
42 pitch. However, because Halpern (1989) and Levitin (1994) are only concerned with
43 long-term memory for absolute pitch, and melodic structure, rhythmic, metric, phras-
44 ing, and interval information in the sung recalls need not be transcribed, the complex-
45 ity of the task is greatly reduced.



1 The Sloboda and Parker recall paradigm

2 The core component of the Sloboda and Parker's (1985) procedure is the request to
3 participants to sing the stimulus melody item back from memory. The experimental
4 procedure, as used in their original study, can be split into four stages:

5 1 A melody is played to the participant up to six times. After each repetition, the par-
6 ticipant is asked to sing into a microphone and recording device those parts of the
7 melody that they can remember.

8 2 The audio recordings are transcribed into music notation by a human expert.

9 3 From the notation, each recall is analysed with regard to different aspects such as
10 metre, breathing breaks, melodic contour, and phrase structure.

11 4 The dependent variables are compared over the values of the independent variables
12 of interest: trial number, participants' musical background, melody type, and pres-
13 entation mode (e.g. with or without lyrics: Ginsborg & Sloboda, 2007).

14 To obtain accuracy scores for different types of errors, in step 3, and ultimately to
15 enable quantitative analyses, Oura and Hatano (1988) developed a more formal scor-
16 ing method which was subsequently used by Zielinska and Miklaszewski (1992),
17 Drake, Dowling, and Palmer (1991) and Ogawa *et al.* (1995) with slight variations. In
18 this scoring procedure, each phrase in the recall where a tonality can be identified is
19 assigned to the best matching phrase in the original melody. Each phrase is then com-
20 pared, in half-bar windows, to its counterpart in the original. If the half-bar window is
21 identical to the original, it is labelled as *correct*. If it differs in either rhythm or contour,
22 it is labelled as *modified* and if it differs in both parameters from the original it is
23 declared *non-identified*. The windows are summed together by category (and error
24 type) and their proportions with respect to the overall number of half-bars are taken
25 as dependent variables for that particular trial.

26 While the method seems superficially straightforward, and the results that can be
27 produced from it are of great interest, much cited, and truly complementary to what
28 can be obtained from a recognition paradigm, it has been used only a few times in the
29 25 years since it was published. This disappointing uptake is due to some inherent dif-
30 ficulties with the method, most of which were discussed in the original article (Sloboda
31 & Parker, 1985); we summarize them as follows.

32 1 The raw data recorded in these studies are extremely 'dirty', requiring expert
33 interpretation. Participants are required to sing, and their singing may be inaccur-
34 ate; in some places, it is necessary to infer which note(s) they *meant* to sing.
35 The participants' singing is recorded, and it is possible that the recording may
36 be imperfect; it is impossible to prevent this without breaking the paradigm, for
37 obvious reasons.

38 2 There is 'no theory of melodic identity' (Sloboda & Parker, 1985, p. 159). Therefore,
39 subjective judgements of the expert transcribers are crucial in doing the analysis.
40 This subjectivity can clearly affect the results, and cannot be controlled.

41 3 It follows that there can be no standard, verified scoring method for these
42 complex response data; that is, there is no theory of melodic *similarity* to give a



**Table 9.1** Numbers of recalled fragments reported in studies applying Sloboda and Parker's recall paradigm

| Study | Recalled fragments |
|--|--------------------|
| Sloboda and Parker (1985) | 48 |
| Oura and Hatano (1988) | 320 |
| Zielinska and Miklaszewski (1992) | 310 |
| Ogawa <i>et al.</i> (1995) | 80 |
| Ginsborg and Sloboda (2007) | 60 |
| Müllensiefen and Wiggins (manuscript in preparation) | 1900 |

1 measure of how close the attempts are to the original. Oura and Hatano (1988)
2 aimed to address this with their procedure, which was described above.

3 4 A simple but very limiting issue is the amount of time it takes to transcribe the data
4 and perform these analyses. No really large studies using the method have yet been
5 published; those studies which have been carried out are listed, with their size, in
6 Table 9.1. This last issue, superficially, is the kind of problem that one would like to
7 solve by using a computer, on the grounds that computers can analyse data very
8 quickly and very accurately. However, because music is fundamentally a perceptual
9 and cognitive construct, any program used for these tasks needs to encode *human*
10 *expert* levels of musical understanding. In the next section, we explore the extent to
11 which this is currently possible, show what can be done, and examine prospects for
12 the future.

13 **Computational methods for researching melodic memory**

14 **Towards automatic support of Sloboda and** 15 **Parker's paradigm**

16 Our central idea in this paper is that the experimental paradigm proposed by Sloboda
17 and Parker (1985) may be greatly facilitated by appropriate application of computa-
18 tional technology. Broadly, it should be possible to use computational methods to
19 process, analyse, and compare recall data, thus allowing the method to be applied
20 more widely.

21 We summarized the drawbacks with the method in the previous section, above; a
22 computational approach can, at least in principle, help with most of them. Digital
23 recording and post-hoc computational audio analysis can greatly help with problem
24 number 1; this technology has been readily and cheaply available for more than a dec-
25 ade, though it is becoming established in music and music psychology research only
26 recently (e.g. Cook, 2008). In particular, as noted above, the speed and repeatability of
27 computer processing could help in principle with problem number 4; however, this is
28 dependent on solutions to the more musical aspects in problems 2 and 3, which are
29 very difficult. Sloboda and Parker's statement that there is 'no theory of melodic iden-
30 tity' (1985, p. 159) holds just as true 25 years on. However, there is now a substantial





1 body of research leading towards computational theories of melodic similarity which
2 are applicable in an analytical context. Progress towards a computational theory brings
3 concomitant benefits beyond the immediate application: in particular, to build a pro-
4 gram that embodies a theory, one must specify that theory to an extreme—indeed
5 *absolute*—level of detail. Once such a specification is given and a program written, it
6 can be tested to destruction against all the data available from studies involving
7 humans, without problems of fatigue, priming, or even ethics. In this way, the compu-
8 tational approach not only facilitates the empirical work, but can make its analysis
9 more rigorous and objective; new knowledge can be created in terms of novel hypoth-
10 eses generated by the models.

