Evidence for two pitch encoding mechanisms using a selective auditory training paradigm

NICOLAS GRIMAULT
UPRESA CNRS 5020, Pavillon Université Hôpital E. Herriot, Lyon, France
and Groupe Entendre GIPA2, Pontchartrain, France

CHRISTOPHE MICHEYL
UPRESA CNRS 5020, Pavillon Université Hôpital E. Herriot, Lyon, France

ROBERT P. CARLYON
MRC Cognition and Brain Sciences Unit, Cambridge, England

and

LIONEL COLLET
UPRESA CNRS 5020, Pavillon Université Hôpital E. Herriot, Lyon, France

The neural mechanisms underlying the perception of pitch, a sensory attribute of paramount importance in hearing, have been a matter of debate for over a century. A question currently at the heart of the debate is whether the pitch of all harmonic complex tones can be determined by the auditory system's using a single mechanism, or whether two different neural mechanisms are involved, depending on the stimulus conditions. When the harmonics are widely spaced, as is the case at high fundamental frequencies ($F0$s), and/or when the frequencies of the harmonics are low, the frequency components of the sound fall in different peripheral auditory channels and are then "resolved" by the peripheral auditory system. In contrast, at low $F0$s, or when the harmonics are high in frequency, several harmonics interact within the passbands of the same auditory filters, being thus "unresolved" by the peripheral auditory system. The idea that more than one mechanism mediates the encoding of pitch depending on the resolvability status of the harmonics was investigated here by testing for transfer of learning in $F0$ discrimination between different stimulus conditions involving either resolved or unresolved harmonics after specific training in one of these conditions. The results, which show some resolvability-specificity of $F0$-discrimination learning, support the hypothesis that two different underlying mechanisms mediate the encoding of the $F0$ of resolved and unresolved harmonics.

Harmonic complex tones, such as musical tones and vowels, generally elicit a strong pitch sensation that corresponds approximately to their fundamental frequency ($F0$). This pitch plays a role of paramount importance in hearing: It conveys melody in music, prosody in speech, and plays an essential part in the perceptual analysis of complex auditory scenes (Hartmann, 1988). The neural mechanisms underlying pitch perception have been debated for over a century (Schouten, Ritsma, & Cardozo, 1962; Von Helmholtz, 1863). At the center of this debate is the question of whether a single neural mechanism can account for the perception of the pitch of all harmonic tones (Cariani &

Delgutte, 1996a, 1996b; Meddis & Hewitt, 1991a, 1991b; Meddis & O'Mard, 1997), or whether more than one mechanism is involved, depending on the stimulation conditions (Carlyon & Shackleton, 1994; Ragot & Crottaz, 1998; Shackleton & Carlyon, 1994; Steinschneider, Reser, Fishman, Schroeder, & Arezzo, 1998). This question is inspired by the fact that the cochlea acts like a bank of parallel bandpass filters and has a finite frequency resolving power, which decreases with increasing frequency. Thus, when the harmonics are widely spaced, as is the case at high $F0$s, and/or when the frequencies of the harmonics are low, the frequency components of the sound fall in different peripheral auditory channels and are then "resolved" by the peripheral auditory system. In this situation, because no single peripheral auditory channel contains unambiguous information about the $F0$ of the sound, the central auditory system must combine the outputs of different auditory channels to derive the pitch. In contrast, when the $F0$ is low, the frequency of the harmonics is high, or both, the components of the sound are "unresolved" by the auditory periphery. Several of them fall within the passband of a single auditory filter. In this situation, the auditory system

can retrieve the pitch by taking advantage of the fact that the auditory filter outputs fluctuate at a rate equal to the $F0$.

Several experimental results argue for the hypothesis that the encoding of the $F0$ of resolved and unresolved harmonics is mediated by two different mechanisms. In particular, it has been shown that the ability to compare the $F0$ of two groups of harmonics filtered in different frequency regions and presented simultaneously is better when both groups are either resolved or unresolved than when one is resolved and the other unresolved (Carlyon & Shackleton, 1994). Furthermore, difference limens (DLs) for $F0$ (DL $F0$s) vary in a qualitatively different way with changes in stimulus duration (Plack & Carlyon, 1995). Finally, another experimental result that has not yet been accounted for by a single-mechanism model of pitch extraction is that DL $F0$s are significantly larger for unresolved than for resolved harmonics (Carlyon, 1998).

On the other hand, it has been proposed that both resolved and unresolved harmonics could be embraced by a single pitch mechanism (Meddis & Hewitt, 1991a, 1991b; Meddis & O'Mard, 1997; Patterson et al., 1991; Slaney & Lyon, 1990). Schematically, this mechanism amounts to the computation by the central nervous system of a summed autocorrelogram, which is obtained by pooling the autocorrelation functions of neural activity within the different peripheral auditory channels. The resulting autocorrelogram exhibits peaks, the largest of which corresponds in most cases to the perceived pitch of the sound. According to Meddis and O'Mard, this unitary mechanism can account for the perceived pitch of both resolved and unresolved harmonics. However, Carlyon (1998) challenged this conclusion, mainly by demonstrating that the results of Meddis and O'Mard could be explained in terms of frequency region, rather than in terms of resolvability per se. Thus, the question of the existence of a single or of two different mechanisms for pitch perception is still posed.

