Bucknell University

[Bucknell Digital Commons](#)

1989

Memory for the absolute pitch of familiar songs

Andrea Halpern
Bucknell University, ahalpern@bucknell.edu

Follow this and additional works at: https://digitalcommons.bucknell.edu/fac_journ

Part of the [Behavior and Behavior Mechanisms Commons](#), and the [Psychological Phenomena and Processes Commons](#)

Recommended Citation

Memory for the absolute pitch of familiar songs

ANDREA R. HALPERN
Bucknell University, Lewisburg, Pennsylvania

Four experiments were conducted to examine the ability of people without "perfect pitch" to retain the absolute pitch of familiar tunes. In Experiment 1, participants imagined given tunes, and then hummed their first notes four times either between or within sessions. The variability of these productions was very low. Experiment 2 used a recognition paradigm, with results similar to those in Experiment 1 for musicians, but with some additional variability shown for unselected subjects. In Experiment 3, subjects rated the suitability of various pitches to start familiar tunes. Previously given preferred notes were rated high, as were notes three or four semitones distant from the preferred notes, but not notes one or two semitones distant. In Experiment 4, subjects mentally transformed the pitches of familiar tunes to the highest and lowest levels possible. These experiments suggest some retention of the absolute pitch of tunes despite a paucity of verbal or visual cues for the pitch.

One of the most salient attributes that we remember about a song is its pitches. Researchers in the psychology of music have spent a good deal of time in showing the extent and characteristics of memory for both isolated pitches, and pitches heard in the context of a tune. For instance, Deutsch (1970, 1972) has shown that memory for a single pitch is more disrupted under some circumstances (presentation of similar pitches during a retention interval) than others (presentation of dissimilar pitches or spoken numbers). Krumhansl (1979) has shown that memory for a single pitch within a tune is affected by whether the tune is tonal (pitches all drawn from the same key) or atonal (pitches drawn from different keys.)

Almost all such studies use unfamiliar tunes and short retention intervals on the order of a few seconds. It is commonly believed that most people will retain memory of a particular pitch for only about as long as the span of short-term memory, and then only under favorable circumstances. Dowling and Harwood (1986) note that even choral singers can have trouble entering with their part after a pause, if the piece has shifted key or if the music is complicated.

This inability to retain the absolute pitch of a tone presents few obstacles to the comprehension or enjoyment of music. Western music is based for the most part on the relationships of pitches to one another—that is, on "relative pitch." Thus, when "White Christmas" is sung correctly, the second note has to be one semitone, or 5.9%, higher in pitch than the first note, regardless of where in the voice the tune is started. Although most classical music is played at pretty much the same pitch in each

rendition, much popular and traditional music is tied to no particular starting pitch. Table 1 shows the variety of notated starting pitches found for some of the songs used here. Each song for which multiple versions were found includes at least two different keys. Indeed, the entire concept of a fixed pitch for a piece is a relatively recent development in Western music (Lindley & Wachsman, 1980).

We may fairly assume that almost all musicians and many people untrained in music have a reasonably well-developed sense of relative pitch. These people can recognize the correctness of a tune and perhaps produce good approximations of tunes they have heard. (Note that an inability to sing "in tune" may reflect a problem only at an output stage of the process. A person may still have a perfectly adequate internal representation of the pitch system.) This ability will also allow the transposition of pieces to higher or lower keys when music is sung, or the acceptance of a transposition when music is listened to. Less anecdotally, a number of researchers have found that memory for relative pitch, or interval sizes, is an important component in memory for music (Cuddy & Cohen, 1976), especially for familiar tunes (Dowling & Bartlett, 1981).

In addition to relative pitch, a small segment of the population possesses the ability to produce, name, and remember pitches on an absolute basis. This is known as perfect or absolute pitch (AP). Popular belief and some writers on the subject (e.g., Bachem, 1955) have suggested that this ability is innate, possessed in an all-or-none fashion. However, a review of the literature (Ward & Burns, 1982) suggests that this ability probably depends on early training. Some adults have successfully developed AP (Cuddy, 1970), and a range of AP ability can be found among self-labeled possessors. To my knowledge, no one has reported a case of a nonmusician "discovering" in adulthood that he or she has AP, again arguing against a nativist position.

Table 1
Starting Notes of Various Notated Versions of the Stimulus Songs
and Average Starting Notes Selected in Experiments 1 (Production) and 2 (Recognition)

Song	Notated Versions	Production		Recognition	
		Males	Females	Musicians	General
Yankee	G,G,A,A,B♭,B♭	C	C♯	B	B
Twinkle	D,G,F	C	D	C♯	B♭
Row	C,D,E	C	D♯	C♯	B
We wish	C,G,E♭,F	C	E	F	D
First Noel	F♯,F♯,F♯,F♯,D	C♯	C♯	B♭	B♭
Joy	B♭,B♭,D,D	D♯	D	B	C♯
Somewhere	B♭	C	C	C♯	B
Puff	not available	D	C♯	C♯	C

Note—In the production task, males sang in the octave below middle C and females in the octave above middle C. In the recognition task, pitches ranged from the B below middle C to the F above middle C.

Although it is not clear why some people learn AP and others do not, the ability seems to consist of an association between a pitch and a name for the pitch. For instance, Siegel (1974) presented a tone followed by interfering tones, and then a to-be-judged tone close in frequency to the first. The subjects had to say whether the last tone was higher or lower in pitch than the first. She found that AP possessors were superior to nonpossessors only in discriminating pitches separated by ¾ of a semitone, at which point the components could be differentially labeled. At separations of ¹⁄₁₀ of a semitone, both groups displayed similarly poor performance. Zakay, Roziner, and Ben-Arzi (1984) found a Stroop effect with AP possessors, such that the subjects had trouble naming a pitch if it was sung with an incongruent note name. Finally, Klein, Coles, and Donchin (1984) found that AP possessors display a smaller P300 evoked potential waveform than do nonpossessors in an auditory task, but not in a visual task. The authors noted that the P300 is associated with the use of working memory. The lack of that waveform may reflect the automatic activation of note names in the possessors.