11 Extant computational methods developed for various musical purposes can be
12 applied in this context. Since musical memory, and hence musical similarity, de facto
13 underpins the vast majority of musical behaviour (Wiggins, 2007), there is a rich seam
14 of more general work to be mined and repurposed to help solve the current problem.
15 There is not space here to review or even to list all the work that could contribute in
16 this way, but some examples follow. Temperley (2001) and Huron (1995, 2006) have
17 developed cognitively informed methods intended for computational music analysis;
18 in particular, attempts to understand structural reference in this context help in our
19 understanding of memory, for it is this which enables such reference to be understood.
20 These essentially musicological approaches can be enhanced by models from mathe-
21 matical music theory (e.g. RUBATO: Milmeister, Mazzola, & Thalmann, 2009), which
22 generally aim to find mathematical systems underpinning perception and cognition,
23 with a view to explaining why either the music or the perception is the way it is.

24 The study of music cognition in general (e.g. Eerola, Järvinen, Louhivuori, &
25 Toiviainen, 2002; Krumhansl, 1990; Thomassen, 1982), from an empirical stand-
26 point, allows different kinds of insight into what may be expected in perception of music,
27 and can inform the construction of heuristics (see below) for helping with analysis of
28 musical data, as well as supplying models of musical competence, such as memory.

29 In a practically motivated context, a significant amount of work in the field of music
30 information retrieval is concerned with understanding of perceived similarity of vari-
31 ous kinds, such as harmonic similarity, usually at the level of whole pieces, (e.g. Pickens
32 *et al.*, 2003) and melodic similarity (e.g. Crawford, Iliopoulos, & Raman, 1998;
33 Müllensiefen & Frieler, 2004a). Finally, since there seem to be shared effects and
34 behaviours between music and language, it has proven useful to consider techniques
35 from computational linguistics in the musical context (e.g. Downie, 2003).

36 Taken together, models and methods developed in these areas can already analyse
37 many aspects of melodies that are of core interest in many psychological studies. These
38 include: melodic identity and similarity (Müllensiefen & Pendzich, 2009); metrical
39 structure, metre induction (Müllensiefen and Frieler, 2004b); (Eck, 2002; Volk, 2008);
40 phrase structure (Pearce, Müllensiefen, & Wiggins, 2010; Temperley, 2001); rhythmic
41 structure (Weyde, 2004); harmonic structure: tonality induction (Krumhansl, 1990;
42 Longuet-Higgins & Steedman, 1971); accent strength (Müllensiefen, Pfeleiderer,
43 and Frieler, 2009), complexity (North & Hargreaves, 1995), expectedness (Pearce &
44 Wiggins, 2006), high-level structure identification (Abdallah, Sandler, Rhodes, &
45 Casey, 2006).





1 While computational methods are nowhere near the position of simulating a human
2 listener as a whole, we suggest that the time is ripe to begin applying these methods to
3 focused problems such as those provided by the Sloboda and Parker paradigm.
4 Through feedback from such application, the methods will themselves be improved.

5 **Computational modelling of perceived melodic similarity**

6 A key issue in executing the Sloboda and Parker paradigm is the comparison of recalled
7 melodies with the original stimulus. This is a hard task for human experts, not simply
8 because it is a difficult task in any case, but in particular because the judgements
9 required can sometimes be very subjective, based on the prior musical bias of the
10 expert. One way one might increase objectivity would be to do the analysis automati-
11 cally: this has the advantage of removing subjectivity based on purely human judge-
12 ment. However, such judgements cannot, in the current state of the art, be made
13 entirely reliably by computer. In this section, we briefly survey the state of that art,
14 with a view to identifying what can be done and what needs further work. We also
15 exemplify this concept with a short description of the *edit distance* (or *Levenshtein*
16 *distance*), an algorithm that is very simple but which has proven to be effective for
17 similarity computations in automatic text processing, computational biology
18 (Gusfield, 1997), and also audio and music computing (e.g. Crawford *et al.*, 1998;
19 Unal, Chew, Georgiou, & Narayanan, 2008).

20 We described the scoring procedure developed by Ogawa *et al.* (1995) in an earlier
21 section. In this method, melodic similarity is measured by (manually) determining the
22 identity of half-bar segments from the original melody and the sung recalls. This
23 approach, while not computerized, is archetypal of computerized approaches. Because
24 musical similarity is a multidimensional thing, it is necessary to be clear about which
25 dimensions one is interested in (and, of course, this may be a parameter of one's
26 study). It follows that, if we are interested in more than one dimension, but we need a
27 linear scale for comparison, we also need a mathematical means of mapping vector
28 distances in a multidimensional space on to scalar distances; fortunately, the mathe-
29 matics required is undaunting to psychologists. Both the dimensions and mapping
30 methods must correspond properly with perceived similarity; where there is ambigu-
31 ity in human response, this should ideally be detectable in the measurement system
32 (for example, two different multidimensional distances might map on to the same
33 scalar distance). To produce a similarity measure which meets this specification in full
34 is a very tall order, and much more research is required to do so.

35 However, we are in a position to begin. Assuming that we have transcriptions of all
36 melodies in a symbolic format to start from, there are two essential steps in measuring
37 melodic similarity: first, the melodies are transformed into one (or several)
38 representation(s) which encode(s) all the dimensions of the music in which one is
39 interested; second, the abstract representations are compared and a numerical value
40 (usually normalized between 0 and 1) is derived to indicate the distance between
41 them.

42 In the first step, the raw melodies (which might be represented, for example, only by
43 a sequence of tuples of pitch and onset values) are transformed into one or more





1 abstract representations that are cognitively more meaningful and of interest with
 2 regard to the similarity comparison. Melodic representations that can be usefully
 3 employed for similarity measurement include melodic intervals, melody contour, a
 4 sequence of durational or rhythmic values, and a sequence of tonality values that are
 5 implied by the melodic pitches. The important point about this step is that the raw
 6 melodic information is *abstracted* and melodies are represented as sequences of numer-
 7 ical or digit symbols. The transformation is in principal independent of the algorithm
 8 that is used to compare symbol sequences subsequently. The choice of transformation
 9 depends rather on whether the researcher deems it important with respect to the simi-
 10 larity measurement or whether it contains an optimal amount of information in an
 11 information-theoretical sense (Pearce & Wiggins, 2004). Of course, various different
 12 representations of the same melody (e.g. one containing pitch intervals and one con-
 13 taining rhythmic values) can be combined for subsequent similarity measurement.

