

Evidence for a central representation of instrument timbre

MARK A. PITT

Ohio State University, Columbus, Ohio

An information processing account of perception seeks to delineate the stages of processing through which a stimulus passes and determine the properties of the representation at each stage. Research in phonetic perception has identified two stages, the second of which is thought to encode abstract acoustic attributes of sounds. The present study provided a further test of this proposal by assessing whether nonphonetic stimuli could yield results similar to those obtained with phonetic stimuli. Five selective adaptation experiments were carried out with a trumpet–piano timbre continuum. Two manipulations were used to measure abstract encoding: cross-ear presentation of adaptor and test series, and the use of adaptors that were acoustically different from the continuum endpoints. The results provide evidence for an abstract representation of timbre. The similarity of the findings to those in the phonetic adaptation literature is discussed.

A fundamental problem in perception is understanding how perceivers maintain object constancy across wide variation in the physical properties of a stimulus. In the auditory domain, this problem may be best known in the form of the acoustic–phonetic variability problem in speech perception, in which there is a lack of one-to-one correspondence between the acoustic properties of the speech signal and the resulting phonetic percept. For example, not only can the same acoustic cue signal different phonemes in different contexts, but different cues can signal the same phoneme in different contexts (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

The question of how listeners perceive object constancy also arises when investigating the perception of nonphonetic sound objects, such as the timbre of musical instruments (Handel, 1989). Tones produced by a saxophone or clarinet are readily perceived as such over most of the instrument's range, yet the acoustic properties of the tones change markedly across this range. Pressing and releasing keys alters not only the fundamental frequency of the tone but also the resonance (filtering) characteristics of the instrument. This variation is particularly large when the register (octave) key is used. Additional variability in the waveform is introduced by changes in the source of production (e.g., vibration amplitude of the reed, tone attack rate; Backus, 1977; Benade, 1976).

As with phonetic percepts, the acoustic cues that signal instrument identity vary across contexts, and no one cue has been found that is necessary and sufficient for

accurate identification. For example, tone onset information provides important cues for identification of some, but not all, instruments (Berger, 1964; Saldanha & Corso, 1964). The relative importance of cues can change when tones are played in different contexts. Steady-state and tone onset information are equally important for identification of some timbres when played in isolation, but steady-state information is more important when tones occur in musical phrases (Grey, 1978; Kendall, 1986). Research in voice identification has yielded similar results: Any of several cues is sufficient for identification, and the importance of a single cue can vary across contexts (Bricker & Pruzansky, 1976; Hecker, 1971; Van Lancker, Kreiman, & Emmorey, 1985).

According to the information processing approach to perception, perceptual constancy is achieved through a series of processing stages over which the stimulus is transformed into an abstract and stable representation. A goal of research in this tradition is to identify the stages of processing and their functional characteristics. One task that has proven useful in this endeavor is selective adaptation (Blakemore & Sutton, 1969; Eimas & Corbit, 1973), in which subjects categorize stimuli along a continuum (e.g., /ba/–/da/) before and after exposure to repeated presentations of another stimulus (the adaptor), frequently an endpoint of the continuum (e.g., /ba/). After adaptation, subjects tend to categorize stimuli in the middle of the continuum as members of the unadapted category (e.g., /da/).

The rationale behind the use of the technique is that repeated presentation of the adaptor slows or fatigues stimulus processing (e.g., by decreasing response intensity) at stages in the perceptual system (see Diehl, 1981; Remez, 1987; Samuel, 1986, for discussions of fatigue effects and the adaptation paradigm). These stages of analysis and their properties are identified by manipulating experimental conditions, such as the adaptor's re-

I thank Gary Starr for testing subjects and Alice Cimino for invaluable assistance in performing the timbral interpolation. A conversation with Arthur Samuel led to the design of Experiment 5. This research was supported by Grant R29 DC01774 from the National Institute on Deafness and Other Communication Disorders. Address correspondence to M. Pitt, Department of Psychology, 1885 Neil Avenue, Columbus, OH 43210.

relationship (e.g., acoustic, phonetic) to the stimuli in the continuum.

At present, experimental evidence suggests that there are two levels of perceptual analysis—a peripheral level, in which simple acoustic properties or dimensions of a stimulus are encoded, and a central level, in which the simple properties from the peripheral stage are integrated into more abstract representations (Eimas & Miller, 1978; Jamieson & Cheesman, 1986; Samuel, 1986; Sawusch, 1986; Sussman, 1993). The peripheral stage is thought to receive input from only one ear and is probably located beyond the cochlea. The central level is located upstream from the peripheral level (perhaps cortically) and receives input from both ears.