As a lack of AP need be no impediment to the enjoyment of music, sometimes the possession of AP can actually interfere with a musical experience. Orchestras may not tune to the same note "A" as the AP listener (modern orchestras tune to an "A" of anywhere between 439 and 444 Hz or even higher), or a chorus may by choice or poor technique sing a whole piece sharp or flat, relative to the written score. Both of these situations can be disturbing to the AP possessor. Vernon (1977) reported that with increasing age, he heard all pitches sharper than they actually were. Because he had made emotional associations with the various keys (another example of verbal labeling), even the character of certain pieces had changed, and they now sounded "wrong" to him.

Thus most people can be said to possess relative but not absolute pitch for the production and perception of music. That is, they are only capable of storing the relationships between pitches. However, in spontaneous singing, people do need to choose some starting pitch. Im-

portant to the topic of the current investigation is that people report having a quasiperceptual experience when asked to imagine a familiar song. They claim to hear the song "running through their heads." Certainly part of this claim suggests a strong experience of hearing the pitches. Could representations of pitch be associated with this mental experience? If so, might these representations be available for retrieval whenever the occasion demands?

Evidence against the ability to retain absolute pitch has already been marshaled above. From this evidence, we should expect that imagined music should not be associated with particular starting pitches, only relative pitches. However, the representation of a song must be associated with some starting pitch on any given production attempt. Perhaps this pitch is in fact associated with the song more permanently, even when note names or other cues are not available.

Such a result has been found with another major component of music: tempo. Like pitch level, the tempo of traditional or popular music can vary and yet still be considered a legitimate version. Clynes and Walker (1982) had musicians tap their fingers to imagined music, and found very stable tapping rates over repeated trials. Halpern (1988b) found that untrained subjects were willing and able to set metronomes to the tempo of imagined tunes. The settings varied for different songs, suggesting that this task was reliably reflecting the different mental tempos of "White Christmas" (slow) and "When the Saints Come Marching In" (fast). These tempo settings were also stable across trials (Halpern, 1988c), although we must note that in this study only musicians were tested.

The following series of experiments explored the possibility that nonpossessors of AP might reliably associate a starting pitch with imagined tunes. Every attempt was made to remove extraneous memory cues for starting pitch in the experimental tasks. The initial expectation in this project was that little evidence for pitch stability would be found. However, this expectation was not met in Experiment 1, which motivated the remaining three experiments. Each experiment used a different method to probe for pitch stability in imagined tunes. No subjects were

AP possessors, so that any such stability could not be attributed to direct coding of note names for particular frequencies.

EXPERIMENT 1: PRODUCTION

In this initial experiment, subjects were asked to indicate the pitch of imagined tunes directly, by humming or singing the first pitch of familiar tunes. Stability was measured by means of asking for repeated productions either within the same session, or across sessions separated by 48 h. If pitch representation is stable, then the same or similar pitches should be produced under repeated observation both between and within sessions.

Method

Subjects. Participants were 22 Bucknell University undergraduates, 9 males and 13 females, who volunteered in exchange for extra course credit. To maximize their chances of being familiar with the stimulus songs, the subjects in all these experiments were required to have been raised in the United States. On the assumption that AP possessors would be aware of their ability by adulthood, the subjects were initially recruited on a sign-up sheet that requested AP possessors not to participate. Upon arrival for the experiment, they were again asked whether they possessed AP. Anyone who answered "yes" would of course have been disqualified. For this study, participants were unselected with regard to musical background, with one exception: People unable to hum a steady pitch long enough to enable accurate transcription were excluded.

Materials. The first phrases of eight songs familiar to this population served as stimuli. Some examples include "Yankee Doodle," "Puff the Magic Dragon," and "Somewhere Over the Rainbow."

Design. The comparison of most interest was the stability of pitch production for particular songs between, as opposed to across, sessions. For this comparison, a given song was probed four times. Either four observations were taken within a session, or two observations were gathered in each of the two sessions. For example, a subject might have to produce the starting note for "Yankee Doodle" four times in one session, or twice in the first session and twice in the second. These two conditions will be called same (i.e., song)-within (i.e., session) and same-between, respectively.

Stability of pitch production might occur artifactually if our subjects happened to be monotone singers. We therefore included two control conditions, in which observations of four different songs formed a quartet of observations. Like the same conditions above, each quartet could be observed within or between sessions, yielding different-within and different-between conditions. The members of each quartet were picked by randomly drawing four songs from the stimulus pool. For example, observations of "Puff," "Rainbow," "Yankee," and "The First Noel" were compared within one session or split between two sessions. If our subjects were simply monotone singers, then the standard deviations of the observations should be low and identical for same and different conditions. Higher standard deviations in the different conditions would indicate that the subjects could reliably produce different starting pitches.

In each session, the subjects produced 20 starting pitches. Each session contained two songs that were repeated four times, four songs that were repeated twice, and four filler songs. Two different sets of materials were constructed so that each song occurred in each condition for an equal number of subjects. For instance, the same number of subjects received "Yankee" four times within a session as received it twice in each session. For the different conditions, four quartets were randomly formed for analysis, subject

to the condition that no song appear more than twice over the set of quartets.