In the present study, this question was addressed by using a transfer-of-learning approach similar to that used in several earlier studies in the visual and somatosensory modalities (Ahissar & Hochstein, 1993, 1996, 1997; Ball & Sekuler, 1982, 1987; Fiorentini & Berardi, 1981a, 1981b; Karni & Sagi, 1990, 1991, 1993; Polat & Sagi, 1994; Sathian & Zangaladze, 1997a, 1997b; Schoups & Orban, 1996; Schoups, Vogels, & Orban, 1995; Shiu & Pashler, 1992; Spengler et al., 1997) and in the auditory modality (Demany, 1986; Irvine, Martin, Klimkeit, & Smith, 2000; Wright, Buonomano, Mahncke, & Merzenich, 1997). The general assumption behind this approach relies on the idea that if perceptual performance on two different tasks or with two different stimulus types is underlain by different neural processes or sets, exclusive practice in one task or with one stimulus type will train, specifically, the neural processes or sets engaged by this task or stimulus type. Consequently, no improvement in performance in the other task or with the other stimulus type should be observed after training. Thus, the general assumption is that neural plasticity selectively affects the units that are active during the performance. Under this assumption, if the neural mechanism underlying pitch perception with resolved

harmonics is distinct from that underlying pitch perception with unresolved harmonics, subjects trained in pitch-discrimination with exclusively resolved harmonics should exhibit little or no performance improvement with unresolved harmonics, and vice versa. The primary objective of the present study was to test further the dual-pitch encoding theory.

Although numerous investigations involving trained subjects have been devoted to $F0$ discrimination, to our knowledge, none of these have provided explicit information either on the extent and time course of improvements in this task with practice or on the generalization of these improvements across $F0$s and frequency regions. The only data on pitch discrimination learning that are currently available in the literature have been obtained in studies involving pure tones, rather than harmonic complex tones (Campbell & Small, 1963; Demany, 1986; Harris, 1952; Spiegel & Watson, 1981; Watson, 1980; Watson, Kelly, & Wroton, 1976; Wyatt, 1945). Since the $F0$ discrimination of harmonic complex tones cannot be reduced simply to the frequency discrimination of pure tones (Moore & Glasberg, 1990, 1991), the characteristics of $F0$ discrimination learning may well differ somewhat from those of pure-tone frequency-discrimination learning. Consequently, besides testing further for the existence of two pitch-encoding mechanisms, a secondary objective of the present study was to document $F0$ discrimination learning.

METHOD

Subjects

Twelve subjects took part in this experiment. The subjects ranged in age between 19 and 28 years ($M = 23.83$, $SD = 3.16$). They all had normal hearing in both ears and absolute pure-tone thresholds at or below 15 dB HL at octave frequencies from 250 to 8000 Hz (American National Standard Institute, 1969). None had prior experience in psychoacoustic tasks. They all were paid an hourly wage for their participation, and all completed the experiment.

Stimuli

The stimuli consisted of harmonic complex tones having durations of 200 msec, including 50-msec cosine ramps. The harmonics were all added in sine (0°) phase and were bandpass-filtered digitally with the use of a filter with a flat top and 48 dB/octave slopes; components to which an attenuation larger than 48 dB would have to be applied were omitted. Following Shackleton and Carlyon (1994) and Carlyon and Shackleton (1994), we used two nominal $F0$s (88 and 250 Hz) and three filtering regions (a low region with lower and upper corner frequencies of 125 and 625 Hz, a mid region of 1375–1875 Hz, and a high region of 3900–5400 Hz), leading to six different $F0$-region combinations. According to Shackleton and Carlyon and Carlyon and Shackleton, the harmonics of both 88 and 250 Hz are resolved in the low region and unresolved in the high region; in the mid region, the resolution status of the harmonics depends on the $F0$ used: They are unresolved at 88 Hz but resolved at 250 Hz. Thus, using these stimuli, the influence of differences in resolvability can be distinguished from that of differences in $F0$ or differences in frequency region. For convenience, the different $F0$-region combinations will hereafter be referred to as LOW88R, LOW250R, MID 88U, MID250R, HIGH88U, and HIGH250U.

The presentation level was determined separately for each subject. All harmonic complex tones were set at 40 dB above the absolute threshold of a reference harmonic complex tone filtered in the corresponding frequency region and having a $F0$ equal to the nominal $F0$

used in the condition being tested. Because the standard deviation of the absolute thresholds was only about 4 dB, the physical level of the harmonic complex tones did not vary widely across subjects.

A pink-noise background with a 3 dB/octave slope was presented continuously throughout all measurements. The overall noise level was set individually 20 dB above the absolute detection threshold for each subject. The purpose of this noise was to prevent the perception by the subjects of combination tones generated by the ear, which might have introduced a bias in some of the test conditions. In particular, in the MID88U, HIGH88U, or HIGH250U conditions, although the physical components of the harmonic complex tones were unresolved, combination tones corresponding to subharmonics of these components or to the $F0$ itself, and falling in a region where the auditory filters were narrower, might have provided spectral cues as to the $F0$ differences between the stimuli being compared. Vliegen and Oxenham (1999) estimated that the levels of combination tones elicited by harmonic complex tones having a level of 70 dB SPL, with $F0$s around 100–200 Hz and harmonics filtered between 2 and 8 kHz did not exceed 30 dB SL at the $F0$ and 20 dB SL at harmonic frequencies. On this basis, they established that a SPL of 25 dB at 1 kHz for the pink noise would be sufficient to ensure that any resolvable distortion product was at least 10 dB below masked threshold. Previous studies concerned with the influence of resolvability on DL $F0$s have in general used a pink-noise background with a level of 15 dB SPL at 1 kHz and harmonic complex tones with a level of 45 dB SPL per component (Carlyon & Shackleton, 1994; Gockel, Carlyon, & Micheyl, 1999; Micheyl & Carlyon, 1998; Plack & Carlyon, 1995; Shackleton & Carlyon, 1994). In the present study, the physical level of the pink noise (corresponding to 20 dB above the absolute detection threshold for each subject) varied little across subjects and was equal on average to about 20 dB SPL at 1 kHz, which corresponded roughly to an overall level of 52 dB SPL. Given that the resulting signal-to-noise ratio was even smaller than in the studies cited above, it appears very unlikely that combination tones have been heard by the subjects in the present experiment.

