Effects of song familiarity, singing training and recent song exposure on the singing of melodies

Steffen Pauws
Philips Research Laboratories Eindhoven
Prof. Holstlaan 4
5656 AA Eindhoven, the Netherlands
steffen.pauws@philips.com

Abstract

Findings of a singing experiment are presented in which trained and untrained singers sang melodies of familiar and less familiar Beatles songs from memory and after listening to the original song on CD. Results showed that adopting the correct pitch of a melody was done better by trained singers, and only after listening to the song. Contours of melodies were equally well reproduced by both trained and untrained singers. In contrast, the intervals of a melody were sung more accurately by trained singers than by untrained singers. These findings demonstrate the dominance of contour for remembering melodies and the poorer interval encoding of melodies or the lack of essential singing skills by untrained singers. When singing from memory, almost two-third of the singing came reasonably close to the actual tempo on the CD. This improved to more than 70% after listening to the song on CD. In general, the singing of less familiar melodies improved after song listening. Implications for 'query by humming' applications are discussed.

1 Introduction

Allowing people to retrieve a song from a large volume by letting them sing a 'hook-line' of that song is often presented as a compelling means to facilitate music selection if only a sense of the melody is known, but no name of the song. However, these so-called 'query by humming' applications are challenged by the singing performance of the general public, if the target user group is the general public. Unfortunately, studies on singing performances are hardly available, though 'query by humming' applications need to capitalize on knowledge about everyday singing.

Therefore, the goal of the current study was to formulate hypotheses from findings of some relevant existing studies and to test these hypotheses in a controlled singing experiment using everyday music. The experiment examined the effects on singing melodies of familiar and less familiar songs of 'The Beatles', effects of being a trained singer or not and effects of singing from memory or after recent song exposure (i.e., CD listening).

To anticipate the results, it was found that familiar songs were better reproduced than less familiar songs by both trained and untrained singers. Both groups of singers had bad memory for the correct pitch of a song; trained singers were better in adopting the correct pitch after listening to the song on CD. Both groups did have good memory for the correct global tempo of a song. Both groups did not perform differently in singing the global motion of the melody (i.e., the contour) correctly. Trained singers were more accurate in singing the more detailed melodic motion (i.e., the intervals). Both groups improved their singing after listening to the original song. These findings have implications for 'query by humming' applications.

1.1 Memory for melodies

What makes it that we can identify a melody as 'heard-once-before', recognize it, distinguish it from others, or even recall and sing it? Empirical work that tries to find answers for the recognition (but not the reproduction) of short, isochronous and rather 'non-musical' melodies has been published (see, e.g., Croonen, 1994; Dowling, 1978; Edworthy, 1985). In these studies, melodies are distinguished along their perceptual properties, such as pitch, loudness, tempo, timbre, contour and rhythm.

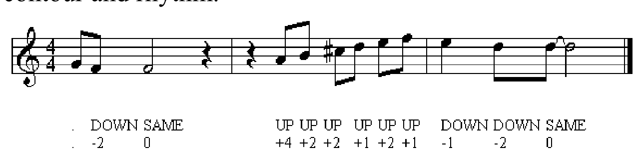

Figure 1: The first three measures of the vocal melody of 'Yesterday' in musical score notation. The melodic contour and musical intervals are shown underneath.

Pitch, loudness, tempo and timbre are human judgments of how high or low, how loud or soft, how fast or slow and in what quality a musical sound is heard. As a standard for western music, the continuous percept of pitch is subdivided into a set of discrete logarithmic steps that hinges on a reference tone (A4 = 440 Hz, the concert pitch). Each step is denoted as a semitone. Musical notes are arranged along these steps as musical scales. The distance between any two notes is called a musical interval and is expressed by the number of semitones between them. A melodic contour is a crude representation of the intervals between successive notes in a melody. It ignores the magnitude of the intervals and retains only the information whether the melody is going up, is going down or does not change in pitch (see Figure 1). Lastly, rhythm is the perception of temporal patterns that (hierarchically) groups a sequence of notes along a steady beat. Beat is known as the perceived pulse in music marking off equal time units in a metrical structure; the latter defines the periodic alternation of strong and weak accents. The question arises now what perceptual properties of a melody are most important in allowing us to encode, recognize, remember and eventually sing melodies correctly.

Many studies report that the perception of a melody changes only slightly under modest transformations along pitch, loudness, tempo and timbre. Except for some extreme variations, it is evident that the perceived melody does not change much if it is played in different keys, louder or softer, faster or slower or by different musical instruments or voices (Dowling and Harwood, 1986).