To summarize, each subject participated in each of four conditions. A subject's score in a condition was the average of four standard deviations. Each standard deviation was based on four pitch productions.

Procedure. Participants attended two half-hour sessions scheduled 48 h apart. At the beginning of the first session, they were shown a list of the stimulus songs and asked to indicate any unfamiliar songs. Volunteers who did not know all of the songs did not participate further. On each trial, they were presented with lyrics from the first phrase of a song, asked to run this phrase through their heads, and asked to imagine the pitch of the first note of the song. When they had comfortably obtained the pitch, they were told to hum or sing the first note into a tape recorder. To help ensure that subjects were accurately producing their imagined pitch, they rated each production on a 1 to 5 accuracy scale. If they were unable to assign a perfect pitch transformation score of 5, they were asked to repeat the trial until they were satisfied.

Several measures were taken to guard against the possibility that subjects might be using their memory of one production to influence the production of another trial (either to produce the same pitch for the next observation of the same song, or to use relative pitch to produce a desired note for a different song). Each observation was separated from the next observation of the same song by at least three intervening trials. Between each trial, a different set of approximately 10 arbitrary interfering notes, each about 750 msec in duration, was played on an electronic keyboard. The beginning notes of the interference series were alternately chosen from the high and low ends of the keyboard. To prevent vocalization of anything other than the starting pitch, the subjects chewed gum (sugarless, with a choice of flavors). And finally, the subjects were asked at the end of the study if they thought their responses on one trial had been influenced by previous responses. No one so indicated.

The pitches on the tape recorder were transcribed into music notation by a musician with extensive piano and singing experience, who was naive with respect to the experiment. Pitches were notated to the nearest semitone. Because subjects unable to maintain a steady pitch had been excluded, the transcription task was relatively easy. During training, the transcriber and the author did not disagree on any pitches. Playback from the tape recorder was matched against the keyboard to ensure that it reproduced pitches accurately.

Results

As noted above, a subject's score in a condition was the average of four standard deviations, where each standard deviation was based on four pitch productions. Note names were translated into arbitrary numerical equivalents for quantitative analysis. Figure 1 shows the results. An analysis of variance (ANOVA) with two within factors (song and session) confirmed that different songs (mean SD = 2.52 semitones) were more variable than same songs (mean SD = 1.28 semitones) [$F(1,21)$ = 46.5, p < .001]. As is obvious from the figure, observations within or between sessions were equally variable, and the two factors did not interact.

Of secondary interest was whether the pitch productions differed due to the gender of the subjects, which was of course expected, or due to the particular song being probed (the data of only 21 subjects were available for this analysis). Males did sing an average of nearly an octave (11 semitones) lower than females [$F(1,20)$ = 242.5,
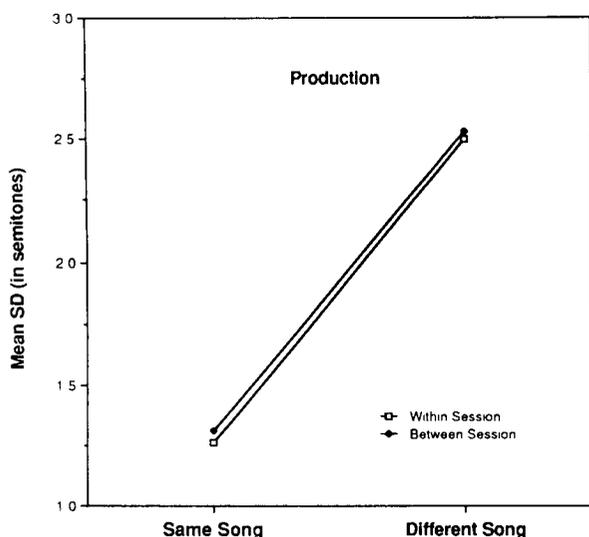
Figure 1. Mean standard deviations of pitch productions for same or different songs, either between or within sessions, in Experiment 1.

$p < .001$], and songs differed from one another [$F(7,140)$ $= 2.1$, $p < .05$]. "Joy To the World" was pitched highest and "Somewhere Over the Rainbow" was pitched lowest. These two factors did not interact; males and females used the same ordering of pitch height for the songs. An attempt was made to further divide the males and females into those with low and those with high voices. Such a division proved to be impossible here and in subsequent experiments, since nearly all participants claimed to have low voices. This claim was verified by our result that the average pitch setting for women was the D above middle C, and for the men, the D sharp in the octave below. These notes are in the ranges of the alto and baritone voices, respectively.

Discussion

The results of this study were very clear and striking. Variability for judgments was low both between and within sessions. Variability among different songs was much higher, so the subjects were sensitive to different starting pitches where appropriate. The lack of an interaction between songs and sessions means that the range in pitch judgments among "Happy Birthday," "Puff," "Somehwere," and "Yankee" was wider than that for four tries with "Happy Birthday," but that the range remained stable within or between sessions.

If singing or humming is an adequate externalization of mental pitch, then we seem to have evidence for the encoding of a form of absolute pitch among ordinary people. (Although the subjects were unselected for musical ability, a questionnaire revealed that 14 of them were nonmusicians, and 8 were musicians, which was defined in terms of their having had at least 5 years of training.) Because of our precautions in excluding proprioceptive and relative memory cues, we may be somewhat confident that the pitch of the songs was being retrieved directly.

However, Experiment 1 did rely on an overt task, which may have added some artifactual stability for good singers or instability for poor singers to the measurements. The next experiment used a recognition paradigm that did not require any vocal response. In addition to a group of subjects unselected for musical background as in Experiment 1, a group of people explicitly recruited for their musical background was also tested.

