

Name that tune: Identifying popular recordings from brief excerpts

E. GLENN SCHELLENBERG

University of Toronto, Mississauga, Ontario, Canada

PAUL IVERSON

University of Washington, Seattle, Washington

and

MARGARET C. MCKINNON

University of Toronto, Mississauga, Ontario, Canada

We tested listeners' ability to identify brief excerpts from popular recordings. Listeners were required to match 200- or 100-msec excerpts with the song titles and artists. Performance was well above chance levels for 200-msec excerpts and poorer but still better than chance for 100-msec excerpts. Performance fell to chance levels when dynamic (time-varying) information was disrupted by playing the 100-msec excerpts backward and when high-frequency information was omitted from the 100-msec excerpts; performance was unaffected by the removal of low-frequency information. In sum, successful identification required the presence of dynamic, high-frequency spectral information.

A song's identity is specified by its pitch and rhythmic structure. Accordingly, these structures have been the primary focus of psychological research on music (e.g., Jones & Yee, 1993; Krumhansl, 1990). Songs are a particularly interesting domain of study because their identity is determined from abstracted information about *relations* between tones, rather than from the tones' *absolute* characteristics. For example, the frequency (pitch) of the initial tone of "Happy Birthday" can be selected arbitrarily, but the song will retain its identity if the relations (intervals) between tones are preserved. Hence, regardless of whether a song is sung with a high or a low voice, it is recognizable if its intervallic structure is maintained. Differences in tone durations (rhythm) work similarly. Songs can be sung fast or slow and still be recognized (within limits; see Warren, Gardner, Brubaker, & Bashford, 1991), if the durational differences between consecutive tones maintain the correct ratios.

By contrast, the sound quality of musical instruments (*timbre*) is irrelevant to a song's identity. "Happy Birthday" is recognizable regardless of whether it is played on a trombone or a piano. Timbre is typically defined by what

it is *not*: characteristics of sounds other than pitch, duration, or amplitude (see, e.g., Dowling & Harwood, 1986; Hajda, Kendall, Carterette, & Harshberger, 1997). Whereas these parameters can be measured on ordinal scales, timbre is multidimensional and difficult to define (Hajda et al., 1997). Nonetheless, we know that listeners' perception of timbre is a function of *static* attributes of tones, such as the steady state frequency distribution of harmonics, and of *dynamic* or time-varying attributes, such as changes in harmonics at tone onsets (see, e.g., Grey, 1977; Iverson & Krumhansl, 1993; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Pitt & Crowder, 1992).

Although a song's identity is defined by relational information, this does not preclude the possibility that absolute information about pitch, tempo, or timbre is also stored in auditory memory. Absolute attributes of voices (e.g., pitch and timbre) are irrelevant to a word's identity, yet talker identity is stored in episodic memory for words (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993). In the experiments conducted by Pisoni and his colleagues, participants typically heard a list of words spoken by different talkers and were asked to identify words that had been presented previously in the list. Consistent with the principle of *encoding specificity* (Tulving & Thomson, 1973), recognition was best if the same talker said the word both times, but relatively poor when the repeated word was said by a different talker. Voice recognition may be somewhat unique, however, in that listeners appear to rely on different cues for different speakers; for example, some famous voices are recognized equally well when they are presented backward or forward, presumably because listeners

Funding for this research was provided by a grant awarded to the first author from the Natural Sciences and Engineering Research Council of Canada. We thank Dennis Phillips for extensive discussions about all aspects of the study, Susan Hall for her assistance in preparing Figure 1, and Andrea Halpern, Dan Levitin, John Wixted, and an anonymous reviewer for their insightful comments on earlier versions of the manuscript. Correspondence concerning this article should be addressed to E. G. Schellenberg, Department of Psychology, University of Toronto at Mississauga, Mississauga, ON, L5L 1C6, Canada (e-mail: g.schellenberg@utoronto.ca).

are using cues other than those based on dynamic spectral information (Van Lancker, Kreiman, & Emmorey, 1985).