Stimulus Generation

The harmonic complex tone signals were generated digitally in the time domain on a PC150 computer and output using a 16-bit digital-to-analog converter (TDT DA1) at a sampling rate of 44.1 kHz. The pink-noise masker was generated digitally, recorded on an audio compact disc (CD), and played out continuously throughout the measurements with the use of the CD-Rom drive (Goldstar CRD 8322B) of another Pentium computer. The signals and noise were individually low-pass filtered at 15 kHz (TDT FT6-2, attenuation more than 60 dB at 1.15 times the corner frequency) and attenuated (TDT PA4). Finally, they were summed (TDT SM3) and led to the right or left earpiece of a Sennheiser HD465 headphone mounted in a 25125 cushion via a headphone preamplifier (TDT HBC). Stimuli were monitored on all sessions using an HP3561A signal analyzer. The subjects were comfortably seated in a sound-treated booth.

Procedure

DLs for $F0$ (DL $F0$s) were measured with a three-interval, two-alternative, forced-choice (3I-2AFC) procedure with a two-down one-up adaptive rule, which tracked the 70.7% correct point on the psychometric function (Levitt, 1971). On each trial, the subject was presented with three successive harmonic complex tones, the second or third of which (with a .5 probability) had a higher $F0$ than the two others. The three harmonic complex tones were separated by 200-msec silent intervals. The subject's task was to indicate, after hearing the three tones, which of the last two sounded different from the others—in other words, to "tell the odd one out." Responses were given by the subject by using the numeric pad of a computer keyboard, with no time limitation. The next trial was initiated after the subject had given a response. Visual feedback, when present, was provided in the form of a change in screen color, which otherwise remained black. Following the subject's response, green indicated a correct response and red an incorrect response.

The $F0$ of two standard harmonic complex tones remained constant at 88 or 250 Hz throughout a run; the $F0$ of the deviant harmonic complex tone was equal to the $F0$ of the standard harmonic complex tones incremented by a variable quantity $\Delta F0$. At the beginning of a run, $\Delta F0$ was set to 40% of the nominal $F0$. It was decreased by a factor of 2 after two consecutive correct responses and was increased by the same amount after any incorrect response until the fourth reversal in the direction of the change in $\Delta F0$; thereafter, a factor of $\sqrt{2}$ was used. The run stopped after 16 reversals, which corresponded to a variable number of trials—in general, around 60. Each run took approximately 1–2 min to complete. The DL $F0$ was then estimated as the geometric mean of the $\Delta F0$s over the last 12 reversals and was expressed as a percentage of the nominal $F0$.

Experimental Design

All 12 subjects first took part in a preliminary test session during which they familiarized themselves with the test procedure and stimuli. During this familiarization session (S0), they performed three runs of the DL $F0$ measurement procedure in each of the six stimulus conditions, in random order. No visual feedback was provided to the subjects during this session. Later, on a different day of the same week, their baseline performance was measured in the first test session (S1). During this session, they performed five runs in each of the six conditions. Again, no visual feedback was provided to the subjects at this point.

In the following week, 8 subjects, who had been chosen at random out of the 12, started training, and the 4 remaining, serving as controls, did not. These 8 subjects who started training were divided into two groups comprising 4 subjects each. One group (TMID88U) received training in the MID88U condition, and the other group (TMID250R) was trained in the MID250R condition. In each experimental group, half of the subjects were tested and trained in the right ear and the other half in the left ear. The training program consisted of 2-h sessions, 3 days per week, for 4 consecutive weeks, corresponding to a total of 24 h. During these training sessions, the subjects had to complete 30 runs. Trial-by-trial visual feedback was provided.

Four weeks after the initial evaluation session (S1)—that is, on the week after the last training session for trained subjects—all subjects took part in a second evaluation session (S2). Finally, 4 weeks later—that is, 5 weeks after the last training session for trained subjects—all subjects took part in a final evaluation session (S3) aimed at checking whether they had retained the same level of performance as that measured on S2. As in S1, during these final two evaluation sessions, five runs in each of the six different $F0$-region conditions were collected, and no visual feedback was provided to the subjects.