EXPERIMENT 2: RECOGNITION

In this study, subjects selected an imagined pitch from a keyboard that was hidden from view. Because hunting for the correct note inevitably required the subjects to play many extraneous notes, we expected this interference to cause higher overall variabilities than had been obtained in the production task.

Method

Subjects. A total of 39 Bucknell University undergraduates participated for extra course credit: 13 musicians (6 males and 7 females), and 26 who were unselected for musical background ("general subjects": 7 males and 19 females). Musicians were defined as people with at least 5 years of musical training (instrument or voice lessons), and they were told during recruitment that we were especially interested in musicians. A questionnaire later revealed that many general subjects had at least some musical training, which is common in this population.

Materials and Design. The same songs and presentation order as in Experiment 1 were used in this experiment. Each participant again attended two half-hour sessions 48 h apart.

Procedure. The procedure began as in Experiment 1, with the subjects asked to indicate familiarity with all the songs. On each trial, the subjects were presented with lyrics from the first phrase of a song. They were told to run this phrase through their heads and imagine the pitch of the first note of the song. When they had comfortably obtained the pitch, they attempted to find it on a Yamaha portable electronic keyboard. To minimize the possibility that subjects would notice and remember either the name of the note they had picked, or the keyboard placement of the selected note, the keyboard was positioned upside down relative to the subject, and hidden from view under a drape. Pilot tests suggested that even the musicians found that this arrangement impeded realization of where they were on the keyboard. The other precautions against extraneous memory cues from Experiment 1 were repeated here: the subjects repeated their attempt on each trial until they were satisfied with their choice, interfering notes were played between each trial, chewing gum was provided, and the subjects were asked whether they had remembered previous notes. There was universal agreement that each trial had to be performed independently.

Results

As before, the main dependent measure was the average standard deviation of the four quartets of observations in the *same-within*, *same-between*, *different-within*, and *different-between* conditions for each subject. Analyses were conducted separately for the musicians and the general subjects.

The results for the musicians and general subjects are shown in Figure 2. The results for the musicians were statistically identical to those in Experiment 1. *Same* songs had a lower mean *SD* (2.36 semitones) than did *different* songs (4.03 semitones) [$F(1,12) = 31.9, p < .001$], but
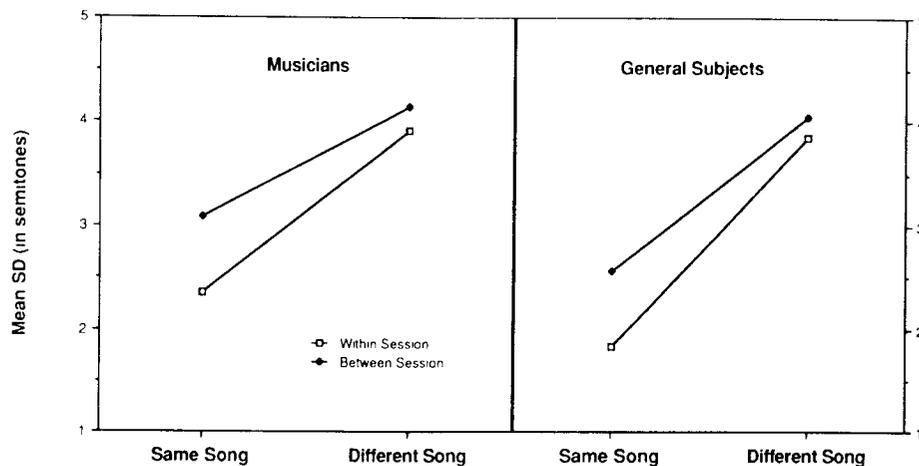
Figure 2. Mean standard deviations of pitch recognitions for same or different songs, either between or within sessions, for musicians and general subjects in Experiment 2.

within and between sessions were equally variable, and the two factors did not interact. Note that, as predicted, standard deviations are about twice as high here as in Experiment 1.

Although the results for the two groups look similar, they in fact differ statistically. Like the musicians, general subjects differentiated between the *same* (mean SD = 2.2 semitones) and *different* songs (3.95 semitones): $F(1,25)$ = 42.3, p < .001. However, the effect of session was also significant: between sessions were more variable (mean SD = 3.31 semitones) than within sessions (2.84 semitones) [$F(1,25)$ = 6.9, p < .05]. These two factors did interact [$F(1,25)$ = 5.0, p < .05]. As is apparent from the figure, the variability of *same* songs increased in between relative to within sessions [$F(1,25)$ = 10.4, p < .01], but *different* songs were equally variable between and within sessions.

As before, a secondary analysis examined whether the pitch productions differed due to the gender of the subject or the song being probed. Gender differences would suggest that pitches were being internally represented with some reference to the characteristics of the vocal mechanism. However, on the average, males and females selected the same pitches, among both musicians and general subjects. As in Experiment 1, songs were also assigned different pitches [$F(7,77)$ = 4.6, p < .001 for musicians; $F(7,168)$ = 2.1, p < .05 for general subjects]. In both groups, "We Wish You a Merry Christmas" was pitched the highest and "The First Noel" the lowest (see Table 1). Interestingly, the actual pitches selected by both males and females were in the same range as the females' production in Experiment 1 (just above and below middle C). The subjects in Experiment 1 and the general subjects in Experiment 2 produced a similar ordering of pitch height over songs [$r(6)$ = .76, p < .05], but the musicians' ordering in Experiment 2 did not correlate with the ordering in Experiment 1 [$r(6)$ = .26].

