Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, 2002
C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.). Adelaide: Causal Productions.

ABSOLUTE MEMORY FOR MELODY: A STUDY WITH CHOIR SINGERS

Kathrin Hahn

Department of Psychology, Humboldt University Berlin, Germany

ABSTRACT

A number of recent recognition and production studies using classical music, pop songs, and folk songs suggest that long-term memory retains more than only relational information on stored melodies: Non-musicians and non-AP-possessors seem to be able to recall absolute pitches and absolute tempo. Here, choir singers were tested on memory for the key of well-rehearsed pieces. Two groups of participants were asked to reproduce choir-pieces from memory 5 and 10 months after the concert, respectively. Results support the notion of accurate starting pitches being stored in long-term memory.

1. INTRODUCTION

Musical practice (e.g., the practice of publishing art songs in more than one key so that singers with different voice ranges can use them) as well as studies using a recognition paradigm (e.g., Dowling & Bartlett, 1981) show that participants usually don't differentiate between exact transpositions of a melody and the original melody, whereas they do notice changes in intervals or contour of a familiar melody. Thus, it has been proposed (e.g., by Carroll-Phelan & Hampson, 1996) that the relational features of a melody, i.e. intervals and rhythm, are stored in long-term memory whereas their key and absolute tempo are not represented unless someone has absolute pitch (AP), i.e. the ability to name or produce tonal pitches without reference to an external standard. However, there are also studies showing that participants *do* seem to remember absolute pitches as well as absolute tempo of familiar melodies. This has been shown by Halpern (1988, 1989) for songs, which are passed on by aural traditions (like "Happy Birthday") and by Levitin (1994; Levitin & Cook, 1996) for pop songs that we hear in the radio or on CD, both using a production paradigm. Vitouch & Gaugusch (2000), as well as Terhardt & Ward (1982) and Terhardt & Seewann (1983) obtained similar results for absolute key recognition. Therefore, musical long-term memory seems to be able to encode relative as well as absolute features, and the features that are recalled in a memory experiment could be task-dependent.

2. AIMS

The aim of this study is to replicate Levitin's (1994) findings for choir music: It examines whether choir singers can recall the key of well-known choir pieces in a production task. The music that is examined here is sung from a score, and therefore has only one correct key, like most pop songs. It is not deliberately learned (no memory techniques are used). The differences from pop music are that choir music is learned primarily by singing, not by hearing a CD, and that the demands on memory may be more sophisticated because while singing the own voice, choir singers also hear the melodies of 3-4 other voices. Another difference is that the choir pieces of a specific concert are heard very often during rehearsal period, but when the concert is over they are rarely heard again, whereas pop songs are more continuously listened to.

3. HYPOTHESIS

The hypothesis is that choir music, which has been sung in the same key very many times, is represented in long-term memory in a way that relative as well as absolute features can be recalled. Thus I expected the choir singers to sing a well-known piece in the key in which it has been practiced. As tempo is less stable during rehearsals and concerts, it is not studied here.

4. METHOD

4.1. Participants

The participants were 26 choir singers from two choir projects. The two projects differed in the amount of time between concert and experiment as well as in the difficulty of the music and in the rehearsal mode. Five participants had taken part in both projects and were therefore tested twice in this study. The two groups of participants are characterized as follows:

The group BARDOU consisted of 13 singers who ranged in age from 18-32 (m = 24.5). On average they had 11.5 years of choir experience, 4 years of ear training, 3 years of voice lessons and 13 years of instrumental lessons. Nine participants were music students or musicians while four participants were non-musicians. Two participants claimed to have AP (but they were not tested on AP). In this group, there were 10 months between concert and experiment. The rehearsed music was sacred music from the 16th to the 20th century (Mundy, Schütz, Zelenka, Brahms, Poulenc, Messiaen), which was in large part sophisticated polyphonic music with an average duration of 2-5 minutes per piece. Rehearsal mode was two weeks of intensive every-day rehearsals.

The group DIDO & AENEAS consisted of 18 singers ranging in age from 18-58 (m = 29.5). On average they had 13.5 years of choir experience, 3.5 years of ear training, 3 years of voice lessons and 14 years of instrumental lessons. Six participants were music students or musicians while 12 participants were non-musicians. Three participants claimed to have AP. In this group there were 5 months between concert and experiment. The rehearsed music was the opera "Dido and Aeneas" by Henry Purcell, which is less difficult than the BARDOU-program, having mostly short (.5-2 min. per piece) homophonic choir numbers. Rehearsal mode was a few weekends in a row for about a month. Some of the pieces were sung by heart, while all of the BARDOU-program was sung with the score.