Absolute attributes also play an important role in memory for popular recordings, despite their irrelevance to a song's identity. When respondents are asked to sing short passages from well-known recordings, they tend to do so at a pitch (Levitin, 1994) and tempo (Levitin & Cook, 1996) that closely approximate those of the original recordings. Anecdotal evidence indicates that listeners can recognize songs rapidly when scanning through radio stations for a song that they like or when participating in radio contests (e.g., "Name that Tune") that require identification of brief excerpts of recordings. Although it is possible that the limited relational information available in these segments is sufficient for recognition, we suggest that such recognition relies more on absolute information based primarily on timbre rather than on pitch or tempo. (*Timbre* can also refer to the global sound quality of the recording and orchestration of a particular song.) Indeed, listeners' ability to perceive differences in timbre is remarkable. For example, sequences of 10-msec tones with identical pitch but different timbres can be distinguished from comparison sequences with the same tones played in a different order (Warren et al., 1991). Moreover, specific musical instruments can be identified in forced-choice tasks involving tones of similarly short durations (Robinson & Patterson, 1995a).

In the present investigation, listeners were asked to identify excerpts from recordings of popular songs that were too brief to contain *any* relational information. We selected five recordings that were highly popular in North America in the months preceding data collection and, therefore, likely to be familiar to undergraduates. Our goal was twofold: (1) to explore the limits of listeners' ability to identify recordings from very brief excerpts and (2) to identify stimulus attributes necessary for successful identification. Although our excerpts contained absolute information about pitch and timbre, their brevity (100 or 200 msec) precluded the possibility of identifying words or multiple tones presented successively. Our hypothesis was that listeners would rely on timbre more than on absolute pitch in these brief contexts. Accordingly, the excerpts were altered in some conditions, to examine which attributes were important for identification. Specifically, we altered the distribution of frequencies in the harmonic spectrum through high-pass (frequencies < 1000 Hz attenuated) and low-pass (frequencies > 1000 Hz attenuated) filtering and the dynamic information by playing the excerpts backward. These alterations affected the timbre of the excerpts but had little impact on their perceived pitch. Thus, differential responding across conditions would indicate listeners' greater reliance on timbre than on absolute pitch.

METHOD

Participants

The listeners were 100 undergraduates enrolled in psychology courses at a medium-sized Canadian university located a few miles

from downtown Detroit. Participation in the experiment took approximately 20 min, for which the students received partial course credit. An additional 10 listeners were recruited but excluded from the testing session for failing to meet the inclusion criterion (see the Procedure section).

Apparatus and Stimulus Materials

We searched through "HOT 100" charts in *Billboard* magazine to select five recordings that were highly popular in North America in the months preceding data collection: (1) "Because You Loved Me," performed by Celine Dion; (2) "Exhale (Shoop Shoop)," performed by Whitney Houston; (3) "Macarena," performed by Los Del Rios; (4) "Missing," performed by Everything But the Girl; and (5) "One Sweet Day," performed by Mariah Carey and Boyz II Men. The extensive airplay accorded these songs ensured that it was likely that anyone who had listened to popular music during this period had been exposed to all of them. The recordings were purchased on compact disc. An excerpt from each disc was digitally copied onto the hard disk of a Macintosh PowerPC 7100/66AV computer in 16-bit format (sampling rate of 22.05 kHz) using the SoundEdit 16 software program. Excerpt onsets were chosen to be maximally representative of the recordings (experimenters' judgment); each started on a downbeat at the beginning of a bar. One of the excerpts ("Macarena") contained no vocals.¹

There were five experimental conditions. In one condition, the excerpts were 200 msec in duration; this duration was selected so that the task would be challenging but not impossible. In a second condition, the excerpts were shortened to 100 msec by deleting the second half. Frequency spectra at 50 msec from excerpt onsets are illustrated in Figure 1. In a third condition, the 100-msec excerpts were played backward (as in Van Lancker et al., 1985), which disrupted the dynamic information but had no effect on the static (steady state) information. In the remaining two conditions, the original (forward) 100-msec excerpts were high-pass or low-pass filtered (following D. L. Halpern, Blake, & Hillenbrand, 1986, but with a cutoff frequency of 1000 Hz, similar to Compton, 1963), using the SoundEdit program.² The stimuli were presented to the listeners binaurally via headphones (Sony CD 550) at a comfortable listening level. Inclusion of 10-msec onset and offset ramps proved to be undetectable to the experimenters, so the excerpts were *not* ramped.