In contrast with the results of Experiment 1, the time taken to complete each session and the reported difficulty were very variable. Time to complete the task ranged from 20 to 45 min. The most difficulty was not always experienced by the general subjects: Some musicians seemed reluctant to assign absolute pitches, perhaps because most musical tasks depend exclusively on relative pitch ability. In contrast, some general subjects seemed to be particularly sure about one or two songs, assigning the same pitch rapidly and confidently in all four observations.

Discussion

This recognition task appeared to be more difficult than the production task, probably due to the interference with the imagined note created by the need to hunt for the correct note on the keyboard. Nevertheless, the musicians showed the same pattern of results as before: The recognition of imagined notes remained equally stable within and across sessions. While standard deviations were larger than in Experiment 1, an interval of 3 semitones (the approximate range of SD shown here) is still extremely small. For the reader not versed in music terminology, an example of that interval is the distance between the first two notes of "Greensleeves."

The general subjects also showed stable settings for the *different* songs, regardless of the observations being split or massed. However, *same* songs did elicit more variability when observed across sessions than between sessions. Examination of the means in this experiment suggests that the interaction of song and session among the general subjects resulted from an aberrantly low value of 1.83 semitones in the *same-within* condition. This mean is .5 semitones lower than the next highest mean in Experiment 2, and the only one comparable to the means in Experiment 1. This result may reflect the confident and consistent assignment of pitches to a few songs by some of the general subjects. In this context, it is interesting

that the nonmusician experimenter, even after running several studies, still maintained that some notes were "correct" starting pitches.

Some people, especially singers, can mimic possession of absolute pitch by using cues from their own voices. Someone who knows that her lowest note is the G below middle C can judge any incoming note against this standard, and use relative pitch to name that note or produce a requested note. Other singers can judge or produce pitches reasonably accurately by using muscular cues from the vocal tract. However, the lack of gender differences in Experiment 2 suggests that the subjects in fact were performing the task without appealing to kinesthetic anchors.

Another interesting observation in this vein is that the highest and lowest pitched songs were not obvious choices. In Experiment 1, "Somewhere" was pitched lowest, and in fact the second note requires the large leap of an octave in pitch from the first note. "Joy to the World" was the highest pitched song, and in fact this song descends a whole octave from the word "Joy" to the word "[Lord is] come." Both choices suggest the influence of the singing mechanism in the production task. However, the highest choice in Experiment 2 was "We Wish," which *ascends* in the first few notes, and overall through the first phrase. The lowest choice of "The First Noel" *descends* in its first three notes. This result again suggests that the recognition task was tapping a mental representation tied only loosely, if at all, to muscular or vocal cues.

The two tasks thus far might have favored people who were good at singing or at note-finding, respectively. Even if mental representations of the pitches were adequate, a person with poor singing might have sung several extraneous pitches before settling on a satisfactory one in Experiment 1. In Experiment 2, a person with poor dexterity or one who was completely unfamiliar with keyboards might have played more pitches per trial before selecting the appropriate one, thereby again generating more interference. Accordingly, a third experiment used a rating task, to remove any requirements for output skills in the main task.

EXPERIMENT 3: RATINGS

To reveal their preferred pitch, the subjects first performed the same task as in Experiment 1 as an initial phase in Experiment 3. Thereafter, they rated the goodness of presented pitches, including their preferred pitches, as starting notes for the songs. If subjects maintain a stable representation of starting pitch, they should rate their previously selected pitch as the most adequate note, and rate notes higher and lower as less adequate.

Of secondary interest was the pattern of ratings. Krumhansl and Shepard (1979) presented listeners with tones in ascending or descending scale patterns. Listeners then heard one of the 13 possible notes in the octave and rated how well the final tone completed the scale sequence. They found that preferences were in accord with

Western music theory, especially among musically trained listeners. For example, notes a perfect fifth (7 semitones) or a major third (4 semitones) away from the beginning note of a scale are considered musically close to that beginning note (e.g., a G or an E relative to a C major scale), even though the notes are fairly distant in frequency from one another. The authors found that these special notes were also preferred over most other notes. Notes from the C major scale (diatonic notes) were preferred over nondiatonic notes in general. This pattern was particularly strong in their one subject with absolute pitch. In another study, Krumhansl (1979) found that diatonic tones are also rated as being more similar to one another than are other pairs of tones close in frequency but far in terms of musical relationship. If the subjects in the current study gave high ratings to notes a third and a fifth away from their preferred note, this would further argue for the psychomusical similarity of those tones. Because Krumhansl found differences between musical and non-musical subjects, Experiment 3 included this subject division as well.

Method

Subjects. A total of 29 volunteers participated in Experiment 3: 15 musicians and 14 nonmusicians. Musicians were defined as people with at least 5 years of training in voice or on an instrument. Nonmusicians were defined as having 0 to 1 year of training. All subjects were students attending summer classes at Bucknell.

Materials. Four songs familiar to this population served as stimuli: "Row Row Row Your Boat," "Joy to the World," "Puff the Magic Dragon," and "We Wish You a Merry Christmas."

Procedure. Experiment 3 was performed in two sessions. The first 10-min session was a miniversion of Experiment 1. The subjects were first asked to indicate their familiarity with the four stimulus songs. They then produced the starting pitch of each song by humming into a tape recorder, following the same instructions as in Experiment 1. Each song was tested twice. The average of the two productions was considered to be the preferred note.

The second session occurred between 2 and 5 days later. Each song was presented for rating with 15 played notes: the preferred note, plus notes from 1 to 7 semitones above, and 1 to 7 semitones below the preferred note. The 15 notes × 4 songs made for 60 trials, and each trial was repeated, for a total of 120 trials. These were presented in the same quasi-random order for each subject. At least 8 trials intervened between repetitions of any given trial.