T-tests and chi^2-tests were performed to find group differences in the amount of musical training or incidence of AP. There were no significant differences except for the BARDOU group containing more people who were formally considered as musicians (chi^2 = 3.94, $p < .05$). But since all participants had an amount of musical experience that by far exceeds regular school education they should all be considered as musicians and the two participant groups are comparable in this respect.

4.2. Procedure

The participants were tested individually. They were given the texts (not the music) of all the pieces of the project in which they had taken part. They were asked to select a piece they could remember very well and to recall this piece to their inner ear. Then they were asked to reproduce their image of the piece by singing a passage of self-chosen length. When finished singing, the participants were given a short questionnaire: They were asked how they would rate their performance in the experiment, whether they had known the piece before rehearsal time, whether they had sung or heard it again between concert and experiment, and how well they liked the piece. They were also asked to name the original starting pitch of the piece and the pitch on which they had just started. The last question was aimed at finding out how well persons without AP could name sung pitches. After these questions the participants were asked about their musical background before starting the second trial of the experiment. In the second trial the participants were not allowed to choose a piece but were given the name of a piece to be imagined. Other than that the procedure was the same as in the first trial, including the questionnaire.

The five singers who were tested twice (as members of both participant groups) were involved in a conversation for about 10 minutes before starting the third and fourth trial, in order to remove the just sung pitches from their short-term memory. The sung melodies were recorded with a minidisk recorder (Kenwood DMC-J7R) connected to a mono microphone (Sony F-V310). A digital recording technique was chosen in order to avoid frequency fluctuations that sometimes occur in analogue recordings.

4.3. Data Analysis

The starting pitches of the sung melodies were analyzed by two choir conductors in two independent sessions who used a tuning fork to name the pitches they heard via headphones (Sennheiser HD 200). They agreed in 57 of the 62 trials, leading to a Cohen's kappa of $\kappa = .91$. The five trials they did not agree on were then fed into a PC via a 16-bit sound card and analyzed with a fast Fourier transformation (using the software Cool Edit), which gave out the exact frequency in Hertz as well as the name of the pitch. After all the starting pitches had been named they were compared with the correct starting pitches as written in the scores. Errors were measured in semitone deviations from the correct pitch. Octave errors were not penalized so that all deviations were in a range of - 6 to + 5 semitones.

5. RESULTS

All participants could remember at least some of the pieces and what they sang was always recognizable as to what part of which piece it was. Typical mistakes were little rhythmic variations and skips of phrases or repeats but only occasionally changes in the intervals. For most of the participants text seemed to be a good cue to recall the melody.

For the comparison of the sung starting pitches with the score, there were 16 cases where it wasn't clear with which score pitch the sung pitch must be compared: In 7 cases of the BARDOU-group the participants sang a fugue entry that had originally been sung by a different voice (e.g. a soprano singing the part of the tenor because the tenor had the first entry in the piece). In 9 further cases of the DIDO-group non-sopranos sang the soprano part (which was usually the melody). For both groups, the sung starting pitches were compared first with the notated pitch of the *own* voice, then with that of the *not-own* voice. Since there was no overall clear pattern as to which notated pitches (*own* or *not-own* voice) the sung pitches were closer to, the data of these 16 cases were not used for further analyses. Two other cases were also excluded because of disruptions during the experiment, so there remained 44 cases, where the deviations between sung pitch and score pitch could unambiguously be obtained:

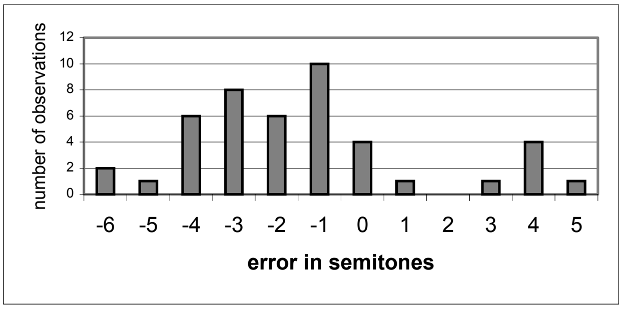

Figure 1: Distribution of errors without octave errors, N = 44

The null hypothesis is that participants can't remember absolute pitches at all. If that were true, the distribution of errors (Fig. 1) should approximate a uniform distribution with 3.6 cases in each of the 12 categories. This is not the case, as a chi²-test ($\chi^2 = 31.27$, $p < .001$) and a Rayleigh test for circular data have shown. Instead, the data fit a normal distribution ($R = .45$, $p < .001$). Looking at the two trials separately however (Figs. 2 and 3), one can see that the null hypothesis can only be rejected for the first trial ($\chi^2 = 24.14$, $p < .05$), but not for the second ($\chi^2 = 13.0$, $p > .05$). This suggests that the choice for the key of the second piece was more or less random.