Research has focused on dissociating peripheral from central processes by using two types of manipulations. In the first, adaptation is assessed with adaptors that are acoustically different tokens of the continuum endpoints. Using whispered and voiced /ba/-/wa/ continua and adaptors, Samuel (1988; see also Ganong, 1978) found that whispered speech adapted voiced speech and vice versa. This result suggests that a relatively abstract representation was being adapted. Other studies have taken this manipulation to an extreme. Samuel and Newport (1979) and Sawusch (1977) found adaptation with adaptors that did not overlap in frequency with stimuli in the test series, suggesting that perceptual commonalities between the stimuli, not acoustic overlap, must be causing the effect. For example, Sawusch (1977) found reliable adaptation by using a three-formant /bae/-/dae/ continuum with endpoint adaptors whose formants had been shifted upward in frequency.

The second type of manipulation involves measuring adaptation interaurally. The adaptor is presented to the ear opposite that of the test series. Cross-ear adaptation is interpreted as evidence of central-level processing, because processing at the peripheral level in the tested ear should have been unaffected by cross-ear presentation of the adaptor (Ades, 1974; Eimas, Cooper, & Corbit, 1973; Jamieson & Cheesman, 1986; Samuel, 1986, 1988; Sawusch, 1977).

Probably the most convincing evidence for multiple processing levels comes from studies that have combined the two manipulations. In the study described above, Sawusch (1977) compared ipsilateral (same-ear) and contralateral (different-ear) adaptation by using continuum endpoints and frequency-shifted endpoints as adaptors. With the continuum endpoints, contralateral adaptation was half the size of that found ipsilaterally, suggesting that adaptation occurred peripherally and centrally. The frequency-shifted endpoints, on the other hand, yielded adaptation of equal magnitude contralaterally and ipsilaterally, suggesting that the central level was the sole locus of the effect; if it were not, ipsilateral adaptation should have been larger. Samuel (1988) found a similar pattern of results when whispered and voiced adaptors were crossed with whispered and voiced test continua. For each adaptor, contralateral adaptation was similar in magnitude regardless of whether the

adaptor and test series matched acoustically (both voiced or whispered). Only in the ipsilateral condition did acoustic match have a large influence on the magnitude of adaptation.

In addition to delineating the levels of analysis, research has also investigated the classes of stimuli (e.g., phonetic or nonphonetic) processed at each level. There is some consensus that the peripheral level processes simple acoustic properties and is indifferent to stimulus class (however, see Liberman & Mattingly, 1985). The weight of the current evidence also suggests that the central level is also indifferent to stimulus class, although there have been proposals that it selectively encodes phonetic information (see Eimas & Miller, 1978). The one manipulation that has been used to explore this issue has been to adapt speech with nonspeech adaptors. Studies have generally found adaptation (Diehl, 1976; Kat & Samuel, 1984; Samuel, 1988; Samuel & Newport, 1979; however, see Remez, Cutting, & Studdert-Kennedy, 1980). In a particularly elegant demonstration of the central locus of this effect, Samuel (1988) showed that an adaptor that resembled the sound of a plucked violin string produced reliable /ba/-like *contralateral* adaptation on a voiced and a whispered /ba/-/wa/ continuum. This finding not only argues for the central level's being acoustic, but also underscores the abstractness of representations at this level.

The argument that abstract acoustic information is encoded at the central level could be strengthened by demonstrating equivalent adaptation effects with nonphonetic stimuli. Such a demonstration is in fact necessary for some theories (e.g., Samuel, 1986, 1988), because processing equivalence is assumed in the claim that both classes of stimuli (phonetic and nonphonetic) are encoded at the same level. I tested this assumption in the present study by trying to measure central-level processing using a nonphonetic continuum. Across five experiments, adaptation on a trumpet-piano timbre continuum was measured in the context of the two manipulations designed to tap the central level (cross-ear adaptation and use of acoustically different adaptors).