For the rating task, the subjects were read the name of one of the four stimulus songs and told to imagine the first note of that song in their minds. The experimenter then played a note on a Yamaha electronic keyboard. The subjects were told to rate the note using a 7-point rating scale, with 1 indicating that the note played was very dissimilar to the note they were imagining, and a 7 indicating that the note was very similar to or exactly the same as the note they were imagining. Pilot work had indicated that the subjects were more comfortable with the high endpoint of the scale indicating extreme similarity *or* identity with the imagined note, rather than identity alone.

The subjects were instructed to use the entire rating scale for their answers. As before, interference notes were played between each trial. Sessions lasted about 20 min.

Results

The dependent measure was the average of the two ratings given to each probe note and song combination by

each subject. A three-way ANOVA was performed on the effects of group, song, and probe note (the latter two being within factors). The song factor was not at all informative, and will not be discussed further. Musicians and nonmusicians performed identically, for neither the main effect nor interaction between probe note and group reached significance. So the result of major interest was whether the ratings of probe notes reliably differed from each other. As can be seen in Figure 3, this result was obtained $[F(14,378) = 5.6, p < .001]$.

The pattern in the figure resembles a series of peaks and valleys. In addition to the preferred note (P), preference peaks occur at -4, $+3$, and $+7$ semitones relative to the preferred note. Because the range of rating choices was quite narrow, at least partly due to the averaging of two ratings, a liberal post hoc test was used to detect any significant differences. Duncan's test confirmed ($p < .05$) that each of the peaks was significantly different from the more distant of its adjacent neighbors. So, for instance, the difference between P and -1 was not reliable, but that between P and $+1$ was. The preference valleys were all significantly different from their nearest peaks, but the peaks did not differ from each other.

Because Session 1 was a miniversion of Experiment 1, we can similarly examine the consistency of the two productions that each subject gave for each song. In fact, for one of the songs, the modal difference between the two productions was one semitone; and for the others, the modal difference was zero semitones. Taken together, differences of zero or one semitone formed the majority of production discrepancies for each song.

Discussion

Both musicians and nonmusicians rated some pitches as better beginning notes for a particular song than other
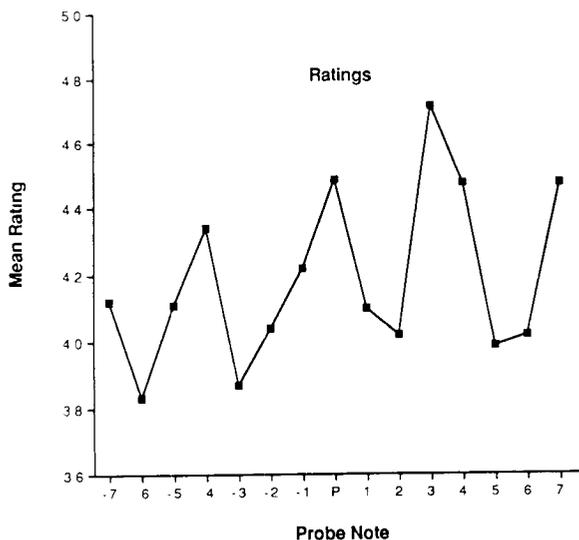

Figure 3. Mean goodness rating on a 1 to 7 scale of subjects' preferred notes (P), and notes one to seven semitones above and below P, in Experiment 3.

pitches, as long as 5 days after producing a preferred starting note. We must keep in mind that no subjects possessed absolute pitch, and that both memory and vocal musculature cues for pitch (From Session 1 to Session 2, and from trial to trial) were minimally, if at all, available.

As predicted, the preferred note from the first session was also rated high in the second session. However, the remaining ratings formed a very regular pattern, which was only somewhat similar to the pattern predicted by Krumhansl's (1979; Krumhansl & Shepard, 1979) work. Notes preferred equally to the P note were a major third below (-4 semitones), and a minor third ($+3$) and fifth ($+7$) above the P note. Of these, only the note a fifth above comes from the same scale as the P note, if the P note is considered to be the tonic, or starting note, of the scale. Two of the least preferred notes (-6 and $+6$) are in fact an augmented fourth away from P, sometimes considered to be the least diatonic interval. But conversely, the scale interval of the major second ($+2$ semitones) was downrated, as were the lower minor third (-3) and lower minor second (-1). In summary, we cannot characterize the preferences as simply being derived from the same scale as the preferred tone.

Instead, the pattern seems to be composed of "preference regions" of pitch height, each separated by three or four semitones. Although subjects seem willing to give a preferred note for a song, they also seem willing (or able) to pitch the entire tune up or down several semitones. They seem not as happy to accept a tune pitched only one or two semitones away from the preferred note. This pattern is consistent with a quasi-absolute representation of song pitch. The ability to distinguish intervals as small as a semitone without verbal or other cues suggests that there is some notion of absolute pitch directly represented in an imagined tune.

Nonimagery Followup

In the previous experiments, we assumed that the best way to capture subjects' ability to retrieve absolute pitch was to ask them to generate an image of the song. In a followup to this experiment, an additional 11 nonselected subjects performed the rating task, but with instructions that omitted any reference to playing the tune inside their heads. The relevant part of the instructions stated: "Your task is to rate the note I play on the 7-point rating chart which you see in front of you. A rating of 7 means that the note played is very similar to or exactly how you would begin the song and a rating of 1 means that the note played is very dissimilar to how you would begin the song."

When interviewed, all 11 subjects claimed that they performed both the initial production session and the second rating session by hearing the first line in their heads or playing the songs to themselves. The resulting rating pattern was very similar to that in the main experiment. The ANOVA confirmed that probe notes differed from one another $[F(14,140) = 3.1, p < .001]$, although fewer probe note differences were significant compared to what

was observed in the main experiment (perhaps due to the small number of subjects here).