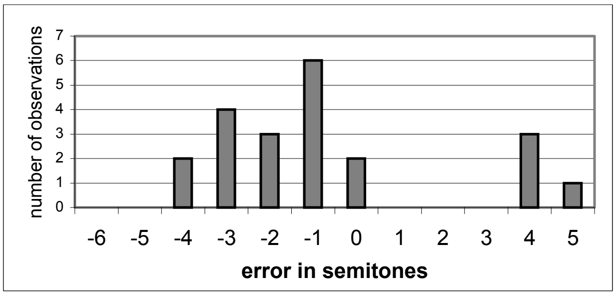

Figure 2: Distribution of errors, 1st trial of experiment, N = 21

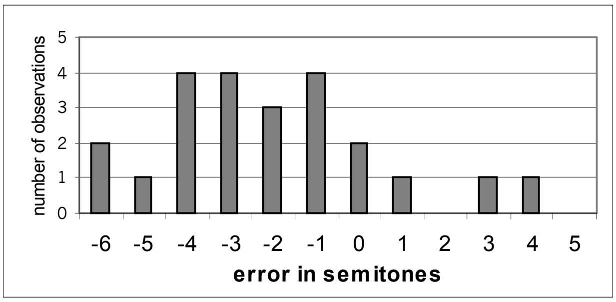

Figure 3: Distribution of errors, 2nd trial of experiment, N = 23

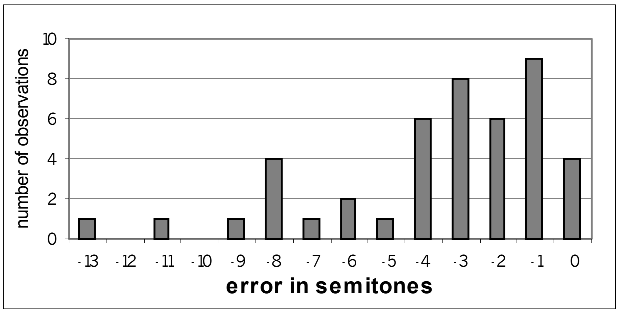

Figure 4: Distribution of errors including octave errors, N=44

When octave errors are included (Fig. 4) it becomes obvious that *all* the sung pitches that weren't right on, were below the notated pitches. Also, the mean of the distribution in Fig. 1 isn't 0 as in Levitin's study, but between -1 and -2.

To find out whether there were "over-represented" starting pitches in the musical literature tested here as well as in the participants' productions, the distribution of notated starting pitches („targets": Fig. 5) as well as the distribution of sung starting pitches (whatever their distance to the targets: Fig. 6) and the distribution of intervals between the starting pitches of first and second trial of every single participant (Fig. 7) were analyzed. If the pieces started on only a restricted number of pitches *and* if participants produced only a restricted number of pitches, then there were two possible interpretations: Either participants have the correct starting pitches stored in long-term memory (absolute memory), or they have formed a more generalized representation of often occurring "standard pitches" which are not connected with any particular melody. The latter interpretation could be supported if participants started on the same pitch in the two trials.