However, since rhythm and pitch are hard to attend to selectively, recognition of a known melody presented in another rhythm is difficult (Jones, Summerell and Marshburn, 1987). Rhythmic accents placed on salient melody events facilitate the remembrance of melodies (Boltz and Jones, 1986). Though a contour maintains only the global motion of 'ups' and 'downs' in a melody, it is one of the main pillars for memorizing melodies. Children (under the age of 5 years) have only the contour as a device to remember melodies. But also for adults, melodies with the same contour are easily confused (Dowling, 1978). Well-known melodies in which the intervals are distorted but the contours are preserved are still easily recognized (Dowling and Harwood, 1986). New melodies or melodies that we have heard only a couple of times are largely remembered by their contour. The intervals of a melody become only more important with increasing song familiarity, after childhood and after musical training (Davidson, 1994; Dowling and Harwood, 1986).

1.2 Singing melodies

Singing means articulating a recalled melody. This task is more complex than it appears to be at first. Besides recalling the contour, intervals and rhythm of a melody, it involves a delicate mastery of several muscle systems controlled by auditory feedback.

Obviously, singing ability and training[1] are deemed to be important factors to reproduce a melody accurately. People without singing training rely heavily on contours, instead of knowledge about intervals and scale structures. Consequently, their singing performances are flawed by large intervals being flattened and by ending a melody line in a different key than that was used at the start. Having a contour as their main recourse, untrained singers can sing a melody only in one way. Changing their singing, including singing in a different key or mode, or reflecting on and improving their own performance, is more problematic for them (Davidson, 1994).

However, it has been shown that people, irrespective of singing ability, are surprisingly good at singing their *favourite* song at the correct pitch and at the correct tempo (Levitin, 1994; Levitin and Cooke, 1997).

2 Experiment

The experiment investigated the influences of singing training (trained and untrained singers), song familiarity (familiar and less familiar songs) and recent song exposure on singing melodies of Beatles songs. Based on findings reported in existing studies, it was expected that

- the correct pitch and the correct tempo can be easily adopted,
- proportionally more intervals of familiar melodies are sung correctly than of less familiar melodies,
- trained singers sing proportionally more correct intervals than untrained singers do, but this difference is not found in contour reproduction,
- trained singers improve their singing after having listened to the original song, while untrained singers do not.

2.1 Material

The stimulus material consisted of 12 Beatles songs as they appeared on the compilation album '1'. This album includes the original (but digitally re-mastered) recordings of the 27 greatest UK and USA Beatles' hit singles. The songs were selected to obtain a variety in time period and status on their No. 1 chart position[2].

It was assumed that songs of the 'The Beatles' can be reliably used in an experiment by simply referring to them by their song titles. Although other artists have performed these songs as well, the compositions themselves can be easily attributed to 'The Beatles' and are known under their original Beatles version (i.e., in a given standard key, tempo and interpretation).

2.2 Participants

Eighteen persons (10 males, 8 females) participated in the experiment. The average age of the participants was 28 years (min: 18, max: 41). Eight of them were students 'Musical theatre' (4) or 'Classical voice' (4) at the Brabant Conservatory in Tilburg, the Netherlands. All students had received at least five years of formal singing training (including their education at the Conservatory). The students obtained a gift token for their participation. The other ten participants never had any singing training or education. They were colleagues at the research laboratory and participated voluntarily.

Most participants were 'coincidental' listeners to the music of 'The Beatles' (e.g., while listening to the radio). Only two

[1] Singing ability refers to the use of theory, vocal skills and aural skills as employed by professional singers. The current study defines 'singing ability' as being formally trained in singing.

[2] The Beatles, 1, EMI Records Ltd, 2000. The following 12 songs were selected: 'Love me do', 'Can't buy me love', 'A hard day's night', 'I feel fine', 'Help!', 'Yesterday', 'We can work it out', 'Yellow submarine', 'Penny Lane', 'Hey Jude', 'Come together' and 'Let it be'.

participants listened sometimes to a 'The Beatles' CD album. All students responded that they had never sung music of 'The Beatles' on an educational or (semi-) professional basis. No one owned the album '1', used in the experiment.

2.3 Procedure

The eight students were invited in a classroom at the Brabant Conservatory. The other ten participants were invited in an office room at the laboratory.

First, participants were asked to complete a small questionnaire for gathering some personal attributes (e.g., name, gender, age), their musical training background and their general familiarity with 'The Beatles'.

Then, participants were asked to perform a card-sorting task in which they had to sort twelve Beatles song titles displayed on small cardboard cards. For that, they read a paper form containing brief instructions on the task. After they had explained the instructions in their own words to check their understanding of the task, they received all twelve cards at random and one-by-one. They were instructed to try to recall the song displayed on the card and to sing a few notes in order to verify that the correct song was recollected. If participants were totally unfamiliar with the song, they were asked to put the card aside. If they happened to know the song, they were instructed to sort all cards list-wise in descending order of familiarity; the most familiar song should appear at the top of the list and the least familiar song at the bottom. No ties were allowed. They were allowed to shift the cards until they were satisfied with the arrangement.