This followup suggests that a true nonimagery control condition may be difficult to engineer in these kinds of tasks. Thinking about the musical characteristics of a song may in fact be equivalent to imaging the song.

In the next study, we explored the flexibility of the pitch representation by explicitly asking the subjects to transform the pitch of imagined tunes.

EXPERIMENT 4: PITCH TRANSFORMATION

Halpern (1988b) found that subjects could transform the tempo of a familiar tune faster and slower relative to a preferred tempo. However, there were limits to such transformations. Upper and lower limits were proportional to the preferred tempo. Thus, for example, although the subjects could imagine "When the Saints Come Marching In" at 175 beats per minute as a fastest tempo, "White Christmas" was imagined only at 143 beats per minute. Neither value is close to the highest possible setting on the metronome that was used for setting the beat.

In this study, the subjects tried to transform imagined tunes to the highest and lowest possible pitches. If limits on such transformations exist, we should again see those limits being proportional to preferred pitches, rather than reflecting an absolute bound on imagined pitches. Recognition was employed for eliciting responses in order to remove any effects of limits imposed by the vocal apparatus.

Method

Subjects. The participants wre 20 Bucknell undergraduates who volunteered their time: 10 males and 10 females.

Materials. The first phrases of 10 songs drawn from the stimulus pool used in the previous experiments served as stimuli.

Procedure. Preferred notes were first determined for all the songs. To locate preferred notes, the recognition procedure from Experi-

ment 2 was followed: after imagining a song, the subjects located the imagined first pitch on a keyboard. The songs were presented in a different random order for each subject.

To determine the highest and lowest imaginable pitch, a modified recognition procedure was used. After announcing the name of a song, the experimenter played the previously selected preferred pitch, and then began playing successively higher (or lower) pitches in semitone increments. As each note was played, the subject used a 4-point scale to rate the difficulty of imagining the song at that pitch: a 1 meant that he or she could easily imagine the song beginning at that pitch, and a 4 meant that it was very difficult or impossible to imagine the song at that pitch. The first note receiving a 4 was recorded as the highest (or lowest) note. The same song was then probed for the lowest (or highest) note.

For half the subjects, a given song was first probed for highest pitch followed by lowest. This order was reversed for the other half of the subjects. As usual, interference notes were played between trials; but the chewing gum was omitted. The sessions lasted about 30 min.

Results

Because this was a new version of the recognition task, responses were analyzed by gender, as well as by song and task. As in the previous recognition task, males and females differed neither overall nor in their behavior with respect to the other two factors. Figure 4 shows the pitch settings for each song in each task.

Pitch settings in the three tasks differed from one another [$F(2,30) = 25.9, p < .001$], as did settings for the 10 different songs [$F(9,135) = 8.0, p < .001$]. As can be seen by the essentially parallel lines in the graph, these two factors did not interact. The three tasks correlated strongly with one another. All three correlation coefficients exceeded .91, with $df = 8$. Because the pitch range was unexpectedly narrow, Duncan's test was used to detect any differences among songs. It showed that only some of the songs so differed ($p < .05$). For instance, "Joy to the World" was higher than all the other songs, "Noel" and "Yankee" were higher than "Row" and "Battle Hymn," and "Star Spangled Banner" and "Silent Night" were higher than "Row." The highest aver-
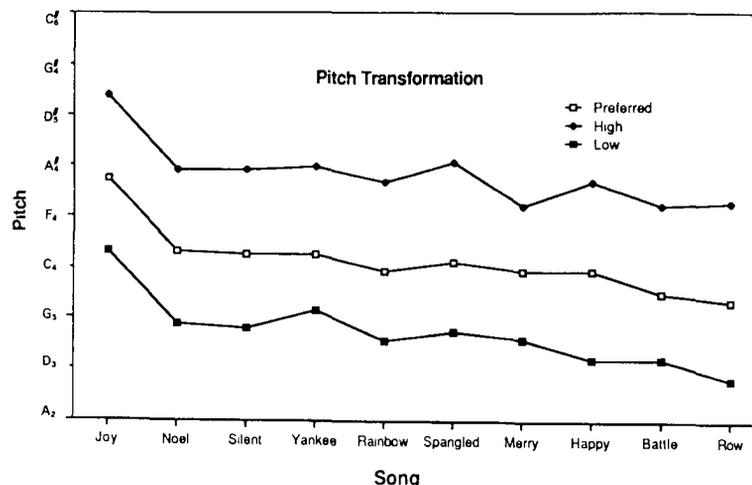

Figure 4. Mean preferred, highest imaginable, and lowest imaginable pitch of 10 familiar songs in Experiment 4.

age setting in the task was F in the second octave above middle C (the top line of the treble staff) for "Joy," and the lowest was the C sharp below middle C (second space on the bass staff) for "Row." However, most highest settings were around the A above middle C; and most lowest settings were around the F below middle C.

Discussion

Evidently, this task was meaningful to the subjects. No one objected to carrying it out, and the results showed consistent differences in pitch selection among the three tasks.

Once again, the absence of gender differences suggests that the subjects were not depending on surreptitious singing or other vocal cues to perform the task. The use of chewing gum in the previous experiments may have been an unnecessary precaution. Pitch settings, especially in the highest condition, were considerably higher than in the other experiments, in which preferred notes were selected. Although some of our subjects may have been physically able to sing that high, it is almost certain that they would have felt extremely uncomfortable doing so. This again supports the contention that the subjects were truly conducting a mental rather than a physical transformation task.