Procedure

The listeners were tested individually; 20 were assigned to each of five conditions. They wore headphones and sat in front of the computer monitor in a quiet room. A SoundEdit file was open on the computer, which allowed the listeners to see the waveforms for each of the five excerpts. (None of the listeners reported any familiarity with waveforms.) The order of the waveforms was randomized separately for each condition. To hear an excerpt, the listeners used a mouse connected to the computer and clicked on one of the waveforms. The listeners were provided with an answer sheet that listed the five artists and song titles (alphabetical order) and were required to match the five excerpts with the five songs on the answer sheet. This method differed from multiple-choice tasks in that the five judgments from any individual listener were not independent (e.g., one error ensured another error). The listeners were allowed to hear the test excerpts repeatedly and in any order they chose.

Prior to the test session, the participants were informed that there would be a pretest, to verify that they were familiar with the five songs used in the experiment. Because many of the students might have been familiar with the recordings but not with the names of the songs, the pretest also served to familiarize or refamiliarize the participants with the song titles and artists, as was required in the subsequent experiment. The pretest involved presenting a single 20-sec excerpt from each of the recordings and requiring listeners to match the five excerpts with the five song titles and artists, as in the actual experiment. The vocals in these excerpts did not reveal the titles of the songs, and the 20-sec excerpts did not contain the excerpts used

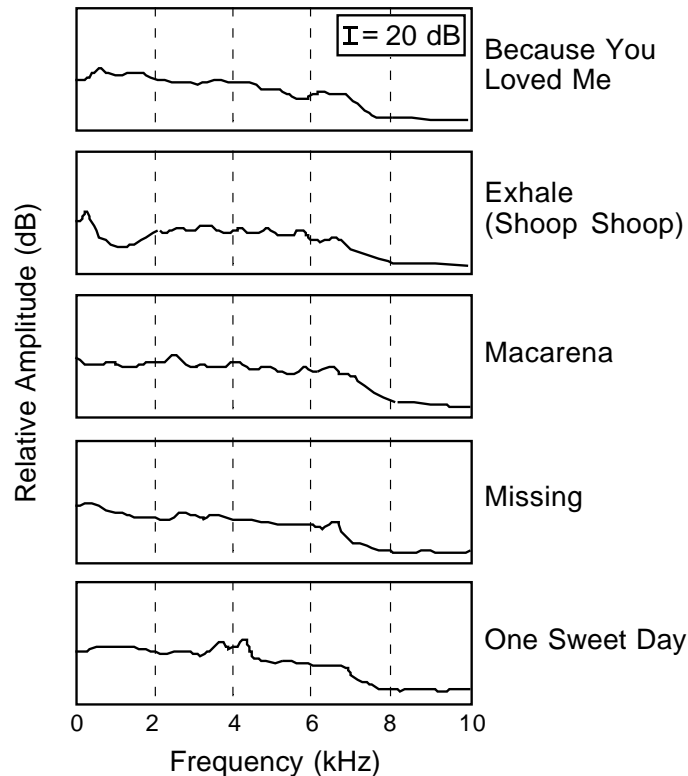


Figure 1. Relative amplitude of frequencies between 0 and 10 kHz in the unfiltered, forward excerpts. Spectra were derived using linear predictive coding (LPC) at 50 msec after the onset of each excerpt.

in the actual experiment. Only listeners who scored 100% were included in the final sample, but all the participants received course credit, even if they failed to meet the inclusion criterion. The listeners were tested individually or in small groups during the screening process. A delay of several minutes between the screening session and the actual experiment prevented the listeners from retaining a representation of the excerpts in working memory.