After the sorting task, participants received another paper form containing the instructions of the experimental task and were again asked to explain the instruction in their own words. The whole procedure was practiced once by using 'The Beatles' song 'All you need is love', before they did the experimental test. The procedure was as follows. First, participants received a card displaying a song title. The two top-most song titles and the two bottom song titles of the personally arranged song list were provided to them at random. They were instructed to imagine the vocal melody of the song given, to forget about the lyrics, and to focus on the original performance by 'The Beatles'. They did not need to recollect the whole vocal melody but were asked to restrict themselves to a passage, which they could easily remember and sing. Then, they were instructed to sing the same passage *twice* in any preferred style of singing (not whistling) without using the song lyrics. No restrictions on the length, tempo, key or content of the passage were given. No instructions were provided how to sing (e.g., breath control, enunciation, volume, body posture). Then, they listened to the original 'Beatles' recording of the song on CD using a portable player with headphones. They could stop listening at any moment in time, but were not allowed to repeat or skip passages. They were already instructed to sing the same passage for the third time, immediately after stopping listening. All singings were recorded for later analyses.

2.4 Processing of the sung melodies

Note onsets, note offsets and beat markers were positioned manually by listening and using a visual representation of the audio waveform of the passage sung, its pitch contour and its wide-band spectrogram. Beat markers demarcate the onset of the human sense of a beat (e.g., foot tapping); the average distance between succeeding beats (inter-beat period) is inversely proportional to the tempo in which the singing is done. In general, an inter-beat period equals the duration of a quarter note. For tempo measurement, also the audio of the original Beatles recordings was processed. The pitch of the sung melodies was measured using sub-harmonic summation in time frames of 40 milliseconds (Hermes, 1988). The pitch, the note onsets and offsets, and the tempo were combined to come at a music score transcription of the sung melody. The median pitch frequency between a note onset and offset was used to quantize the musical pitch value for each note using the equally tempered musical scale tuned at A-440. The resulting transcription was encoded in a MIDI representation for listening and visual inspection in music notation software.

For all sung melodies, the music transcription of the corresponding vocal melody line on the original recordings of 'The Beatles' was collected. Songbooks were used as a source; transcriptions with *exact* timing were checked on their correctness for being used as a reference.

2.5 Measures

Four different measures were defined to assess the sung melodies. Measures were calculated using the original Beatles melody as a reference (i.e., the actual melody).

The measure *tuning* was assessed by the semitone deviation of the first tone as produced by the participant from the corresponding, correct pitch as found on the original Beatles recording. This deviation was corrected for octave.

The measure *contour* was defined as the proportion of correctly sung pitch movements. As shown in Figure 1, a contour representation of a melody is a sequence of three different pitch movements: *down*, *same* and *up*. So, if one sings a melody by going up six times and going down twice and five of these pitch movements happen to be at the correct position, the *contour* measure states: 0.625 (5/8).

Quite similar to the *contour* measure, the *interval* measure was defined as the proportion of correctly sung intervals. As shown in Figure 1, musical intervals are the semitone distances between two succeeding notes in a melody. This measure was corrected for the fact that the change of one note affects two corresponding intervals.

To compute the *contour* and *interval* measures, an optimal alignment in the number of matches between elements has to be established between the sung melody and the actual melody. This was done by approximate string matching in a dynamic programming framework (see, e.g., (Wagner and Fisher, 1974)).

The measure *tempo* was expressed in beats per minute (bpm) and computed using the average inter-beat period measured.

All alignments needed for the measurements were checked and corrected, if necessary. For the analyses, all proportional data (i.e., *contour* and *interval*) were *logit*-transformed.

2.6 Results

All reported multiple analyses of variance (MANOVA) with repeated measures were conducted with the within-subject variables *song familiarity* (familiar vs. less familiar), *song* (2) and *trial* (3), unless stated otherwise. Between-subject variable was *singing training* (trained vs. untrained).

Recall that participants had to sing two familiar and two less familiar songs and that, on trial 1 and 2, they had to do this from memory without an external reference. Trial 2 was meant to assess any change in singing by solely reflecting on the performance on trial 1; the experiment did not reveal any effects of this sort. Just before trial 3, they had listened to the original song on CD.

At one particular occasion, one participant was 'totally lost' in the music while listening to it and found himself not able to reproduce a melody on the third trial. For the *contour*, *interval, tempo* and *timing* measures, this missing data was replaced by the mean value of the first two trials. For the *tuning* measure, it was replaced by the maximum semitone deviation (6 st.).