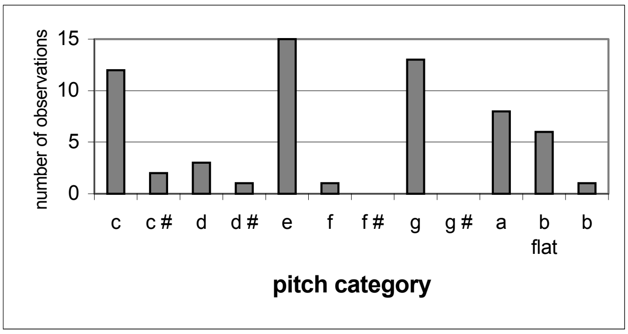

Figure 5: Distribution of target pitches for all 62 cases

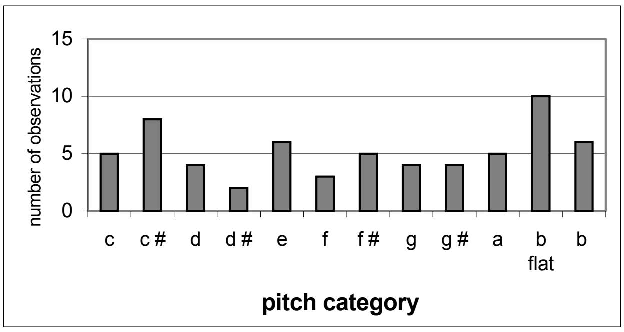

Figure 6: Distribution of sung starting pitches for all 62 cases

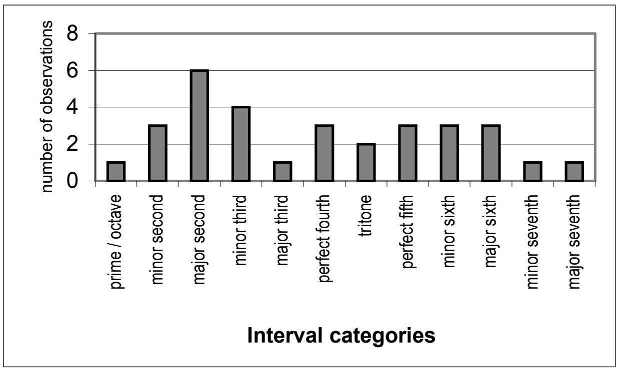

Figure 7: Distribution of intervals between starting pitches of 1st and 2nd trial of every participant, all 62 cases

Results show that while the null hypothesis of uniformity had to be rejected for the target notes ($chi^2 = 64.58$, $p < .001$), it was not rejected for the sung starting pitches ($chi^2 = 10.0$, $p > .05$) as well as for the intervals between the pitch of the first and second trial of every single participant ($chi^2 = 9.65$, $p > .05$). This means that, although there is only a restricted number of often occurring targets, the participants overall didn't sing only a few pitches, and they didn't start twice on the same pitch. This result speaks against a generalized representation of "standard pitches" and supports the interpretation of Fig. 1 as absolute memory for particular melodies.

To measure consistency across trials, the errors of first and second trials were correlated as well as tested for significant differences. Although there were no differences between error-sizes of the first and second trials (paired-samples T-test), there was a significant negative correlation between the errors of the first and second trials ($r = -.484$, $p < .05$), meaning for example, that one who sang very flat during first trial, sang sharp (but not more correct) during second trial. A comparison of hits and misses in both trials revealed that half of the participants show consistent performance in both trials (if one includes semitone errors into the "hit" category), whereas the other half shows improvement or worsening to the same degree.

An important question concerns the issue whether those participants with better performance have a certain skill (AP or "latent" AP) which might or might not be related to musical expertise, or if "piece factors" such as familiarity with the piece determine how well the correct pitches can be recalled. To find out whether the error sizes correlate with any of the musical education variables or factors asked for in the questionnaire, correlational analyses, T-tests, and ANOVAs were performed.

The error sizes did not correlate with any of the musical education variables. The two groups BARDOU and DIDO (who differed in the amount of time since the practice period as well as in the difficulty of the music and the rehearsal mode) did not differ in the average error size, but the DIDO-participants seemed to have fewer problems with the second piece and spontaneously sang a few bars more. It has to be kept in mind, though, that participants were not asked to sing as much as possible, but only as much as they wanted to. The better recall of a given piece could be caused by the shorter time interval between concert and experiment as well as by the overall easier music. One-way ANOVAs produced no significant effects for the dependent variable error-size and the independent variables voice part (soprano, alto etc), knowledge or liking of the piece, ability to name the notated starting pitch, number of times that the pieces had been sung or heard again, and the deliberate effort to sing the correct pitch. However, there is the interesting observation that participants who claimed to have AP came closer to the accurate pitch in the first trial ($r = -.5$, $p < .05$), but not so in the second trial.

6. DISCUSSION

The results of this experiment suggest that choir singers can recall absolute as well as relative features when asked to imagine a piece from their repertoire: They don't start singing on a randomly chosen pitch but instead on a pitch that is close to the originally practiced pitch notated in the score. This seems to be true only for the first song that participants imagined, whereas the second song seems to have been sung in a randomly chosen key. An explanation for this could be interference processes: The break between the two trials might have been too short to remove the key of the first song from participants' short-term memory, so that it interfered with the attempt to imagine the key of the second piece. An alternative explanation is that it is more difficult to imagine a given song (2nd trial) than a self-chosen song (1st trial).