RESULTS

For each condition, there were 120 ($5 \times 4 \times 3 \times 2 \times 1$) possible response combinations, each of which was equally likely if the listeners were guessing. The average number of correct responses for these 120 possibilities was one. Because the distribution of scores (number correct) based on chance levels of responding was not normal, the data were analyzed with nonparametric tests. Individual listeners were classified according to whether or not they performed better than chance (score > 1 or score ≤ 1). The probability of getting more than one correct response (two, three, or five correct)³ was 31/120 if listeners were guessing. Thus, only about 1 in 4 listeners (i.e., 5.17 out of 20 in each condition) should score better than chance, if listeners as a group were guessing. Figure 2 illustrates the number of listeners who performed above chance separately for each condition. Mean scores for each condition

(provided below the figure) make it clear that dichotomizing the outcome variable did not affect the overall response pattern.

Chi-square goodness-of-fit tests were used separately for each condition, to examine whether the number of listeners with scores greater than 1 exceeded chance levels. Performance was much better than chance in the 200-msec condition [$\chi^2(1, n = 20) = 49.89, p < .001$], with 19 of 20 listeners performing above chance. Group responding remained above chance for the even briefer 100-msec stimuli [$\chi^2(1, n = 20) = 8.87, p < .005$]. Performance was also better than chance in the 100-msec high-pass filtered condition [$\chi^2(1, n = 20) = 15.99, p < .001$], but not in the low-pass filtered or backward conditions. A chi-square test of independence confirmed that the number of listeners performing above chance differed across conditions [$\chi^2(4, N = 100) = 30.29, p < .001$].

Performance in the 200-msec condition was superior to levels observed in the 100-msec condition [$\chi^2(1, n = 40) = 8.53, p < .005$]. This effect was evident for each of the five recordings and implies that successful identification of the recordings required the presence of dynamic information in the frequency spectrum, because the static (steady state) information and the absolute pitch of the excerpts would have been very similar for the 200-

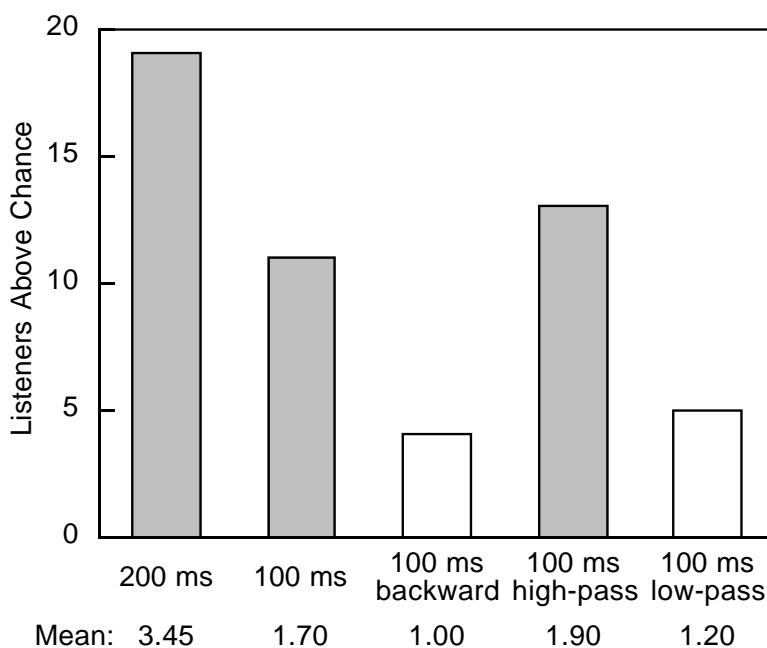


Figure 2. Number of listeners exceeding chance levels (>1 correct response) for each testing condition ($n_s = 20$). Hatched bars indicate conditions in which group performance was significantly better than chance. Mean scores (number of songs identified correctly) are provided below the figure.

and the 100-msec excerpts. This hypothesis was tested directly in the next comparison, which showed that performance was poorer in the backward 100-msec condition than it was in the forward 100-msec condition [$\chi^2(1, n = 40) = 5.23, p < .05$]. This decrement was evident for four of the five songs (all but “One Sweet Day”). Because static spectral information and absolute pitch were *exactly* the same in these two conditions, inferior performance with the backward excerpts provides confirmation of listeners’ reliance on dynamic information in the frequency spectrum.