14 In the second step, from the abstract representations of two melodies a similarity
 15 value from the interval $\in [0, \dots, 1]$ is computed where 1 indicates maximal similarity
 16 or identity and 0 indicates no similarity between the two melodies. In recent years,
 17 several different measures, and algorithms to implement them, have been proposed
 18 and tested in various applications needing to determine the similarity between melo-
 19 dies. These include geometric measures (Aloupis *et al.*, 2003; O Maidin, 1998), string
 20 matching techniques such as edit distance (Crawford *et al.*, 1998; Mongeau & Sankoff,
 21 1990), n-gram measures (Downie, 2003; Uitdenbogerd, 2002), and hidden Markov
 22 models (Meek & Birmingham, 2002) from speech recognition, as well as the Earth
 23 Movers Distance algorithm (Typke, Wiering, & Veltkamp, 2007) from computer
 24 vision, and hybrid algorithms that combine the output of several different similarity
 25 measurement procedures (Müllensiefen & Frieler, 2004a). Here, by way of example,
 26 we briefly explain the edit distance measure, which, in recent years has proven to be a
 27 surprisingly effective comparison or benchmark algorithm that appears to be similar
 28 to informal and rather intuition-guided music-analytic applications of melodic simi-
 29 larity concepts (for a comparative discussion of quantitative and qualitative approach-
 30 es to melodic similarity, see Müllensiefen & Pendzich, 2009).

31 The main idea behind the edit distance, or Levenshtein distance, is to treat the mini-
 32 mum number of operations ('edits') needed to transform one string into another as a
 33 measure of the distance between them. The permitted operations are insertion, dele-
 34 tion, and substitution. Since the operation of the algorithm is independent of the
 35 denotation of the symbols in the strings compared, we use letter strings to demon-
 36 strate the edit distance in our example. Consider the two letter strings SCHOENBERG
 37 and SCHUBERT. These strings can be aligned optimally by applying the series of
 38 operations applied to each letter that is recorded in the *edit transcript* shown. The
 39 symbol D denotes the deletion of a letter, I stands for insertion, S for a substitution,
 40 and M for a match of two letters, i.e. no operation is carried out.

$$\begin{array}{r}
 \text{String 1} \quad \text{S C C H O E N B E R G} \\
 \text{String 2} \quad \text{S C H _ _ U B E R T} \\
 \hline
 \text{Edit Transcript} \quad \text{M M M D D S M M M S}
 \end{array} \tag{9.1}$$



1 Since there are two deletions and two substitutions, the edit distance in this case is
2 4—if deletions and substitutions are given the same weight, though these amounts
3 may be varied to achieve different matching behaviours. To arrive at a similarity value,
4 rather than a distance metric, and to confine that value to the interval (0,1), we divide
5 the distance value by the maximal distance (the length of the longer string) and sub-
6 tract the result from one:

$$7 \quad \sigma(s_1, s_2) = 1 - \frac{d_{edit}(s_1, s_2)}{\max(|s_1|, |s_2|)} \quad (9.2)$$

8 For our example, we thus obtain an edit distance similarity value of $1 - 4/10 = 0.6$. Of
9 course, this same measure works with symbols representing pitch, intervals or rhyth-
10 mic categories; all that we need is the ability to decide whether two symbols are the
11 same, regardless of their meaning. What makes it attractive for music researchers is
12 the fact that the quality of the operations (insertions, deletions, and substitutions)
13 have parallels in musical composition, where it is common practice to produce varia-
14 tions or related pieces of music by, for example, inserting ornamental notes or by
15 replacing certain structurally important notes by others that may serve a similar—or
16 different—harmonic purpose. This neat simplicity is necessarily lost if we wish to
17 consider more than one musical dimension at once, because of the intimate relation-
18 ship between the dimensions of music, but Mongeau and Sankoff (1990) have adapted
19 the measure to work better in these circumstances.

20 The description above, however, covers only the measure, and not the algorithm
21 required to compute it. For edit distance to be useful, two strings must be aligned
22 *optimally*—that is, with the smallest possible distance. This is often achieved, in $O(n^2)$
23 time and memory, using *dynamic programming* (Gusfield, 1997), whose implementa-
24 tion is straightforward. This, and the fact that the measure itself and its outputs are
25 mostly easy to understand, have made edit distance a popular tool in computational
26 music analysis in recent times.

27 Revisiting Sloboda's and Parker's results

28 Motivation

29 We now revisit the results of Sloboda's and Parker's paper as discussed in their nine
30 conclusions (Sloboda & Parker, 1985, pp. 159–160). The primary purpose of reanalys-
31 ing the musical data from the original study is not to refute or support the authors'
32 findings, but rather to show how analytical algorithms can be usefully employed in
33 cognitive studies, how they might help to cope with large amounts of difficult-to-an-
34alyse musical data, and how they might shed a different light on some analytical pro-
35 cedures that are commonly executed by human (as opposed to computational)
36 effort.

37 As musical data, we used the 48 transcribed recalls that are given in appendix
38 by Sloboda and Parker (1985). We applied similarity measurements and other algo-
39 rithmic analysis methods to the research questions and results summarized at the

1 beginning of their conclusion section. We limit ourselves to discussion of six out of
2 the nine results, where the application of computational methods seem most straight-
3 forward and productive, and omit Sloboda and Parker's main findings 4, 5, and 7.

4 **Main finding 1: 'No recall is perfect'**

5 We measured the similarity between the target melody and each recall using the edit
6 distance algorithm (see above). To have a 'second opinion' from a similarity algorithm
7 that makes use of musical background information from a corpus, we also applied an
8 asymmetric similarity algorithm proposed by Müllensiefen and Pendzich (2009),
9 which is based on a conception of feature similarity proposed by Tversky (1977), the
10 *Tversky.target.only* algorithm. We applied both algorithms only to the *pitch sequences*
11 of the melodies. The maximal similarity values were 0.93, from edit distance, and 0.98,
12 from the Tversky similarity algorithm. Both were measured for participant 5 on her
13 sixth trial. Sloboda and Parker's assertion that, even for this simple folk tune and with
14 some musically trained participants, there is no single 'note-for-note perfect' recall
15 among the 48 attempts. The computational analysis corroborates this surprising limi-
16 tation of human melody recall.