2.6.1 Tuning

In Figure 2, participants' deviation from the correct pitch of the first tone is shown for all three trials[3]. If participants could not remember, produce or approach the correct pitch at all, it would be reasonably to expect that the deviations were evenly distributed at each side of the correct pitch. Therefore, a Kolmogorov-Smirnov one-sample test was performed to compare the observed deviations from the semitone scale with a uniform distribution of deviations. For trials 1 and 2, we could not reject the null hypothesis for both trained and untrained singers that the deviations are thought to have been drawn from a uniform distribution (on trial 1, $Z = 1.06$, $p = 0.21$ (trained), $Z = 0.69$, $p = 0.73$ (untrained); on trial 2, $Z = 0.74$, $p = 0.65$ (trained), $Z = 0.87$, $p = 0.48$ (untrained)). For trial 3, the hypothesis for uniformity was rejected for the trained singers, not for the untrained ones ($Z = 1.71$, $p < 0.01$ (trained), $Z = 0.88$, $p = 0.43$ (untrained)).

A Cochran Q test was conducted to assess whether or not the starting tone of a particular song changed in having the correct pitch across trials. For the trained singers, song reproductions changed significantly in having correct tones on trials 1, 2 and 3 ($Q = 10.53$, $p < 0.01$). On trial 1 and 2, the trained singers made no error on 5 out of a total of 32 reproductions (16%). On trial 3, they made no error at 15 out of a total of 32 reproductions (47%); 21 reproductions (66%) were within 1 semitone. Also for the untrained singers, song reproductions changed significantly in having the correct pitch under the three trials ($Q = 7.17$, $p < 0.05$). On trials 1 and 2, the untrained singers made no error for 3 (8%) and 2 (5%) out of a total of 40 reproductions, respectively. On trial 3, they tuned accurately at 9 out of 40 reproductions (23%) and 16 reproductions (40%) were within 1 semitone.

A Fisher exact test revealed that the number of song reproductions starting at the correct pitch on trial 3 was significantly higher for trained singers than for untrained singers ($p = 0.027$). Also when allowing a pitch deviation of 1 semitone, there was a significant difference between reproductions of trained and untrained singers ($p = 0.009$).

[3] For the first three successively produced tones, it was found that semitone deviation did not depend on tone position and the average deviation did not differ significantly from the individual deviations. Consequently, analysis was simplified by providing only analyses of the participant's first produced tones.

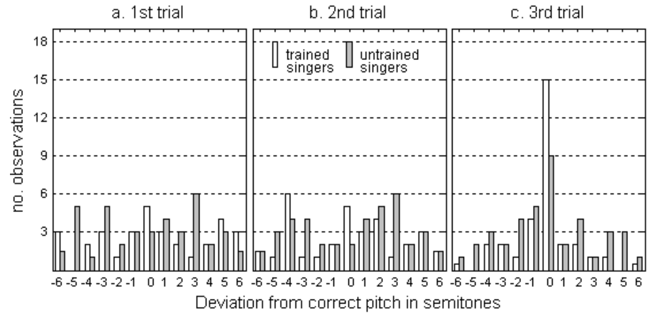

Figure 2: Tuning. Observed deviation from the correct pitch in semitones of trained and untrained singers singing their first note of a Beatles melody on three trials. The deviation was corrected for octave. Note that a deviation of 6 semitones yields the same pitch class as a deviation of –6 semitones. Observations were equally distributed amongst these two categories.

In summary, both trained and untrained singers attempted to adopt the correct pitch after having heard the original recording (i.e., on trial 3); the trained singers were better at doing this.

Also, participants' tuning performance for familiar and less familiar songs has been analyzed. For the familiar tunes, Kolmogorov-Smirnov one-sample tests found out that the observed deviations differed from being drawn from a uniform distribution for all three trials (For trial 1, $Z = 1.41$, $p < 0.05$; for trial 2, $Z = 1.70$, $p < 0.01$; for trial 3, $Z = 1.56$, $p < 0.05$). However, only on trial 3, the data peaks at zero deviation (matching the correct pitch). On trial 1 and 2, peaks occurred at other places. For the less familiar tunes, observed deviations were likely to be drawn from a uniform distribution (for trial 1, $Z = 1.35$, $p = 0.053$; for trial 2, $Z = 1.18$, $p = 0.12$; for trial 3, $Z = 1.00$, $p = 0.27$).

A Cochran Q test found that the song reproductions differed significantly in having the correct pitch on trials 1, 2 and 3 ($Q = 19.11$, $p < 0.001$). On trial 1 and 2, familiar songs were with no error in 3 (8%) and 1 (3%) cases out of a total of 36 cases, respectively. On trial 3, 15 out of 36 (42%) familiar songs were started at the correct pitch; 23 (64%) were within 1 semitone. For the less familiar songs, the number of song reproductions did not differ in having the correct pitch on trials 1, 2 and 3 ($Q = 2.00$, $p = 0.37$). On trial 1 and 2, less familiar songs were started without error in 5 (14%) and 6 (17%) out of a total of 36 cases, respectively. On trial 3, 9 (25%) less familiar songs were without error; 14 (39%) were within 1 semitone.