In the next set of analyses, differences in performance as a function of the presence of low-frequency or high-frequency information were examined. Performance in the high-pass filtered condition was no different from levels observed in the original 100-msec condition; the number of listeners scoring above chance increased for two songs (“Because You Loved Me” and “Missing”), decreased for two songs (“Exhale” and “Macarena”), and remained unchanged for one song (“One Sweet Day”). Significant performance decrements were observed, however, in the low-pass condition, as compared with the original 100-msec and the high-pass conditions [$\chi^2(1, n = 60) = 6.54, p < .05$]; indeed, the low-pass condition had the fewest above-chance listeners for all the songs but one (“One Sweet Day”). Thus, successful identification of the excerpts depended on the presence of high-frequency, but not on low-frequency, spectral information.

To examine the possibility that listeners were relying solely on vocal cues, rather than the timbre of the overall recordings, we examined song-by-song responding for each of the three conditions in which performance was better than chance. In each condition, absolute levels of performance were highest for the excerpt that did *not* contain any vocals (“Macarena”).

DISCUSSION

Our listeners were able to identify recordings of popular songs from excerpts as brief as 0.1 sec, provided that dynamic, high-frequency information from the recordings was present in the excerpts. The observed pattern of findings cannot be attributed to absolute-pitch cues or to recognition of specific voices. Rather, the spectra in Figure 1 show that the excerpt with the highest levels of performance (“Macarena,” no vocals) had the densest concentration of energy between 1000 and 8000 Hz, which may have contributed to its relative distinctiveness. Listeners may also have been more familiar with “Macarena” than with the other recordings.

Listeners’ ability to identify complex musical stimuli from a minimal amount of perceptual information is similar to their abilities with speech. For example, 10-msec vowels can be identified reliably (Robinson & Patterson, 1995b; Suen & Beddoes, 1972), as can individual voices from vowel samples as brief as 25 msec (Compton, 1963).

When respondents are asked to identify famous voices from a set of 60 different voices, performance starts to exceed chance levels with samples of 250 msec (Schweinberger, Herholz, & Sommer, 1997). The capacity to identify speech stimuli from a minimal amount of information appears to be general enough to extend to other auditory domains—such as music—where the adaptive significance is much less obvious (Roederer, 1984). Although our findings do *not* imply that recognition of popular songs typically occurs in 100 msec, they provide unequivocal evidence that excerpts this brief contain information that can be used for identification. Moreover, our results reveal that such information is timbral in nature and independent of absolute-pitch cues or changes in pitch and tone durations.

Our results extend those of Levitin (1994; Levitin & Cook, 1996; see, also, A. R. Halpern, 1989), who reported that memory representations for popular recordings contain absolute information about pitch and tempo. With very brief presentations, however, identification of recordings is primarily a function of timbre rather than of absolute pitch or tempo. Although information about tempo was unavailable in our brief excerpts, pitch is perceptible from tones as brief as 10 msec (Warren et al., 1991). Nonetheless, performance was at chance when our 100-msec excerpts were played backward or low-pass filtered. Because both manipulations would have dramatically disrupted attributes that are critical to timbre (dynamic and static information, respectively) while having little impact on perceived pitch, it appears that timbre is more important than absolute pitch for identifying popular recordings from very brief excerpts. This finding converges with others involving music and speech, which show that timbre (i.e., a specific musical instrument or vowel) is better identified than is pitch when stimuli are extremely brief (Robinson & Patterson, 1995a, 1995b).