17 **Main finding 2: 'Recalls are highly related to the original in 18 many respects'**

19 We measured the similarity between recalls and the target melody using our simple
20 edit distance algorithm, but feeding it different types of musical information, viz.,
21 sequences of pitches, sequences of interpolated contour values, sequences of catego-
22 rized rhythmic (duration) values, and sequences of implied tonalities at the bar level
23 as derived from the Krumhansl-Schmuckler algorithm for tonality induction
24 (Krumhansl, 1990). How these different types of musical information are obtained,
25 through appropriate transformations from the raw melodic pitch and onset data, is
26 documented by Müllensiefen and Frieler (2004a). We obtained mean similarity val-
27 ues, averaged over all 48 recalls, of 0.37 for pitch similarity, 0.21 for contour similarity,
28 0.56 rhythmic similarity, and 0.20 for similarity of implied tonalities. Taking into
29 account that these values are averaged over all six trials and therefore include low simi-
30 larity values from initial (and so mostly incomplete) trials, it seems fair to say that the
31 recorded recalls were related to the target melody in many different respects. In par-
32 ticular, the similarity is found to some extent in all the dimensions tested. Thus, the
33 original findings are supported.

34 **Main finding 3: 'Metrical structure is preserved in almost all 35 recalls'**

36 We used the beat and metre extraction model of Frieler (2004) to induce the metre of
37 the target melody and the recalls. The model uses temporal smoothing with Gaussian
38 kernels, accent rules based on note durations and autocorrelation to determine the
39 beat level, metre, and metrical phase from monophonic input. The model induced a
40 2/4 metre for the target melody, which is not unreasonable; the melody being actually

1 notated in 4/4 may simply be because 4/4 metres are more common in Western
 2 musical styles. Eighty one per cent of the recalls were classified as having a 2/4 metre
 3 and 19% were classified as 4/4. So all trials had a simple duple metre, like the target
 4 melody. However, since 4/4 is the default for most (popular) Western music, this
 5 result is not so surprising; it would be interesting to apply the approach to recalls
 6 where the target was in triple and/or compound meter.

7 **Main finding 6: 'Subjects vary significantly regarding their** 8 **melodic and harmonic accuracy'**

9 We compared the recalls of musical novices (participants 1–4) and musicians (parti-
 10 cants 5–8) with regard to their melodic and harmonic accuracy (i.e. similarity to the
 11 target melody), as did Sloboda and Parker. We measured pitch edit distance to esti-
 12 mate pitch accuracy. Novices achieved a mean similarity of 0.30 over all trials and
 13 musicians reached a mean of 0.45, which proved to be a significant difference ($t(30.4)$
 14 $= 2.86$, $p = 0.004$). Similarly, mean harmonic similarity as measured by the edit dis-
 15 tance over implied tonalities differed significantly between novices ($\bar{x}_{novices} = 0.09$) and
 16 musicians ($\bar{x}_{musicians} = 0.31$; $t(28.5) = 3.17$, $p = 0.002$). Although the difference between
 17 novices and musicians is significant for both measures, the it is much larger for the
 18 harmonic comparison. This supports Sloboda and Parker's conclusion that 'memory
 19 for harmonic structure seems to be related to musical expertise' (1985, p. 160).

20 We also compared rhythmic similarity values between the two groups. The resulting
 21 t-test shows no significant difference between the two groups of participants ($\bar{x}_{novices} =$
 22 0.59 ; $\bar{x}_{musicians} = 0.54$; $t(45.5) = -0.75$, $p = 0.77$). Sloboda and Parker observed that
 23 while there are considerable differences between listeners regarding the memory
 24 retention of harmonic structure, metrical structure was abstracted well by all partici-
 25 pants. This finding seems now to extend to rhythmic structure as well.

26 **Main finding 8: 'Musicians and non-musicians differ in** 27 **retention of harmony'**

28 We looked at the development of the representation of harmonic structure in both
 29 groups of participants over the six trials, with main finding 6 (above) in mind. We
 30 averaged the harmonic similarity values, as measured by the edit distance on the
 31 implied tonalities, for each trial number over each participant group. The develop-
 32 ment is illustrated in Figure 9.1.

33 Figure 9.1 shows a clear increase in harmonic similarity between the musicians'
 34 recalls and the target melody from trial 4 on while the harmonic similarity between
 35 recalls of novices and the target melody stay at the same low level of similarity over all
 36 six trials. In contrast, no learning effect or improvement of accuracy in the mental
 37 representations seems to take place for the novices. Accordingly, a dependent sample
 38 t-test between novices' and musicians' mean recall similarity values indicated a signifi-
 39 cant difference ($\bar{x}_{differences} = 0.22$; $t(5) = 3.16$; $p = 0.01$). Therefore, it seems evident
 40 that the perceptual and cognitive abstraction and refinement of harmonic structure is
 41 achieved only by participants with some musical training.

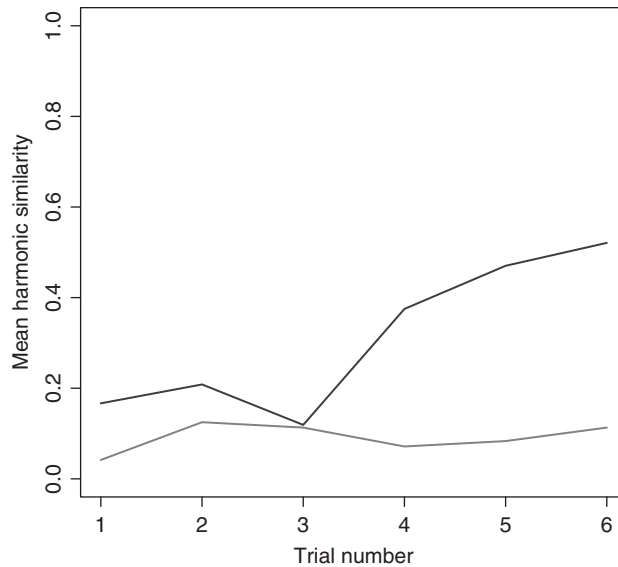


Fig. 9.1 Mean similarity between participants' recalls and target melody over trials. Blue: musicians; red: novices.