A Fisher exact test found no significant difference in the number of correctly pitched tones between familiar and less familiar songs on trial 3 ($p = 0.11$). For pitch errors within 1 semitone, there was a significant difference between familiar and less familiar songs on trial 3 ($p = 0.029$). Only when considering an allowance of 1 semitone deviation, more reproductions of familiar songs (23 out of 36) were started at the correct pitch than the reproductions of less familiar songs (14 out of 36).

2.6.2 Contour

The results of the proportion of correctly sung pitch movements for familiar and less familiar songs across trials are shown in Figure 3.

In the multivariate analysis of the *logit*-transformed *contour* data, a main effect for *trial* was found to be not significant ($F(2,15) = 3.40$, $p = 0.061$), while the univariate test found a significant main effect for *trial* ($F(2,32) = 3.72$, $p < 0.05$). Further analysis found out that most variance was explained by the difference in contours produced on trial 3 and contours produced on trial 1 and 2 combined ($F(1,16) = 3.89$, $p = 0.066$). On trial 3, contours tended to contain proportionally more correct pitch movements than on trial 1 and 2 (mean proportions: 0.79, 0.78, 0.81 on trial 1, 2, and 3, respectively).

An interaction effect for *song familiarity* and *trial* was found to be significant ($F(2,15) = 4.34$, $p < 0.05$). The sung contours of less familiar songs contained proportionally more correct pitch movements on trial 3 as compared to trial 1 and 2 than the contours of familiar songs ($F(1,16) = 7.00$, $p < 0.05$). As shown in Figure 3, the reproduction of contours of less familiar songs improved on trial 3. In contrast, the contour reproduction of familiar songs did not improve (mean *contour* for familiar songs across trials: 0.82, 0.82, 0.81; mean *contour* for less familiar songs across trials: 0.77, 0.74, 0.82).

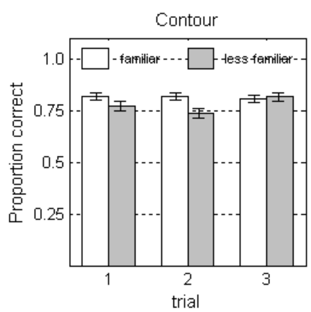

Figure 3: Contour. Mean proportion of correctly sung pitch movements of familiar and less familiar songs of 'The Beatles' across three trials. The crossbars represent standard error of the mean.

No other effects were found to be significant; for instance, contour reproduction was not affected by *singing training* ($F < 1$, $p = 0.47$) or *song familiarity* ($F < 1$, $p = 0.81$).

2.6.3 Intervals

The results of the proportion of correctly sung intervals for familiar and less familiar songs and by trained and untrained singers across trials are shown in Figure 4.

In the analysis of the *logit*-transformed *interval* data, a main effect for *singing training* was found ($F(1,16) = 6.00$, $p < 0.05$). Trained singers sang 62% of their intervals correctly, whereas untrained singers sang 56% of their intervals correctly. As shown in panel (b) of Figure 4, trained singers sang better than untrained singers did in terms of musical intervals.

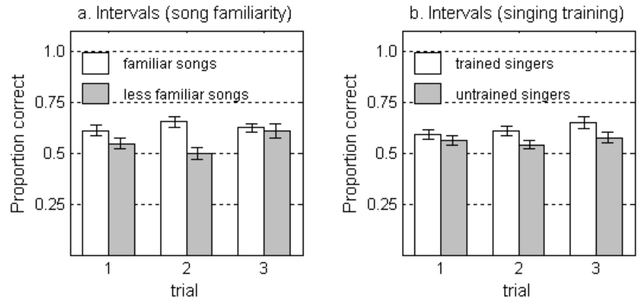

Figure 4: Interval. Mean proportion of correctly sung musical intervals. The left-hand panel (a) shows the mean proportions for familiar and less familiar songs of 'The Beatles' across three trials. The right-hand panel (b) shows the mean proportions for trained and untrained singers. The crossbars represent standard error of the mean.

Though an interaction effect for *song familiarity* and *trial* was not found to be significant in the multivariate analysis ($F(2,15) = 2.15$, $p = 0.15$), it was found to be significant in the univariate test ($F(2,32) = 3.79$, $p < 0.05$). Further analysis revealed that most variation was explained by the reproductions of less familiar songs being more accurate on trial 3 than on trials 1 and 2, while this did not hold for the familiar songs ($F(1,16) = 4.01$, $p = 0.062$). As shown in panel (a) of Figure 4, less familiar songs were better sung at the interval level on trial 3 (mean across trials: 0.55, 0.50, 0.61). In contrast, the singing of the familiar songs did not improve (mean across trials: 0.61, 0.65, 0.62).