The listeners' dependence on timbre rather than on absolute pitch in the present investigation could stem from (1) the importance of timbral cues (i.e., voice qualities other than pitch) in speech, (2) the relative unimportance of absolute, as compared with relative pitch in music listening, or (3) both of these factors. Although voices vary in pitch as well as in timbre, differences in pitch (i.e., average fundamental frequency) between talkers of the same sex are relatively small; in a group of 12 women tested by Miller (1983), the *SD* was less than 2.5 semitones. Nonetheless, most people can rapidly identify many different female (or male) voices, despite similarities in pitch. Because of the multidimensional nature of timbre, voice-quality cues are more distinctive than those based on pitch. Extensive experience discriminating voices on the basis of timbre could, in turn, influence processing in the musical domain.

We also know that the ability to perceive musical pitch in an absolute manner is limited to a relatively small proportion of the population (approximately 1 in 10,000; see Takeuchi & Hulse, 1993). Absolute-pitch possessors can

identify a note by name (e.g., C, F#, etc.) when it is played in isolation (an ability that is qualitatively different than remembering the pitch of a recording). Because such absolute-identification abilities tend to be automatic, they can interfere with relational processing strategies that are more relevant to music listening (Miyazaki, 1993). Moreover, other evidence implies that absolute-pitch processing is actually a relatively primitive auditory strategy. For example, elevated prevalence levels have been reported among mentally retarded individuals, and absolute- rather than relative-pitch processing is the norm for nonhuman vertebrates (Ward & Burns, 1982).

At present, it is unclear why the portion of the spectrum above 1000 Hz is more important for song recognition than the portion below 1000 Hz. The high-pass filtered excerpts differed quantitatively from the low-pass excerpts (e.g., they had more spectral information, because most of the harmonics in the excerpts were above 1000 Hz; see Figure 1), and qualitative differences may also have played a role (e.g., the high frequencies may have been more distinctive). It is also possible that high-frequency timbral information is either perceived or encoded in memory with better detail, as compared with low-frequency information. Interestingly, Compton (1963) used speech samples that were low-pass and high-pass filtered much like our musical excerpts (cutoff frequency of 1020 Hz, rather than 1000 Hz) and reported results similar to ours. His respondents, who were asked to identify the talker, showed marked deficits in performance for low-pass filtered samples, but not for high-pass samples.

Performance levels in the present study were undoubtedly inflated by two factors: (1) allowing the excerpts to be heard repeatedly, which would have enhanced perceptual fluency for the repeated items (Jacoby & Dallas, 1981), and (2) the pretest session, which would have primed listeners' memories of the songs. Indeed, exposure to the pretest excerpts could have allowed above-chance levels of performance to emerge even among listeners who had limited familiarity with the songs prior to the experiment. These listeners may have met the pretest inclusion criterion by recognizing one or two of the singers, by a process of elimination, by luck, or by a combination of these factors, all of which may have influenced performance in the subsequent test session as well. Because the listeners received course credit even if they failed to meet the inclusion criterion (which excused them from the test session), however, it is unlikely that they falsely claimed familiarity with the tunes. Moreover, the time frame of the experiment prevented the listeners from retaining one or more of the excerpts in working memory. By definition, then, the task required the listeners to rely primarily on representations in *long-term* memory of greater or lesser permanence. For example, such representations would be relatively permanent (or consolidated) for listeners with extensive familiarity with the tunes, but more temporary (or less consolidated) for other listeners, being retrievable only for the length of the experiment. Regard-

less, the results make it clear that (1) the brief stimuli contained information that listeners could compare with their representations of the recordings and (2) this information was primarily timbral in nature. Future research could examine the generalizability of these findings with a broader selection of excerpts and a less constrained task. For example, different results might be obtained with recordings of soft-rock tunes or orchestral symphonies or with individual recordings in which the overall timbre is less distinctive. Representations that vary in degree of consolidation could also differ in the way timbre is encoded.

It is important to clarify that absolute attributes in memory representations for popular songs would be stored in combination with the relational information that defines the songs. Adult, child, and infant listeners recognize similarities between sequences of pure tones presented in transposition (different absolute pitch, same pitch and temporal relations; Schellenberg & Trehub, 1996a, 1996b). It is safe to assume, then, that our listeners would recognize previously unheard versions of, say, "Macarena," performed by different singers, on different instruments, and in a key and tempo different from the original recording. Nonetheless, our results provide converging evidence that memory representations for complex auditory stimuli contain information about the absolute properties of the stimuli, in addition to more meaningful information abstracted from the relations between stimulus components. Indeed, in contexts with an extremely limited amount of information, listeners may rely primarily on the sound quality of the stimuli for successful identification and recognition.