1 Main finding 9: 'Subjects do not show improvement on any 2 measure over the six trials'

3 This is one of Sloboda and Parker's most interesting conclusions, since it seems to
4 imply that human memory for melodies does not increase in accuracy with repeated
5 exposure. To test whether there is actually no such learning effect, we averaged for
6 each trial similarity for pitch, harmonic, rhythmic, and overall similarity (measure
7 *opti2* combining duration weighted pitch values and a similarity measure on short
8 pitch motives, see Müllensiefen & Frieler, 2004a, for details) over all participants and
9 computed Pearson's correlation, r , between trials number and means similarity for
10 each similarity measure. We found high and significant values of r for all similarity
11 measures as given in Table 9.2. Figure 9.2 illustrates these correlations.

12 These correlations do suggest that a learning effect is taking place over repeated tri-
13 als. The effect seems to be less strong for the rhythmic aspect of the memory represen-
14 tation than for the pitch and harmonic aspects. So, while Sloboda and Parker assert
15 that recalls get longer over trials (starting at a mean of 16.5 notes for trial 1 and monotonically
16 increasing to 26.625 notes for trial 6), we found that recalls also get more
17 accurate and that participants also seem to refine their memory representations.

18 This is in line with results of Zielinska and Miklaszewski (1992), who found (using
19 10 instead of 6 trials) an increase in the number of correctly recalled half-measures.

20 Discussion

21 The translation of the six main analyses discussed in the previous section into compu-
22 tational methods proved to be very simple and straightforward. Translating main

Table 9.2 Values of Pearson's correlation coefficient r with corresponding p values for different similarity measures measuring the correlation between trial number and mean similarity with the target melody

| Similarity measure | r | p Value |
|----------------------------------|------|-----------|
| Edit distance, pitch | 0.93 | 0.008 |
| Edit distance, implied tonality | 0.93 | 0.008 |
| Edit distance, rhythmic category | 0.77 | 0.08 |
| <i>opti2</i> | 0.91 | 0.01 |

1 findings 4, 5, and 7 into computational form requires some more detailed argumenta-
 2 tion; they can probably be realized in more than one way.² Therefore, we defer these
 3 for future work, because the purpose of this chapter is merely to support our claim
 4 that analytical tasks in music cognition research can be solved computationally.

5 For main finding 4, 'subjects preserve phrase structure', one possible analysis would
 6 have been to compute, for each recall, the similarity between all possible pairs that can
 7 be created from the three phrases, A_1 , A_2 , and B . The overall similarity between pair
 8 (A_1 , A_2) should then be higher than the similarity between pairs (A_1 , B) and (A_2 , B).

9 The translation of main finding 5, 'original rhythms are often replaced by metrical
 10 equivalents', would require the measurement of rhythmic similarity at the phrase

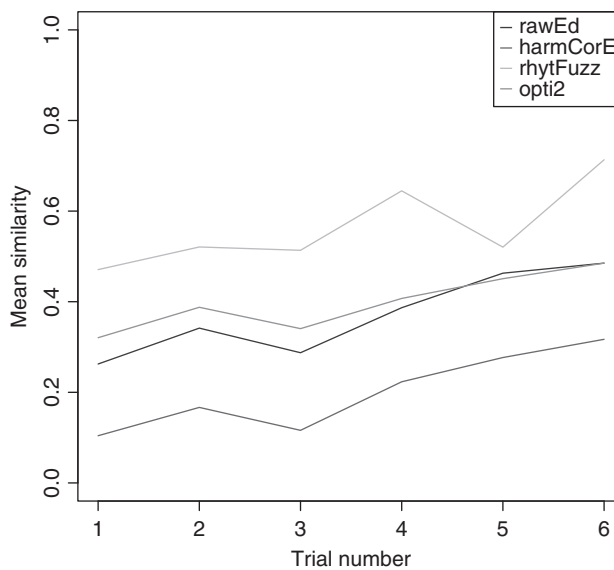


Fig. 9.2 Mean similarity over trials according to different similarity measures; colours correspond with the different measures.

² Specifically, for main findings 4 and 5 the phrases all recalls need to be identified initially to enable the subsequent phrase-level analyses.



1 level, between original and recalled phrase, and a metre induction procedure giving
 2 both metre and phase information. The test would then be that, for most phrases, the
 3 induced metre and phase is identical to the original but rhythmic similarity to the
 4 original varies to a greater extent.

5 The claim from main finding 7, 'harmonic structure is preserved even when exact
 6 melodic structure is lost', is similar to main finding 5: Sloboda and Parker observe that
 7 while the recall of melodic details shows a greater variation, the more abstract har-
 8 monic structure, having fewer degrees of freedom, is less likely to vary and is often
 9 close to the original. This might be demonstrated by comparing the variation in har-
 10 monic with melodic similarity measurements, within a series of repeated
 11 measurement.

12 **Conclusion and outlook**

13 Several central questions motivated Sloboda and Parker's original study, including:

- 14 ♦ How do memory representations build up over repeated exposure to the same
 15 novel melody?
- 16 ♦ What melodic parameters or dimensions are easier to grasp and to encode and
 17 which ones or more difficult?
- 18 ♦ How do people with and without musical training differ with respect to building
 19 up representations of melodies in memory?

20 In this chapter, we have demonstrated how a computational approach can help in
 21 answering these questions, and how the nature of the answers changes when a compu-
 22 ter is involved as a tool for analysis and modelling. We have seen answers become
 23 necessarily very precise and quantitative; this we consider to be an advantage, espe-
 24 cially if experiments are repeated with the same method but on different data or with
 25 different stimuli. However, one must take care not to confuse this necessary precision
 26 with greater validity—indeed, the converse can easily be the case. Therefore, in order
 27 to obtain valid results, the questions that the computer is supposed to answer must
 28 also be asked very precisely. This is no bad thing: the necessity to make explicit every
 29 tiny step from the formulation of the question (via the choice and implementation of
 30 a formal procedure) to the interpretation of the results can often win us additional
 31 insight into the phenomena under study. As can be seen from our discussion of
 32 melodic similarity algorithms, the use of computational models and algorithms offers
 33 the researcher a huge choice of analytic procedures—but, in consequence, that
 34 researcher must take responsibility for their decision and argue it. Though unavoida-
 35 bly painful at the beginning, this extra rigour in several aspects of methodology, can
 36 ultimately only be good for the field.