Two questions remain to be answered. First: Why did the majority of participants sing flat? Is this result due to production deficits, meaning the participants have the right pitch in mind but can't sing it? Or do they have a flat pitch stored in long-term memory? A possible explanation for production deficits could be arousal, i.e. the participants (who were not vocally warmed up) did not put as much effort into singing as they normally do in a rehearsal or concert. The pieces sung here, though, did not for the most part start in a very high register for any voice, so that they must have been singable for the participants. Still it has to be kept in mind that choirs generally tend to rather sing flat than sing sharp. On the other hand, a memory deficit seems also possible: The starting pitches of any one's part are perhaps not as clearly represented in memory as those of pop songs because in choir music there is not only the own part that one hears and sings, but also the three or four other parts, so that memory seems to be more complex. Furthermore the pop songs of Levitin's study were participants' *favorite* songs meaning they have heard them hundreds of times in their lives, whereas the choir pieces have been practiced intensively for a few weeks but then rarely been heard again until the experiment.

Second: Are there correlates of more accurate pitch memory, and what are the correlates? The answer of this experiment seems to be that only the possession of AP serves as a predictor for pitch memory, whereas neither musical background nor the amount of time between training and experiment or the familiarity with a piece seem to be connected with the accuracy of pitch memory. Some factors have to be kept in mind, though: First of all the sample used here was very homogeneous regarding musical background, so there may be an effect that could not be found here. The questionnaire also did not generate enough variance in the answers, so that e.g. the liking of and familiarity with the pieces was very homogeneous.

For follow-up studies it would be interesting to use a more heterogeneous sample as well as a questionnaire with higher data level to find out if it is some people's ability to better remember absolute pitches (latent AP) or if it is in fact dependent on "piece factors" if a key is accurately remembered. Other "person factors" with possible influence could be the ability to sight-read, the recall-strategy (trying to "see the score" or to "hear the song"), and the amount of concentration during the experiment. The reason why I want to distinguish between person factors and piece factors is the implication for the theoretical explanation of memory for absolute pitches: If it turns out to be the capability of just some people then a skill such as latent AP seems to be the appropriate explanation, whereas if all people have this ability and it turns out to be dependent on the number of times a piece has been heard I would suggest looking for explaining memory mechanisms such as perceptual learning (Goldstone, 1998).

7. REFERENCES

1. Carroll-Phelan, B., & Hampson, P.J. (1996). Multiple Components of the Perception of Musical Sequences: A Cognitive Neuroscience Analysis and Some Implications for Auditory Imagery. *Music Perception, 13(4)*, 517-561.

2. Dowling, W.J., & Bartlett, J.C. (1981). The Important Interval Information in Long-term Memory for Melodies. *Psychomusicology, 1*, 30-49.

3. Goldstone, R.L. (1998). Perceptual Learning. *Annual Review of Psychology, 49*, 585-612.

4. Halpern, A.R. (1988). Perceived and Imagined Tempos of Familiar Songs. *Music Perception, 6(2)*, 193-202.

5. Halpern, A.R. (1989). Memory for the Absolute Pitch of Familiar Songs. *Memory and Cognition, 17(5)*, 572-581.

6. Levitin, D.J. (*1994*). Absolute Memory for Musical Pitch: Evidence from the Production of Learned Melodies. *Perception & Psychophysics, 56(4)*, 414-423.

7. Levitin, D.J., & Cook, P.R. (1996). Memory for Musical Tempo: Additional Evidence that Auditory Memory is Absolute. *Perception & Psychophysics, 58(6)*, 927-935.

8. Terhardt, E., & Seewann, M. (1983). Aural Key Identification and its Relationship to Absolute Pitch. *Music Perception, 1*, 63-83.

9. Terhardt, E., & Ward, W.D. (1982). Recognition of Musical Key: Exploratory Study. *JASA, 72*, 26-33.

10. Vitouch, O., & Gaugusch, A. (2000). Absolute Recognition of Musical Keys in Non-Absolute-Pitch-Possessors. In Woods, C., Luck, G.B., Brochard, R., O'Neill, S.A., & Sloboda, J.A. (Eds), *Proceedings of the Sixth International Conference on Music Perception and Cognition*. Keele, Staffordshire, UK: Department of Psychology. CD-ROM.

ISBN 1 876346 39 6