Analyzing the *interval* data produced on trials 1 and 2, a main effect of *song familiarity* was found to be significant ($F(1,16) = 6.81$, $p < 0.05$). As shown in panel (a) of Figure 4, the singing of familiar songs contained proportionally more correct intervals on trials 1 and 2 (mean: 0.63) than the singing of less familiar songs did (mean: 0.52). The singing of the less familiar songs improved on trial 3 (mean: 0.62).

No other effects were found to be significant.

2.6.4 Tempo

The measured tempo as produced by the participants is compared to the actual tempo (from the original Beatles recordings) in bivariate scatter plots in Figure 5. In case of singing a familiar song from memory, participants came close to the original tempi indicated by the high correlation between the sung tempo and the actual tempo ($r = 0.91$, $p < 0.01$; trial 1 and 2 combined). They picked up the actual tempo after having heard the familiar song ($r = 0.97$, $p < 0.01$; trial 3). In case of singing a less familiar song from memory, the tempi sung are more dispersed from the actual tempi ($r = 0.81$, $p < 0.01$; trial 1 and 2 combined). However, song listening made participants pick up the actual tempo of the less familiar song ($r = 0.92$, $p < 0.01$; trial 3).

Tempo matching performance did not differ among *singing training*. Trained and untrained singers performed equally well (on trials 1 and 2, $r = 0.89$, $p < 0.01$; on trial 3, $r = 0.95$, $p < 0.01$).

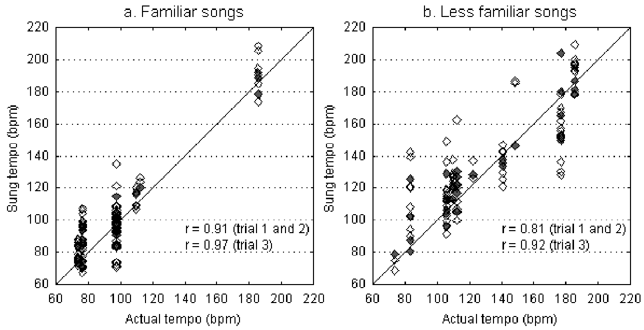

Figure 5: Actual tempo versus sung tempo in beats per minute. The left-hand panel (a) shows the tempo deviations for the familiar songs. The right-hand panel (b) shows the tempo deviations for the less familiar songs. The white diamonds represent data on trial 1 and 2 combined. The black diamonds represent data on trial 3.

To put current results into context, it is important to know that humans have only a finite resolution in perceiving a difference between sensations caused by physical stimuli. The smallest difference in sensation that can be noticed is called a 'just noticeable difference' (JND). Though different JNDs for tempo perception have been reported (caused by different tasks and instructions), we use a rather conservative JND of +/-6% (Allen, 1975). This means that tempo deviations within this 6% range, while singing a song from memory or after recent exposure, went largely unnoticed by the participants. In Figure 6, the deviation from the actual tempo expressed as percentages are shown in equal steps of 6%.

It appeared that, when singing a familiar song from memory on trails 1 and 2, 23/72 (32%) of the singing fell within one JND of the actual tempo and 46/72 (64%) came within two JNDs. On trial 3, 17/36 (47%) came within one JND and 27/36 (75%) came within two JNDs.

Likewise, when singing a less familiar song from memory on trials 1 and 2, 20/72 (28%) of the singing fell within one JND and 41/72 (57%) within two JNDs. On trial 3, 18/36 (50%) and 26/36 (72%) came within one and two JNDs, respectively.

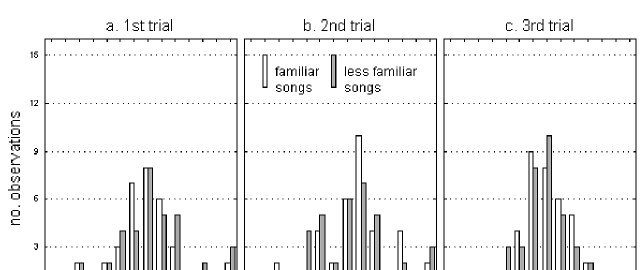

Figure 6: Percent deviation of tempo sung from actual tempo of familiar and less familiar songs of 'the Beatles' across three trials. The percent deviation is divided up into steps of 6%.