37 While we certainly do not suggest that the edit distance measure is necessarily the
 38 best measure of melodic similarity in the context of this experimental data, we chose
 39 it deliberately because it is easy to understand, and it is at least loosely connected to
 40 semi-formal procedures that musicologists use to determine similarity relations and
 41 derivative procedures between melodies. However, it would be interesting to reanalyse
 42 this experimental dataset with more complex similarity models that are claimed to



1 have more cognitive validity (e.g. Müllensiefen and Frieler, 2004a; Müllensiefen and
2 Pendzich, 2009). Then, employment of a computational procedure for determining
3 melodic similarity would change status, from merely using an adequate tool, to mod-
4 elling human behaviour with a cognitive-computational model—an epistemic transi-
5 tion for which we have argued in detail elsewhere (Pearce, Müllensiefen, & Wiggins,
6 2009; Wiggins, Pearce, & Müllensiefen, 2009).

7 Apart from answering Sloboda and Parker's questions in a computational way that
8 is of a different quality, the use of computers to analyse and model data resulting from
9 their experimental paradigm also enables us to ask some additional questions. Because
10 these questions entail comparison across different melodic stimuli, we conducted a
11 new experiment with Sloboda and Parker's design but using 14 different melodies and
12 30 subjects with acceptable singing abilities. As a result, we obtained about 1900 usable
13 sung recalls which have been transcribed by a professional human transcriber. These
14 data are about to be analysed using the computational approach proposed here. The
15 additional questions that we hope to answer from the analysis of this new dataset con-
16 cern mainly features of melodic structure, and they aim at being generalizable to the
17 memory processing of Western popular melodies in general:

- 18 ♦ How do musical features affect the recall of melodies?
- 19 ♦ What makes a melody easy or difficult to recall?
- 20 ♦ Which parts of a melody are represented first and most accurately?
- 21 ♦ Does commonness or rarity of melodic features play a role?
- 22 ♦ Are melodies in their song context and as audio excerpts recalled better than sin-
23 ggle-line melodies or is this vice versa?

24 To facilitate this type of feature analysis, we have developed an open source software
25 toolbox, called FANTASTIC,³ which computes summary and sequence-based features
26 of monophonic melodies. FANTASTIC also enables researchers to model melody
27 perception and cognitive processing in the context of a corpus of melodies and, there-
28 fore, it can take into account previous listening experience given a suitable corpus of
29 music. The software has already proven useful in the analysis of experimental data
30 from a recognition paradigm similar to the one used by Halpern and Müllensiefen
31 (2008), and preliminary results suggest that objective measures of implicit and explicit
32 memory performance, as well as subjective memory measures including false alarms
33 and misses, can be explained to a certain degree by features of the melodic structure
34 (Halpern, Müllensiefen, & Wiggins, 2008). Once our recognition dataset is fully ana-
35 lysed, we aim to model our recall data with the same feature approach, to see whether
36 recall and recognition memory can be explained by the same structural features. This
37 would supply strong evidence for the hypothesis that the performance of musical
38 memory is indeed dependent on musical structure, and that we can explain and pre-
39 dict cognitive behaviour, at least partially, from the structure of the music itself.

³ Feature ANalysis Technology Accessing SStatistics (In a Corpus); download from our project
website: <http://www.doc.gold.ac.uk/isms/mmm/?page=Software\%20and\%20Documentation>.

1 Acknowledgements

2 This research was supported by EPSRC project EP/D0388551, *Modelling Musical*
3 *Memory and the Perception of Melodic Similarity*. It was inspired by the work of John
4 Sloboda.

5 References

- 6 Abdallah, S., Sandler, M., Rhodes, C., & Casey, M. (2006). Using duration models to reduce
7 fragmentation in audio segmentation. *Machine Learning, Special Issue on Machine Learning*
8 *for Music*, 65(2–3), 485–515.
- 9 Aloupis, G., Fevens, T., Langerman, S., Matsui, T., Mesa, A., Rappaport, D., &
10 Toussaint, G. (2003). Computing the similarity of two melodies. In *Proceedings of*
11 *the 15th Canadian Conference on Computational Geometry*. Halifax, Canada: Dalhousie
12 University.
- 13 Anderson, C., & Borkowski, J. G. (1978). *Experimental psychology*. Glenview, IL: Scott,
14 Foresman and Company.
- 15 Atkinson, R., & Shiffrin, R. (1968). Human memory: a proposed system and its control
16 processes. In K. Spence, & J. Spence (Eds.), *The psychology of learning and motivation*. New
17 York, NY: Academic Press.
- 18 Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.
- 19 Baddeley, A. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.
- 20 Calkins, M. (1894). Association. *Psychological Review*, 1, 476–483.
- 21 Cook, N. (2008). Techniques for analysing recordings. In N. Cook, E. F. Clarke, D. Leech-
22 Wilkinson, & J. Rink (Eds.), *The Cambridge companion to recorded music*. Cambridge, UK:
23 Cambridge University Press.
- 24 Crawford, T., Iliopoulos, C., & Raman, R. (1998). String-matching techniques for musical
25 similarity and melodic recognition. *Computing in Musicology*, 11, 73–100.
- 26 Davies, J., & Yelland, A. (1977). Effects of two training procedures on the production of
27 melodic contour, in short-term memory for tonal sequences. *Psychology of Music*, 5(2),
28 3–9.
- 29 Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory.
30 *Science*, 168, 1604–1605.
- 31 Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*,
32 11(6), 411–412.
- 33 Deutsch, D. (1974). Generality of inference by tonal stimuli in recognition memory for pitch.
34 *Quarterly Journal of Experimental Psychology*, 26, 229–234.
- 35 Deutsch, D. (1975a). The organization of short-term memory for a single acoustic attribute.
36 In D. Deutsch, & J. Deutsch (Eds.), *Short term memory*. New York, NY: Academic Press.
- 37 Deutsch, D. (1975b). Two-channel listening to musical scales. *Journal of the Acoustical Society*
38 *of America*, 57, 1156–1160.
- 39 Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception*
40 *& Psychophysics*, 28(5), 381–389.
- 41 Deutsch, D., & Feroe, J. (1975). Disinhibition in pitch memory. *Perception & Psychophysics*, 17,
42 40–51.
- 43 Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music.
44 *Psychological Review*, 88, 503–522.