REFERENCES

- COMPTON, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers, aurally. *Journal of the Acoustical Society of America*, **35**, 1748-1752.
- DOWLING, W. J., & HARWOOD, D. L. (1986). *Music cognition*. San Diego: Academic Press.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, **61**, 1270-1277.
- HAJDA, J. M., KENDALL, R. A., CARTERETTE, E. C., & HARSHBERGER, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253-306). Hove, U.K.: Psychology Press.
- HALPERN, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, **17**, 572-581.
- HALPERN, D. L., BLAKE, R., & HILLENBRAND, J. (1986). Psychoacoustics of a chilling sound. *Perception & Psychophysics*, **39**, 77-80.
- IVERSON, P., & KRUMHANSL, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, **94**, 2595-2603.
- JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.
- JONES, M. R., & YEE, W. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69-112). Oxford: Oxford University Press, Clarendon Press.
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- LEVITIN, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, **56**, 414-423.
- LEVITIN, D. J., & COOK, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, **58**, 927-935.
- MCADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, **58**, 177-192.
- MILLER, C. L. (1983). Developmental changes in male/female classification by infants. *Infant Behavior & Development*, **6**, 313-330.
- MIYAZAKI, K. (1993). Absolute pitch as an inability: Identification of musical intervals in a tonal context. *Music Perception*, **11**, 55-72.
- NYGAARD, L. C., & PISONI, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, **60**, 355-376.
- NYGAARD, L. C., SOMMERS, M. S., & PISONI, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.
- PALMERI, T. J., GOLDINGER, S. D., & PISONI, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 309-328.
- PITT, M. A., & CROWDER, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 728-738.
- ROBINSON, K., & PATTERSON, R. D. (1995a). The duration required to identify an instrument, the octave, or the pitch chroma of a musical note. *Music Perception*, **13**, 1-15.
- ROBINSON, K., & PATTERSON, R. D. (1995b). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America*, **98**, 1858-1865.
- ROEDERER, J. G. (1984). The search for a survival value of music. *Music Perception*, **1**, 350-356.
- SCHELLENBERG, E. G., & TREHUB, S. E. (1996a). Children's discrimination of melodic intervals. *Developmental Psychology*, **32**, 1039-1050.
- SCHELLENBERG, E. G., & TREHUB, S. E. (1996b). Natural intervals in music: A perspective from infant listeners. *Psychological Science*, **7**, 272-277.
- SCHWEINBERGER, S. R., HERHOLZ, A., & SOMMER, W. (1997). Recognizing familiar voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, & Hearing Research*, **40**, 453-463.
- SUEN, C. Y., & BEDDOES, M. P. (1972). Discrimination of vowel sounds of very short duration. *Perception & Psychophysics*, **11**, 417-419.
- TAKEUCHI, A. H., & HULSE, S. H. (1993). Absolute pitch. *Psychological Bulletin*, **113**, 345-361.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.
- VAN LANCKER, D., KREIMAN, J., & EMMOREY, K. (1985). Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices. *Journal of Phonetics*, **13**, 19-38.
- WARD, W. D., & BURNS, E. M. (1982). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 431-451). New York: Academic Press.
- WARREN, R. M., GARDNER, D. A., BRUBAKER, B. S., & BASHFORD, J. A. (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception*, **8**, 277-290.

NOTES

1. Although the recording of "Macarena" contained vocals, the excerpt did not.
2. Filtering is actually *gradual* rather than absolute; some frequencies on the unwanted side of the cutoff point are present with monotonically decreasing amplitude (D. J. Levitin, personal communication, August 1998).
3. A score of four correct was impossible: In the present matching task, one error ensured another error.

(Manuscript received June 23, 1998;
revision accepted for publication December 11, 1998.)