RESULTS

Figure 1 represents the mean DL $F0$s obtained in the two training groups (TMID88U and TMID250R) on the preliminary (S0), pretraining (S1), training (T1–T12), and posttraining sessions. Each data point represents the geometric mean of the mean DL $F0$s measured in the 4 subjects of a given group on a given test or training session. For each training session, the mean DL $F0$ for each subject was itself computed as the geometric mean of 30 estimates corresponding to the 30 runs of the adaptive procedure that a subject had to perform. For each test session, individual mean DL $F0$s were computed as the average of the five collected thresholds. The geometric standard errors around the geometric mean DL $F0$ ratios were computed with the following formula (Micheyl & Carlyon, 1998):
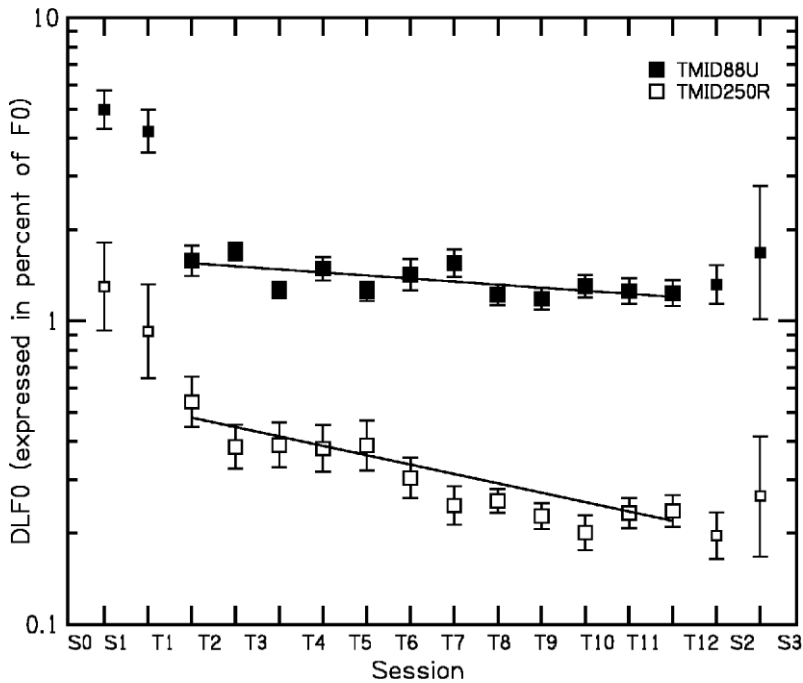
Figure 1. DL $F0$s measured on the different test (small symbols) and training sessions (large symbols) in the two training groups. The filled squares correspond to TMID88U data, empty squares to TMID250R data. The error bars refer to the intersubject variability. They represent the geometric standard error around the geometric mean DL $F0$ across subjects, computed as the standard error of the log-transformed geometric mean DL $F0$s of the 4 subjects of each group. The regression lines fitted to the data are superimposed (see text for details).

$$1 + \sqrt{\left(10^{\frac{\sigma_{s1}}{\sqrt{4}}} - 1\right)^2 + \left(10^{\frac{\sigma_{s2}}{\sqrt{4}}} - 1\right)^2},$$

where σ_{s1} represents the standard deviation of the four log-transformed DL $F0$ estimates obtained for the 4 subjects of each group in the pretraining session (S1), and σ_{s2} represents the standard deviation of the four log-transformed DL $F0$ estimates obtained for the 4 subjects of each group in the first posttraining session (S2). Basically, this formula was derived from the idea that the standard error of the ratio between two means is equal to the square root of the sum of the squares of the standard errors of the two means.

Before looking into more detail at the differences in DL $F0$s between the pre- and posttraining sessions in the two groups, it is worth considering how performance improved during the training sessions (T1–T12). The graph in Figure 1 indicates a protracted decrease in thresholds over the 12 training sessions in the TMID250R group. For the TMID88U group, after a marked decrease in thresholds between S1 and T1, little improvement was observed across training sessions. These results were submitted to an analysis of variance (ANOVA) with the log-transformed DL $F0$s (always expressed in percents of the nominal $F0$) measured during the training sessions as the dependent variable, the training session number ("session") as the re-

peated measures factor, and the training group ("group") as the independent measures factor. The results of the ANOVA reveal the existence of a significant difference in thresholds across training sessions [$F(11,66) = 4.25, p < .001$]. Throughout the article, all main effects and interactions not mentioned were not statistically significant. Since the session $\times$ group interaction only just failed to reach the .05 threshold [$F(11,66) = 1.91, p = .054$], we decided to check whether significant differences across sessions were effectively present in both groups; session was found to have had a significant effect in the TMID250R group [$F(11,33) = 4.55, p < .001$] but not in the TMID88U group [$F(11,33) = 1.13, p = .371$]. The regression lines fitted to the log-transformed training data of the two groups in a least-squares algorithm are also shown in Figure 1. The slopes of these lines were compared with the use of a t test as described in Edwards (1973). They proved to be significantly different [$t(20) = -4.30$; slope $= -0.03$ with $R^2 = .99$ for TMID250R and slope $= -0.01$ with $R^2 = .93$ for TMID88U].

In regard to changes in thresholds between the pre- and posttraining sessions, Figure 2 shows the mean variation in DL $F0$ measured between S1 and S2 in the six conditions for the three experimental groups. These variations were obtained by dividing the geometric mean DL $F0$s measured across subjects and the runs on S1 by those measured on
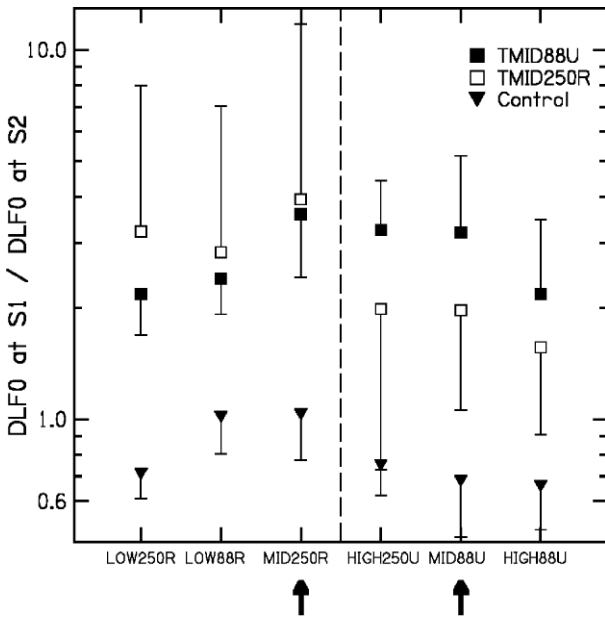
Figure 2. Variations in DL F0s between S1 and S2 in the different F0-region conditions. Each datapoint was obtained by dividing the geometric mean DL F0 obtained by the subjects of a given group in a given F0-region condition on S1 by the corresponding geometric mean DL F0 on S2. The error bars refer to the geometric standard errors around the geometric mean ratio. Open squares correspond to data from the group trained with resolved harmonics. Filled squares correspond to data from the group trained with unresolved harmonics. Filled triangles correspond to data from the control group. The F0-region conditions indicated on the abscissa are sorted by order of decreasing harmonic resolvability from left to right. The three rightmost conditions involved unresolved harmonics, whereas the three leftmost involved resolved harmonics; the separation between the resolved and unresolved conditions is materialized by a vertical dashed line. The two training conditions are indicated by arrows.