- 1 DeWitt, L., & Crowder, R. (1986). Recognition of novel melodies after brief delays. *Music*
2 *Perception*, 3(3), 259–274.
- 3 Dowling, W., Tillmann, B., & Ayers, D. (2002). Memory and the experience of hearing music.
4 *Music Perception*, 19(2), 136–149.
- 5 Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and
6 retrograde inversion. *Perception & Psychophysics*, 12(5), 417–421.
- 7 Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for
8 melodies. *Psychological Review*, 85(4), 341–354.
- 9 Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term
10 memory for melodies. *Psychomusicology*, 1, 30–49.
- 11 Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory
12 for melodies. *Journal of the Acoustical Society of America*, 49(2, Part 2), 524–531.
- 13 Downie, J. S. (2003). *Evaluating a simple approach to music information retrieval. Evaluating a*
14 *simple approach to music information retrieval. Conceiving melodic n-grams as text.* Ph.D.
15 thesis. London (Ontario), Canada: Faculty of Information and Media Studies, University of
16 Western Ontario.
- 17 Drake, C., Dowling, W., & Palmer, C. (1991). Accent structures in the reproduction of simple
18 tunes by children and adult pianists. *Music Perception*, 8(3), 315–334.
- 19 Ebbinghaus, H. (1885). *U das Gedächtnis. Untersuchungen zur experimentellen Psychologie.*
20 Amsterdam: Bonset.
- 21 Eck, D. (2002). Finding downbeats with a relaxation oscillator. *Psychological Research*, 66(1),
22 18–25.
- 23 Edworthy, J. (1983). Towards a contour-pitch continuum theory of memory for melodie.
24 In D. Rogers, & J. Sloboda (Eds.), *The acquisition of symbolic skills* (pp. 263–271).
25 New York, NY: Plenum.
- 26 Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, 2(3),
27 375–388.
- 28 Eerola, T., Järvinen, T., Louhivuori, J., & Toivianen, P. (2002). Statistical features and
29 perceived similarity of folk melodies. *Music Perception*, 18(3), 275–296.
- 30 Frieler, K. (2004). Beat and meter extraction using gaussified onsets. In C. Lomeli Buyoli, &
31 R. Loureiro (Eds.), *Proceedings of the 5th International Conference on Music Information*
32 *Retrieval* (pp. 178–183). Barcelona: Universitat Pompeu Fabra.
- 33 Ginsborg, J., & Sloboda, J. A. (2007). Singers' recall for the words and melody of a new,
34 unaccompanied song. *Psychology of Music*, 35(3), 421–440.
- 35 Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics.* New York, NY: Wiley.
- 36 Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: Computer science and*
37 *computational biology.* Cambridge: Cambridge University Press.
- 38 Halpern, A. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*,
39 17(5), 572–581.
- 40 Halpern, A., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for
41 music. *Quarterly Journal of Experimental Psychology*, 61 (9), 1371–1384.
- 42 Halpern, A., Müllensiefen, D., & Wiggins, G. (2008). *Modelling memory responses in a memory*
43 *recognition task.* Talk at the 10th International Conference for Music Perception and
44 Cognition (ICMPC10), Sapporo (Japan).
- 45 Huron, D. (1995). *The humdrum toolkit: Reference manual.* Menlo Park, CA: Center for
46 Computer Assisted Research in the Humanities.





- 1 Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA:
2 MIT Press.
- 3 Idson, W., & Massaro, D. (1978). A bidimensional model of pitch in the recognition of
4 melodies. *Perception & Psychophysics*, 24(6), 551–565.
- 5 Kantowitz, B. H., Roediger, H. L. I., & Elmes, D. G. (1994). *Experimental psychology:
6 understanding psychological research*. Minneapolis: West.
- 7 Keller, T., Cowen, N., & Saults, J. (1995). Can auditory memory for tone pitch be rehearsed?
8 *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(3), 635–645.
- 9 Kluwe, R. (1990). Gedächtnis und Wissen. In H. Spada (Ed.), *Allgemeine psychologie*
10 (pp. 115–187). Stuttgart: Hans Huber.
- 11 Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford Psychology Series 17.
12 Oxford: Oxford University Press.
- 13 Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of
14 learned melodies. *Perception & Psychophysics*, 56(4), 414–423.
- 15 Long, P. (1977). Relationships between pitch memory in short melodies and selected factors.
16 *Journal of Research in Music Education*, 25(4), 272–282.
- 17 Longuet-Higgins, C., & Steedman, M. (1971). On interpreting Bach. *Machine Intelligence*, 6,
18 221–242.
- 19 Mauch, M., Müllensiefen, D., Dixon, S., & Wiggins, G. (2008). What's a work in harmony? In:
20 K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, and M. Tsuzaki (Eds.), *Proceedings of the
21 10th International Conference for Music Perception and Cognition* (pp. 683–689), 25–29
22 August 2008, Sapporo, Japan (CD). Adelaide: Causal Productions.
- 23 Meek, C., & Birmingham, W. (2002). Johnny can't sing. a comprehensive error model for sung
24 music queries. In M. Fingerhut (Ed.), *Proceedings of the Third International Conference on
25 Music Information Retrieval* (pp. 124–132). Paris: IRCAM.
- 26 Miller, G. (1956). The magical number seven plus or minus two: some limits on our capacity
27 for processing information. *Psychological Review*, 63, 81–97.
- 28 Milmeister, G., Mazzola, G. B., & Thalmann, F. (2009). *The Rubato composer musicsoftware*.
29 Berlin: Springer.
- 30 Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the
31 Humanities*, 24, 161–175.
- 32 Mueller, J. (1970). Response properties of the position indicant in serial learning. *Journal of
33 Experimental Psychology*, 84, 35–39.
- 34 Müllensiefen, D., & Frieler, K. (2004a). Cognitive adequacy in the measurement of melodic
35 similarity: Algorithmic vs. human judgments. *Computing in Musicology*, 13, 147–176.
- 36 Müllensiefen, D., & Frieler, K. (2004b). Optimizing measures of melodic similarity for the
37 exploration of a large folk-song database. In *Proceedings of the 5th International Conference
38 on Music Information Retrieval* (pp. 274–280). Barcelona: Universitat Pompeu Fabra.
- 39 Müllensiefen, D., & Pendsch, M. (2009). Court decisions on music plagiarism and the
40 predictive value of similarity algorithms. *Musicae Scientiæ, Discussion Forum 4B*, 257–295.
- 41 Müllensiefen, D., Pfeleiderer, M., & Frieler, K. (in press). The perception of accents in pop
42 music melodies. *Journal of New Music Research*, 38(1), 19–44.
- 43 Murdock, B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*,
44 64, 482–488.
- 45 Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18,
46 251–269.