Good performance may be biased by a strong memory for a small range of tempi that happened to be in the test. It is argued that this was not the case because of the following. In total, participants sang eleven different Beatles songs with tempi ranging approximately between 70 and 190 beats per minute; this range spans from 'moderately slow' to 'very fast'. As shown in panel (a) of Figure 5, six different familiar songs were sung that were not characterized by a fast tempo; five out of the six songs were not faster than 120 beats per minute. As shown in panel (b) of Figure 5, the material of less familiar songs was made out of ten different pieces; the tempi were widely distributed.

Good performance may also be biased by a 'personally preferred' tempo that is easy to remember. Though a 'personally preferred' tempo varies in different contexts, it often points between 80 and 100 beats per minute (Dowling and Harwood, 1986). We found three observations that contradict the existence of this bias. First, the familiar song 'Yesterday' was reproduced in a wide range between 71 and 135 beats per minute, though its original tempo is a moderate one of 97 beats per minute. Second, each participant sang a wide variety of tempi and hence did not rely on a 'personally preferred' tempo. The tempi sung for the four songs on the first trial by a single participant did not correlate significantly. The mean difference between the slowest produced tempo and the fastest produced tempo sung by a participant on the first trial was 79 beats per minute (min: 19 bpm, max: 133 bpm). Third, participants had not sung songs with similar original tempi. The actual tempi of the four songs sung in a trial did not correlate significantly. The mean difference of the actual tempi between the slowest song and the fastest song was 78 beats per minute (min: 33 bpm, max: 112 bpm).

2.7 Discussion and conclusion

This experiment evaluated the effects of song familiarity, singing training and recent song exposure on the singing performance of 'The Beatles' melodies. Participants were not instructed how to sing or how to improve their singing; so to the extent they performed well, they did so on their own. It became clear that some perceptual properties of a melody were more accurately reproduced from memory than others. Improvement after song listening was not found for all properties and differed among song familiarity and singing training.

It was expected that singing at the correct pitch and at the correct tempo is quite easy. As found by Levitin (1994), Levitin and Cook (1997), there is some evidence that people have a sense of absolute memory for pitch and tempo for their *favourite* pop-rock songs.

In the current experiment, no support for absolute memory of pitch for familiar and less familiar songs was found. It was found that trained and untrained singers tend to approach the correct pitch from the original Beatles recording, but only after having heard the recording. Trained singers were significantly better in tuning their singing to the song on the CD; 47% of the reproductions of trained singers were at the correct pitch whereas it was 23% for the untrained singers.

In the current experiment, support for absolute memory of tempo for familiar and less familiar songs was indeed found, irrespective of singing ability. Though the tempi of less familiar melodies were somewhat less accurately matched with the originals, almost one out of three reproductions came within one JND for tempo; almost two out of three came within two JNDs. This even improved after CD listening.

It is clear that both trained and untrained singers are aware that a song has a particular correct pitch and a particular correct tempo. Start singing at the correct pitch is however likely to be hampered by a lack of singing experience. It is known that people have a preferred octave in which they sing best and that trained singers are better able to use their full vocal range. If the original recording of the song has a pitch that happens to be out of this vocal range, people have to transpose their singing one whole octave to match this pitch. This octave switching requires training.

In addition, the pitch of familiar melodies was more easily adopted than that of less familiar ones. Since less familiar melodies were reproduced worse on other aspects as well, matching the correct pitch for these melodies was probably given a lower priority.

It was expected that familiar melodies are sung more accurately than less familiar ones at the interval level. Indeed, the memory reproduction of less familiar melodies contained proportionally fewer correct intervals. However, both trained and untrained singers improved the singing of intervals after having heard the original song. This finding confirms the inadequate encoding of interval information for less familiar melodies.

It was expected that trained singers sing proportionally more correct intervals than untrained singers do, in contrary to the reproduction of contours. Indeed, trained singers sang 62% of the intervals correctly, while untrained singers sang 56% of the intervals correctly. The contour reproduction was not affected by singing training; about 80% of the pitch movements were correctly reproduced. This finding confirms the dominance of contour in the remembering and, in turn, the singing of melodies for both trained and untrained singers. The bad performance on singing correct intervals by untrained singers may be due to the intervals of melodies being badly encoded in memory or to a lack of essential singing skills.

It was expected that trained singers improve their singing after having listened to the original song, in contrast to untrained singers. As found by Davidson (1994), untrained singers are less able to reflect on and improve their own singing performance. This hypothesis has to be rejected, however, considering current results. Both trained and untrained singers improved their singing of less familiar songs on many aspects, after having listened to the CD recording. To the extent that untrained singers truly reflected on their own performance or that they simply took over the melody in the original song needs further scrutiny.

The current study did not address the beauty or professionalism of a singing performance or the special liking or willingness to sing; it only addressed the accuracy in which one sings. With respect to the aesthetics of singing, trained singers are assumed to have a better control of timing, breath, vibrato, enunciation and a more pronounced quality of their voice, amongst other things. The love for singing is best demonstrated by the observation that trained singers sang a far longer passage of a familiar song than untrained singers did[4]. Some singers were even more than happy to sing the complete song three times.