S2. The fact that most datapoints from the trained groups were above 1 indicates, at first glance, a general improvement in performance for these two groups. In contrast, most datapoints from the control group were slightly below 1, indicating either a slight decrease or no substantial change in performance between S1 and S2. These results first were submitted to a four-way ANOVA with log-transformed DL F0 as the dependent variable, session number (S1 vs. S2), frequency region (low, mid, and high), and F0 (88 and 250 Hz) as the repeated-measures factors, and group (TMID88U, TMID250R, and control) as the independent measures factor. The results of this analysis reveal a significant effect of session [$F(1,9) = 20.49, p = .001$], as well as an interaction between session and group [$F(2,9) = 10.26, p = .005$]. Frequency region and F0 both had a significant main effect [$F(2,18) = 160.00, p < .001$, and $F(1,9) = 225.29, p < .001$] and interacted significantly [$F(2,18) = 16.29, p < .001$]. When the analysis was performed on the data of the two trained groups alone (leaving aside those of the control group), a significant difference between S1 and S2 was still observed [$F(1,6) = 31.34, p = .001$]. In the

control group, no significant difference was observed between the two test sessions [$F(1,3) = 2.48, p = .213$].

By looking in a more detailed way at the results for the two trained groups, it can also be seen in Figure 2 that although in the three resolved conditions (LOW88R, LOW250R, and MID250R) the datapoints of the subjects from the TMID250R group lay above those of the TMID 88U group, the opposite pattern of results is apparent in the three unresolved conditions (MID88U, HIGH88U, and HIGH250U). Figure 3 shows the mean threshold variations in the *untrained* resolved and unresolved conditions in the three experimental groups. Excluding the data obtained in the trained conditions allows the elimination of a possible bias in the interpretation, which is caused by the expected fact that the improvements were largest in the trained conditions. In order to check the effect of resolvability per se, these data were submitted to a post hoc analysis contrasting the log-transformed DL F0s obtained in the two training groups (TMID88U vs. TMID250R) in the two conditions (resolved and unresolved). More precisely, this contrast analysis was realized by setting coefficients in the contrast matrix to -1 for the untrained–resolved conditions, to 1 for the untrained–unresolved conditions, and to 0 for trained conditions. The results of the contrast analysis indicate that the subjects trained in the resolved (MID250R) condition showed larger improvements in the two other resolved conditions (LOW88R and LOW250R) than did the subjects trained in the unresolved (MID88U) condition [$F(2,6) = 8.45, p < .05$], and vice versa [$F(2,6) = 27.59, p < .001$].
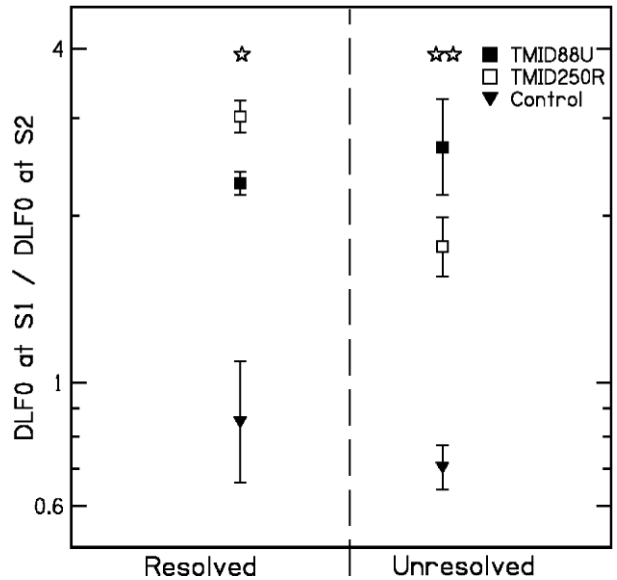

Figure 3. Mean relative changes in DL F0s between S1 and S2 in the two untrained resolved and the two untrained unresolved conditions. The stars indicate the statistical significance (one star for $p < .05$ and two stars for $p < .001$) in the contrast analysis (see text for details). The error bars refer to the intercondition variability. They represent the geometric standard error around the geometric mean DL F0 across untrained resolved or untrained unresolved conditions.
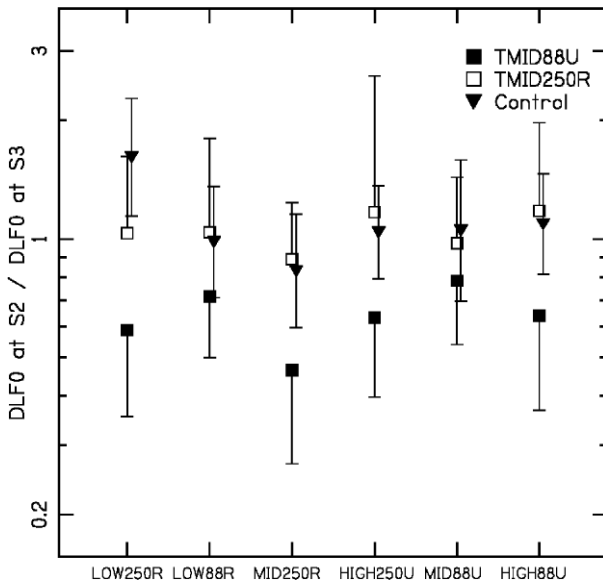
Figure 4. Relative changes in DL F0s between S2 and S3 in each of the six F0-region conditions. This figure follows the same conventions as used for Figure 2.