- 1 North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for
2 popular music. *Psychomusicology*, 14, 77–93.
- 3 O Mardin, D. (1998). A geometrical algorithm for melodic difference in melodic similarity.
4 *Computing in Musicology*, 11, 65–72.
- 5 Ogawa, Y. T. K., & Mito, H. (1995). Modification of musical schema for Japanese melody:
6 A study of comprehensible and memorable melody. *Bulletin of the Council for Research in*
7 *Music Education*, 127, 136–141.
- 8 Oura, Y., & Hatano, G. (1988). Memory for melodies among subjects differing in age and
9 experience in music. *Psychology of Music*, 16, 91–109.
- 10 Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the*
11 *cerebral cortex* (G. V. Anrep, Trans.). London: Oxford University Press.
- 12 Pearce, M. T., Müllensiefen, D., & Wiggins, G. (2009). "Melodic grouping in music
13 information retrieval: New methods and applications. In R. Zbigniew, & A. Wiczorkowska
14 (Eds.), *Advances in music information retrieval*, Studies in Computational Intelligence.
15 Berlin: Springer.
- 16 Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). Melodic grouping in music
17 information retrieval: new methods and applications. In: R. Zbigniew & A. Wiczorkowska
18 (Eds.), *Advances in music information retrieval. studies in computational intelligence*, vol. 274
19 (pp. 365–390). Berlin: Springer.
- 20 Pearce, M. T., & Wiggins, G. A. (2004). Improved methods for statistical modelling of
21 monophonic music. *Journal of New Music Research*, 33(4), 367–385.
- 22 Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and
23 learning. *Music Perception*, 23(5), 377–406.
- 24 Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a
25 working-memory model. *Memory & Cognition*, 20(3), 314–320.
- 26 Pickens, J., Bello, J. P., Monti, G., Sandler, M. B., Crawford, T., Dovey, M., & Byrd, D. (2003).
27 Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling
28 approach. *Journal of New Music Research*, 32(2), 223–236.
- 29 Rundus, D. (1974). Output order and rehearsal in multi-trial free recall. *Journal of Verbal*
30 *Learning and Verbal Behaviour*, 13, 656–663.
- 31 Schulkind, M. D. (2000). Perceptual interference decays over short unfilled intervals.
32 *Memory & Cognition*, 28(6), 949–956.
- 33 Schulkind, M. D. (2004). Serial processing in melody identification and the organization of
34 musical semantic memory. *Perception & Psychophysics*, 66(8), 1351–1362.
- 35 Schulkind, M. D., Posner, R. J., & Rubin, D. C. (2003). Musical features that facilitate melody
36 identification: How do you know it's 'your' song when they finally play it? *Music*
37 *Perception*, 21(2), 217–249.
- 38 Shepard, R. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal*
39 *Learning and Verbal Behaviour*, 6, 156–163.
- 40 Shiffrin, R., & Nosofsky, R. (1994). Seven plus or minus two: A commentary on capacity
41 limitations. *Psychological Review*, 101(2), 357–361.
- 42 Sloboda, J. A., & Parker, D. H. (1985). Immediate recall of melodies. In P. I. C. Howell, &
43 R. West (Eds.), *Musical structure and cognition* (pp. 143–167). London: Academic Press.
- 44 Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.
- 45 Thomassen, J. (1982). Melodic accent: Experiments and a tentative model. *Journal of the*
46 *Acoustical Society of America*, 71, 1596–1605.





- 1 Tulving, E. (1968). When is recall higher than recognition? *Psychonomic Science*, 10, 53–54.
- 2 Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
- 3 Typke, R., Wiering, F., & Veltkamp, R. (2007). Transportation distances in human perception
4 of melodic similarity. *Musicae Scientiae, Discussion Forum 4A*, 153–181.
- 5 Uitdenbogerd, A. L. (2002). *Music information retrieval rechnology*. Ph.D. thesis, RMIT
6 University of Melbourne, Australia.
- 7 Unal, E., Chew, E., Georgiou, P., & Narayanan, S. (2008). Challenging uncertainty in query by
8 humming systems: A fingerprinting approachenging. *IEEE Transactions on Audio, Speech,*
9 *and Language Processing*, 16(2), 359–371.
- 10 Underwood, B., Rundquist, W., & Schulz, R. (1959). Response learning in paired-associate lists
11 as a function of intralist similarity. *Journal of Experimental Psychology*, 58, 70–78.
- 12 Volk, A. (2008). Persistence and change: Local and global components of metre induction
13 using inner metric analysis. *Journal of Mathematics and Computation in Music*, 2(2),
14 99–115.
- 15 Weyde, T. (2004). Modeling rhythmic motif structure with fuzzy logic and machine learning in
16 music query: Methods, models, and user studies. *Computing in Musicology*, 13, 35–50.
- 17 Wiggins, G., Pearce, M., & M`ullensiefen, D. (2009). Computational modelling of music
18 cognition and musical creativity. In R. Dean (Ed.), *The Oxford handbook of computer music*
19 *and digital sound culture*. Oxford: Oxford University Press.
- 20 Wiggins, G. A. (2007). Models of musical similarity. *Musicae Scientiae, Discussion Forum 4A*,
21 315–337.
- 22 Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2006). Music in working memory?
23 examining the effect of pitch proximity an the recall performance of nonmusicians.
24 In M. Baroni, R. C. Addressi, & M. Costa (Eds.), *Proceedings of the 9th International*
25 *Conference of Music Perception and Cognition (ICMPC9)* (pp. 1581–1589). Bologna:
26 Society of Music Perception and Cognition (SMPC) AND European Society for the
27 Cognitive Sciences of Music (ESCOM).
- 28 Young, R. (1962). Tests of three hypotheses of the effective stimulus in serial learning. *Journal*
29 *of Experimental Psychology*, 63, 307–313.
- 30 Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social*
31 *Psychology Monograph Supplement*, 9, 1–27.
- 32 Zielinska, H., & Miklaszewski, K. (1992). Memorising two melodies of different style.
33 *Psychology of Music*, 20, 95–111.