Were the subjects simply selecting an arbitrary upper and lower limit for all the songs? If they were, we would still have some interesting evidence that subjects could consistently select that designated note in the absence of verbal or visual cues. Settings for the different songs did not vary as much in this study as they did in the comparable study with tempos (Halpern, 1988b). However, we do have several informative examples of differentiation among songs. For "Joy to the World," the preferred setting was much higher than for any other song (in the range of other songs' highest settings), as were its highest and lowest settings. "Row" had the lowest preferred setting, and its lowest setting was two semitones below that for the next lowest song. Thus, we may conclude that there does not exist an absolute upper or lower bound for the range of imagined pitches used here. Of course, all the pitches selected here were well within the thresholds of hearing of approximately 20 and 20,000 Hz. It is likely that people would have a great deal of trouble imagining more extreme pitches.

GENERAL DISCUSSION

To summarize, four experiments required people of varying musical backgrounds to externalize the pitch of an imagined tune. The tunes were all popular or traditional, and they had very probably been heard and performed at many pitch levels by the participants in the past. Considering that people without absolute pitch have difficulty remembering isolated tones, or tones in unfamiliar melodies, performance in these experiments was indicative of considerable memory for an arbitrary starting pitch of familiar tunes.

Familiar tunes thus may be stored with much exact information. People apparently needn't rely solely on such musical abstractions as contour or scale structure, as they do when processing unfamiliar tunes (Dowling, 1978). Undoubtedly, both more literal and more abstract aspects of tune representation are used in musical tasks. Although one may have a definite idea of an appropriate literal starting pitch and tempo for a tune, memory for the interval, contour, or key might help one retrieve a note from a less familiar part of the tune, correct a vocal error, or add a harmony line or improvise. In general, storage of all the exact pitches of a familiar tune would seem to be inefficient and even deleterious to our ability to transpose, as noted earlier.

Although the argument in this article is that ordinary people store absolute pitch to a greater extent than is commonly believed, this should not be taken to mean that everyone has a mild case of what is traditionally called AP. Clearly, verbal labels for notes were not used in these tasks, and consistency of pitch judgment was not up to the level shown by true AP possessors. When AP possessors make errors in pitch naming, they are frequently octave errors or errors of one semitone (e.g., Lockhead & Byrd, 1981). Although the tasks in these experiments were not conducive to octave errors, Experiment 3 showed that subjects were not inclined to equate tones separated by a semitone in preference ratings (although production attempts were sometimes only one semitone apart).

Throughout this project we have assumed that the subjects were in fact imagining the song in response to our instructions. Certainly people reported their subjective impressions of imagery even in the nonimagery control condition of Experiment 3. However, we cannot verify this representation as strongly as we can when a time-dependent process is being measured. For instance, Weber and Brown (1986) showed that the time taken to describe the pitch contour of melodies was about the same whether the tune was perceived or imagined. Halpern (1988a) found that time to complete tasks requiring comparison of two parts of a familiar tune increased when those two parts were distant from one another in the actual tune. In these experiments, we depended for the most part on stability of pitch memory as an indirect indication of the inclusion of starting pitch in the tune representation. Thus, although this work cannot conclusively be said to have established the existence of the "song in the head," further work along these lines could include time-dependent tasks to more rigorously pursue the specific nature of imagery for pitch.

REFERENCES

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America, 27,* 1180-1185.

Clynes, M., & Walker, J (1982). Neurobiologic functions of rhythm, time, and pulse in music. In M Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 171-216). New York: Plenum.

Cuddy, L L. (1970). Training the absolute identification of pitch. *Perception & Psychophysics, 8,* 265-269.

Cuddy, L. L., & Cohen, A. J. (1976) Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28,* 255-270.

DEUTSCH, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, **168**, 1604-1605

DEUTSCH, D. (1972). Mapping of interactions in pitch memory store. *Science*, **175**, 1020-1022.

DOWLING, W. J. (1978) Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, **85**, 341-354.

DOWLING, W. J., & BARTLETT, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, **1**, 30-49.

DOWLING, W. J., & HARWOOD, D. L. (1986). *Music cognition*. Orlando, FL: Academic Press.

HALPERN, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **114**, 434-443.

HALPERN, A. R. (1988b). Perceived and imagined tempos of familiar songs. *Music Perception*, **6**, 193-202.

HALPERN, A. R. (1988c). [Stability of mental tempo]. Unpublished raw data.

KLEIN, M. COLES, M. G. H., & DONCHIN, E. (1984). People with absolute pitch process tones without producing a P300. *Science*, **223**, 1306-1309.

KRUMHANSL, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, **11**, 346-374.

KRUMHANSL, C. L., & SHEPARD, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 579-594.

LINDLEY, M., & WACHSMAN, K. (1980). Pitch. In S. Sadie (Ed), *The new Grove dictionary of music and musicians* (Vol. 14, pp. 779-786) London Macmillan

LOCKHEAD, G. R., & BYRD, R (1981). Practically perfect pitch *Journal of the Acoustical Society of America*, **70**, 387-389

SIEGEL, J. A. (1974). Sensory and verbal coding strategies in subjects with absolute pitch. *Journal of Experimental Psychology*, **103**, 37-44

VERNON, P. E. (1977). Absolute pitch: A case study. *British Journal of Psychology*, **68**, 485-489.

WARD, W. D , & BURNS, E. M. (1982) Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp 431-451). New York: Academic Press.

WEBER, W. J., & BROWN, S (1986). Musical imagery. *Music Perception*, **3**, 411-426.

ZAKAY, D., ROZINER, I., & BEN-ARZI, S. (1984). On the nature of absolute pitch *Archive für Psychologie*, **136**, 163-166.