Finally, when one considers how thresholds varied between the two posttraining test sessions (S2 and S3), as is shown in the graph of Figure 4, thresholds remained overall approximately constant in the TMID250R and the control groups, for which most of the S2/S3 threshold ratios were close to unity. In the TMID88U group, the mean threshold ratios were all below 1, indicating at first glance, a slight decrease in performance between S2 and S3. The results of an ANOVA (session $\times$ group) on the log-transformed DL F0s of all three groups failed to reveal any statistically significant difference between the data of S2 and S3 [$F(1,9) = 2.10, p = .181$]. However, a significant interaction was observed between the session and group [$F(2,9) = 4.79, p < .05$]. In order to obtain further insight into this outcome, we analyzed the data of the two training groups separately from those of the control group. The results of these separate analyses failed to show any significant change from S2 to S3 in any of the experimental groups taken individually.

DISCUSSION

Comparison of Thresholds With Previous Data

The posttraining DL F0s obtained in this study are consistent with those measured in earlier studies using similar stimuli and trained subjects (e.g., Carlyon & Shackleton, 1994; Gockel et al., 1999; Micheyl & Carlyon, 1998; Plack & Carlyon, 1995; Shackleton & Carlyon, 1994). Earlier studies found respective DL F0s of about 0.5% and 2%–5% for resolved and unresolved harmonics, which compare well with those measured in our posttraining ses-

sion—namely, around 0.3% in the MID250R condition and around 2% in the MID88U condition.

Learning Curves

Although many earlier studies have been devoted to pitch discrimination and have involved trained subjects with the implicit assumption that performance in this task improved with practice, this is to our knowledge the first study in which learning in pitch discrimination with harmonic complex tones is explicitly documented. A main result evidenced by this study corresponds to the fact that DL F0s improved significantly over the training period in the TMID250R group, but not in the TMID88U group. A tentative interpretation of this observation is that most, if not all, of the long-term improvement in F0-discrimination performance in the resolved condition was mediated by an improvement in the encoding of the individual frequencies of the harmonics (Faulkner, 1985), which, when the harmonics were unresolved, as was the case for the TMID88U group, cannot take place. Although the subjects could in theory rely on changes in the frequency of a single harmonic to perform the F0-discrimination task when these harmonics are resolved, there exists fairly strong evidence in the literature that they actually compare the pitch sensations elicited by the whole harmonic complex tones rather than the frequencies of individual harmonics (Moore & Glasberg, 1990, 1991). However, since the encoding of the virtual pitch of resolved harmonics presumably involves as a first stage the encoding of the individual frequency of these harmonics (Goldstein, 1973; Terhardt, 1972; Wightman, 1973), it is conceivable that all or part of the improvements in F0-discrimination performance that were observed in the resolved conditions of this study were mediated by improvements in the accuracy with which the frequencies of the different components of the harmonic complex tones were encoded in the auditory system. If this were the case, parallels should be observed between the learning curves obtained here for resolved harmonics and those obtained for frequency discrimination for pure tones in previous studies. Several previous investigations have shown that the ability to detect small differences in frequency between successive pure tones improves markedly with practice over several hours, days, or even months, depending on the intensity of the training (Campbell & Small, 1963; Demany, 1986; Harris, 1952; Moore, 1973; Turner & Nelson, 1982; Wyatt, 1945). Results obtained by Campbell and Small (1963) indicated that frequency-discrimination thresholds (FDTs) measured for 800-msec long tones at 2 kHz using a 2AFC procedure were divided by a factor of about 2 on average after 6 consecutive days of training, corresponding to about 3,900 trials. In their study, an asymptote was reached after about 4 h of training, which corresponded to approximately 2,600 trials. Watson (1980) later reported that between 1,600 and 4,200 trials were necessary for frequency discrimination performance to reach asymptote. In the present study, the subjects were trained for 24 h, in 12 two-hour long sessions, and performed 30 runs per session, leading

to approximately 21,600 trials. Despite this very large number of training trials, performance did not clearly reach asymptote near the end of the training period in the resolved harmonics condition. This suggests that there is something more to the learning of pitch discrimination with resolved harmonic complexes than to the learning of pitch discrimination with pure tones.

Although the difference observed in the training data of the TMID88U and TMID250R is generally consistent with the hypothesis that pitch discrimination is mediated by several different mechanisms for resolved and unresolved harmonics, it cannot be taken as a demonstration for this hypothesis. Indeed, the observed difference in the learning curves in the two training conditions could be related to the fact that the nominal $F0$s were different in these two conditions. It might be the case that less learning takes place at low $F0$s such as 88 Hz than at high $F0$s such as 250 Hz.

Transfer of Learning

A stronger argument, however, can be found when one looks at how learning is transferred between resolved and unresolved conditions. When one considers now the changes in thresholds between the pre- and posttraining test sessions—namely, S1 versus S2—it must first be noted that no significant change was observed in the control group. Thus, the changes in thresholds that took place between these two test sessions in the two trained groups can be interpreted to be a result of the specific training that these groups received. Second, it is worth noting that although the two training groups were each trained in a single condition involving solely the mid frequency region, a significant main effect of session in the statistical analyses indicates a general improvement in DL $F0$s across the different conditions. This observation reveals that the improvement in $F0$ discrimination that resulted from training in one condition generalized, in part, to the other conditions that were tested. Such an overall increase in performance across frequency regions and $F0$s might reflect procedural learning, which refers to the learning surrounding the response demands of the task and does not depend on the specific characteristics of the stimuli used, as opposed to stimulus learning (Robinson & Summerfield, 1996). The subjects, who had no prior experience in psychoacoustical tests, had perhaps not completely familiarized themselves with the experimental procedure or task by the end of the preliminary (S0) and first (S1) test sessions and benefited from further training before S2. The lack of extra training for the control group would then explain the absence of overall improvement between S1 and S2 observed in the latter group. Another possible interpretation is that, although the subjects had become perfectly familiar with the task and test procedure by the end of the pretraining test (S1) session, the intensive training that the subjects of the TMID88U and TMID250 groups received following this session improved the functioning of some generic underlying $F0$-discrimination mechanism that does not strictly depend on