The current study did not assess the effects of song complexity or music idiomatic differences (e.g., genre, stylistic period, instrumentation). The Beatles' songs used in the current study are simple; none of the participants made any comment on the songs as being too complex to sing. It may be evident that some songs in the world are more difficult to remember and to reproduce than others due to melodic, metrical and rhythmical complexities.

No reliable statistics was obtained on what parts of a melody are most likely to get recalled and sung. This experiment showed that lyrics and melody are considerably tied (Serafine, Davidson, Crowder and Repp, 1986); participants were cued by the song title on the card to recall the passage to sing. Consequently, the melody of almost all 72 reproductions, except for four 'Let it be' reproductions, had the song title as one of the very first words in its lyrics.

3 Implications for 'query by humming'

Current findings impose constraints on search algorithms for 'query by humming' given the song unfamiliarity and the diversity of singing abilities and training amongst the general public. Nowadays, most search algorithms are best described as optimal solvers of a precisely stated mathematical problem in a discrete domain. The general problem is stated as finding an optimal alignment between the pitches and the note durations of the sung melody and the melodies in the database. While computing this alignment, typical human errors in singing are taken into account[5].

First, as people are unlikely to start singing at the correct pitch, the use of a transposition invariant melody representation is a prerequisite for search efficiency. The use of a contour or interval representation is already common practice.

Second, the extent to which a person is familiar with a song and the extent to which a person is a good or a trained singer are crucial to how well a melody will be reproduced. A contour is a dominant feature that aids in memorizing a melody and in reproducing it directly from memory, irrespective of how familiar one is with the song or how good one is in singing. Intervals of a melody are more readily available to trained singers for reproduction and if songs are more familiar. Though a contour only provides limited information about a melody, searching for the pitches should be based on a weighted combination of contours and intervals dependent on how well a person can sing and how familiar a person is with the song of the sung melody.

Third, people have a good memory for global tempo. However, trained singers are deemed to explicitly use the metrical structure of the song in their singing resulting in a better use of timing and smoother local time fluctuations. The extent to which the errors in note durations (or their durational

---

[4] Trained singers sang, on average, 45.4 notes of a familiar song, whereas untrained singers sang, on average, 27.7 notes. This difference was found to be significant ($F(1,16) = 16.16$, $p < 0.005$).

[5] In addition to singing errors, the front-end of any 'query by humming' application that transcribes the singing into a symbolic query introduces errors in the search process. The implications of these errors are not further discussed.

ratios) contribute to the overall search should be based on the rhythmic or metrical complexity of the song at hand, how well a person can sing and the measured global tempo, since the latter partly determines the timing employed.

Last, besides practice, mere listening to the original song makes people improve their singing. This calls for adequate visual and auditory feedback mechanisms on the performance of people singing, to enable them to improve or change their singing.

Acknowledgement

Thanks go to the Brabant Conservatory in Tilburg, the Netherlands, for their co-operation.

4 References

Allen, G.D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics, 3*, 75-86.

Boltz, M. & Jones, M. R. (1986). Does rule inversion make melodies easier to reproduce? If not, what does? *Cognitive psychology, 8*, 389-431.

Croonen, W.L.M. (1994). Effects of length, tonal structure and contour in the recognition of tone series*, Perception & Psychophysics, 55*, 597-606.

Davidson, L. (1994). Song singing by young and old, In Aiello, R., and Sloboda, J.A. (Eds.), *Music Perceptions*. (pp. 99-130). New York: Oxford University Press,.

Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*, 351-354.

Dowling, W.J., & Harwood, D.L. (1986). *Music cognition*, London: Academic Press.

Edworthy, J. (1985). Melodic contour and musical structure. In Howell, P., Cross, I., and West, R. (Eds.), *Musical structure and cognition*. (pp. 169-188). New York: Academic Press.

Hermes, D. (1988). Measurement of pitch by subharmonic summation, *Journal of the Acoustical Society of America, 83, 1*, 257-264.

Jones, M.R., Summerell, L. & Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology, 39*, 89-121.

Levitin, D.J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies, *Perception & Psychophysics, 56, 4*, 414-423.

Levitin, D.J. & Cook, P.R. (1997). Memory for musical tempo: Additional evidence that auditory memory is absolute*, Perception & Psychophysics, 1996, 58,* 927-935.

Serafine, M.L., Davidson, J., Crowder, R.G. & Repp, B.H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language, 25*, 123-135.

Wagner, R.A. & Fisher, M.J. (1974). The string-to-string correction problem, *Journal of the Association of Computing Machinery, 21, 1*, 168-173.