the frequency region and the $F0$ being tested. Several existing models of pitch perception involve two stages: a first stage that corresponds to the local extraction of the frequencies of the different harmonics, and a second stage that corresponds to the computation of the $F0$ on the basis of this information (e.g., Goldstein, 1973; Terhardt, 1972). Since the second mechanism operates across auditory channels, a refinement in its functioning under the influence of training might produce a global improvement in $F0$-discrimination performance.

An important qualification to the finding that $F0$ discrimination generalizes across $F0$s and regions comes from the observation of significantly larger transfer of learning between harmonic complex tones of the same resolution status (i.e., both resolved or both unresolved) than between conditions involving harmonic complex tones with a different resolution status. The second finding absolutely cannot be explained in terms of procedural learning and argues for the hypothesis that the perceptual encoding of the $F0$ of harmonic complex tones, as reflected by $F0$-discrimination performance, is mediated by different mechanisms depending on whether the harmonics are resolved or unresolved by the peripheral auditory system. The present finding can be paralleled with the earlier observation by Carlyon and Shackleton (1994) that the ability to compare the $F0$s of two simultaneous groups of harmonics filtered in different frequency regions is larger when both groups have the same resolution status than when one is resolved and the other is unresolved.

A final point of the results of the present study that deserve mention is in regard to the lack of a significant difference between DL $F0$s measured on the two posttraining sessions (S2 and S3). This finding reveals that the learning effects that occurred in this study were retained over a period of at least 5 weeks after the end of the training. This finding can be paralleled with earlier results in the visual modality, which indicate that perceptual learning effects are long lasting (Ball & Sekuler, 1982; Karni & Sagi, 1993).

CONCLUSION

In conclusion, besides providing data on the learning of $F0$ discrimination with harmonic complex tones, the present study provides evidence based on a transfer-of-learning paradigm, in which two different mechanisms underlie the perception of pitch depending on whether the harmonics are resolved or unresolved at the auditory periphery. The strongest argument in favor of this dual-pitch mechanism hypothesis is that subjects trained in $F0$ discrimination with resolved harmonics later show improved performance in *other* resolved conditions versus unresolved conditions, and vice versa. This effect cannot be accounted for by existing models of pitch perception in which a unitary mechanism processes the pitch of all harmonic complex tones. A consistent interpretation of these results is that in spite of the apparent unity of the sensory attribute of harmonic complex tones known as virtual pitch, owing to constraints

imposed on the central nervous system by the peripheral auditory organ, two separate neural mechanisms underlie the encoding of the $F0$ of resolved and of unresolved stimuli.

REFERENCES

AHISSAR, M., & HOCHSTEIN, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences,* **90**, 5718-5722.

AHISSAR, M., & HOCHSTEIN, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research*, **36**, 3487-3500.

AHISSAR, M., & HOCHSTEIN, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, **387**, 401-406.

AMERICAN NATIONAL STANDARD INSTITUTE. (1969). *Specification for audiometers* (No. S3.6-1969). New York: ANSI.

BALL, K., & SEKULER, R. A. (1982). Specific and enduring improvement in visual motion discrimination. *Science*, **218**, 697-698.

BALL, K., & SEKULER, R. A. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, **27**, 953-965.

CAMPBELL, R. A., & SMALL, A. M. (1963). Effect of practice and feedback on frequency discrimination. *Journal of the Acoustical Society of America*, **35**, 1511-1514.

CARIANI, P. A., & DELGUTTE, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, **76**, 1698-1716.

CARIANI, P. A., & DELGUTTE, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology,* **76**, 1717-1734.

CARLYON, R. P. (1998). Comments on "A unitary model of pitch perception." *Journal of the Acoustical Society of America*, **104**, 1118-1121.

CARLYON, R. P., & SHACKLETON, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *Journal of the Acoustical Society of America*, **95**, 3541-3554.

DEMANY, L. (1986). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America*, **78**, 1118-1120.

EDWARDS, A. E. (1973). *Statistical methods* (3rd ed.). New York: Holt, Rinehart & Winston.

FAULKNER, A. (1985). Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations? *Journal of the Acoustical Society of America,* **78**, 1993-2004.

FIORENTINI, A., & BERARDI, N. (1981a). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research*, **21**, 1149-1158.

FIORENTINI, A., & BERARDI, N. (1981b). Perceptual learning specific for orientation and spatial frequency. *Nature*, **287**, 43-44.

GOCKEL, H., CARLYON, R. P., & MICHEYL, C. (1999). Context dependence of fundamental frequency discrimination: Lateralized temporal fringes. *Journal of the Acoustical Society of America*, **106**, 3553-3563.

GOLDSTEIN, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, **54**, 1496-1516.

HARRIS, J. D. (1952). Pitch discrimination. *Journal of the Acoustical Society of America*, **24**, 750-755.

HARTMANN, W. M. (1988). Pitch perception and the segregation and integration of auditory entities. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 623-645). New York: Wiley.

IRVINE, D. R. F., MARTIN, R. L., KLIMKEIT, E., & SMITH, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of the Acoustical Society of America*, **108**, 2964-2968.

KARNI, A., & SAGI, D. (1990). Texture discrimination learning is specific for spatial location and background orientation. *Investigative Ophthalmology & Visual Science*, **31**(Suppl.), 562.

KARNI, A., & SAGI, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, **88**, 4966-4970.

KARNI, A., & SAGI, D. (1993). The time course of learning a visual skill. *Nature*, **365**, 250-252.

LEVITT, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

MEDDIS, R., & HEWITT, M. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I. Pitch identification. *Journal of the Acoustical Society of America*, **89**, 2866-2882.

MEDDIS, R., & HEWITT, M. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery: II. Phase sensitivity. *Journal of the Acoustical Society of America*, **89**, 2883-2894.

MEDDIS, R., & O'MARD, L. J. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, **102**, 1811-1820.

MICHEYL, C., & CARLYON, R. P. (1998). Effect of temporal fringes on fundamental-frequency discrimination. *Journal of the Acoustical Society of America*, **104**, 3006-3018.

MOORE, B. C. J. (1973). Frequency difference limens for short duration tones. *Journal of the Acoustical Society of America*, **54**, 610-619.

MOORE, B. C. J., & GLASBERG, B. R. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *Journal of the Acoustical Society of America*, **87**, 2163-2177.

MOORE, B. C. J., & GLASBERG, B. R. (1991). Effect of signal-to-noise ratio on the frequency discrimination of complex tones with overlapping or nonoverlapping harmonics. *Journal of the Acoustical Society of America*, **89**, 2858-2865.

PATTERSON, R. D., ROBINSON, K., HOLDSWORTH, J., McKEOWN, D., ZHANG, C., & ALLERHAND, M. (1991). Complex sounds and auditory images. In Y. Cazals, L. Demany, & K. Horner (Eds.), *Auditory physiology and perception* (pp. 429-446). Oxford: Pergamon.

PLACK, C. J., & CARLYON, R. P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between harmonic complex tones consisting of resolved and unresolved harmonics. *Journal of the Acoustical Society of America*, **98**, 1355-1364.

POLAT, U., & SAGI, D. (1994). Spatial interactions in human vision: From near to far via experience dependent cascades of connections. *Proceedings of the National Academy of Sciences,* **91**, 1206-1209.

RAGOT, R., & CROTTAZ, S. (1998). A dual mechanism for sound pitch perception: New evidence from brain electrophysiology. *NeuroReport*, **9**, 3123-3127.

ROBINSON, K., & SUMMERFIELD, A. Q. (1996). Adult auditory learning and training. *Ear & Hearing*, **17**, 51S-65S.

SATHIAN, K., & ZANGALADZE, A. (1997a). Perceptual learning in tactile hyperacuity: Complete intermanual transfer but limited retention. *Experimental Brain Research*, **118**, 131-134.

SATHIAN, K., & ZANGALADZE, A. (1997b). Tactile learning is task specific but transfers between fingers. *Perception & Psychophysics*, **59**, 119-128.

SCHOUPS, A. A., & ORBAN, G. A. (1996). Interocular transfer in perceptual learning of a pop-out discrimination task. *Proceedings of the National Academy of Sciences,* **93**, 7358-7362.

SCHOUPS, A. A., VOGELS, R., & ORBAN, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *Journal of Physiology*, **483**, 797-810.

SCHOUTEN, J. F., RITSMA, R. J., & CARDOZO, B. L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, **34**, 1418-1424.

SHACKLETON, T. M., & CARLYON, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *Journal of the Acoustical Society of America*, **95**, 3529-3540.

SHIU, L.-P., & PASHLER, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, **52**, 582-588.

SLANEY, M., & LYON, R. F. (1990). A perceptual pitch detector. In *Proceedings of the 1990 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 357-360). Piscataway, NJ: IEEE Press.

SPENGLER, F., ROBERTS, T. P., POEPPL, D., BYL., N., WANG, X., ROWLEY, H. A., & MERZENICH, M. M. (1997). Learning transfer and neuronal plasticity in humans trained in tactile discrimination. *Neuroscience Letters*, **232**, 151-154.

SPIEGEL, M. F., & WATSON, C. S. (1981). Factors in the discrimination of tonal patterns. III. Frequency discrimination with components of well-learned patterns. *Journal of the Acoustical Society of America*, **69**, 223-230.

STEINSCHNEIDER, M., RESER, D. H., FISHMAN, Y. I., SCHROEDER, C. E., & AREZZO, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving

pitch perception. *Journal of the Acoustical Society of America*, **104**, 2935-2955.

TERHARDT, E. (1972). Zur Tonhöhenwahrnehmung von Klängen: II. Ein Funktionsschema. *Acustica*, **26**, 187-199.

TURNER, C. W., & NELSON, D. A. (1982). Frequency discrimination in regions of normal and impaired sensitivity. *Journal of Speech & Hearing Research*, **24**, 34-41.

VLIEGEN, J., & OXENHAM, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, **105**, 339-346.

VON HELMHOLTZ, H. L. F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* Braunschweig: F. Vieweg & Sohn.

WATSON, C. S. (1980). Time course of auditory perceptual learning. *Annals of Otology, Rhinology & Laryngology*, **89**, 96-102.

WATSON, C. S., KELLY, W. J., & WROTON, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America*, **60**, 1176-1186.

WIGHTMAN, F. L. (1973). The pattern transformation model of pitch. *Journal of the Acoustical Society of America*, **54**, 407-416.

WRIGHT, B. A., BUONOMANO, D. V., MAHNCKE, H. W., & MERZENICH, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, **17**, 3956-3963.

WYATT, R. F. (1945). Improvability of pitch discrimination. *Psychological Monographs*, **58** (2, Whole No. 267).