A few researchers have used timbre continua successfully in adaptation experiments. Cutting, Rosner, and Foard (1976) and Remez et al. (1980) found reliable adaptation by using endpoint adaptors on a pluck-bow continuum. Similarly, Remez (1979; see also Remez, 1980) found adaptation with an /ae/-buzz continuum with endpoint adaptors. The present study is similar to these in purpose: It was designed to find parallels in speech and nonspeech processing. It builds on these studies by investigating the levels of analysis in nonspeech processing.

Conclusions about the locus of processing have been based almost exclusively on a single measure—changes in proportion identification after adaptation. Samuel (1986; Samuel & Kat, 1994) showed that analysis of identification response times (RTs) can provide insight into the nature of the mechanism responsible for adaptation and corroborating evidence about the locus of pro-

cessing. For example, RTs to steps in the endpoint regions of the continuum increase after adaptation to that endpoint, suggesting that repeated presentation of the adaptor causes a slowdown in stimulus processing. Such an outcome fits well with the operation of a fatigue-like mechanism. As with the identification measure, Samuel (1986) found larger RT effects with manipulations intended to tap peripheral rather than central processes, suggesting that the magnitude of the RT effect can serve as another method of distinguishing between processing at the two levels. Both measures of adaptation were used in the present study.

EXPERIMENT 1

Binaural Adaptation Using Endpoint Adaptors

The aim of the first experiment was to demonstrate adaptation with a trumpet–piano continuum by using endpoint adaptors. This was a necessary precondition for subsequent experiments, because adaptation serves as the index of processing at a particular level of analysis. The current manipulation is the most likely of any to yield adaptation. A failure to find it would make additional experiments pointless.

The design was a standard within-series selective adaptation experiment with subjects identifying stimuli on the timbre continuum before and after adaptation to the endpoints. I expected to find adaptation, given the success of past studies in producing reliable effects with timbre continua (e.g., Cutting et al., 1976).

Method

Subjects

Eight undergraduates at Ohio State University participated in exchange for pay. None reported hearing problems.

Stimuli

Piano and trumpet tones were chosen as the continuum endpoints because the instruments are readily recognizable (Pitt, in press) and the timbres are perceptually dissimilar (Iverson & Krumhansl, 1993), which should have minimized instrument confusability.

An 11-step trumpet–piano continuum was created by digitally combining a trumpet and a piano tone in varying proportions. The tones (E_4 ; 333 Hz) were recordings of real piano and trumpet tones taken from the *McGill University Master Samples* compact disk (Opolko & Wapnick, 1987). Stimuli were transferred to computer disk at a 44-kHz sampling rate. The first step in creating the continuum was to decompose the two tones into their individual partials and time-varying amplitude values by using Phase Vocoder software (Moore, 1990; Portnoff, 1976). The continuum was then synthesized by combining, partial by partial, the trumpet and piano tones in varying proportions by using the *cmusic* sound synthesis software (Moore, 1993). Only the first 20 partials (harmonic and inharmonic) of each tone were used in the synthesis. Each partial maintained a constant frequency over the duration of the tone, and amplitude phase relations were synchronized across partials at tone onset. The trumpet endpoint was synthesized by using a trumpet proportion of .999 and a piano proportion of .001. The second step consisted of trumpet and piano proportions of .90 and .10, respectively. At each successive step, the trumpet proportion decreased by .10 and the piano proportion increased by the same amount. The piano endpoint was synthesized by using trumpet/piano proportions of .001 and .999. Blends of the two tones

rather than the original tokens (e.g., 100% trumpet) were used as endpoints to ensure that each step was synthesized from the two timbres. All steps were then truncated to a duration of 311 msec, with the amplitude of the last 11 msec of the tone ramped to zero. Perceptually, the continuum sounded like one instrument timbre transitioning into the other, with the middle steps sounding halfway between the endpoints.

Amplitude waveforms and spectrograms of the continuum endpoints (Steps 1 and 11) are shown in the top part of Figure 1. As can be seen, the trumpet and piano tones differ in both dynamic and spectral characteristics. Amplitude onset begins abruptly in the piano tone and then decays steadily across its duration. The onset of the trumpet is much more gradual, and amplitude stays relatively constant once a maximum level is reached. Spectrally, the piano consists of harmonics below 2 kHz, the most intense of which is the fundamental. Harmonics extend over a much wider frequency range (up to 4 kHz) for the trumpet. In addition, many harmonics are prominent above 1 kHz and the third and fourth harmonics are the most intense.

The continuum construction procedure that was used differs from that in typical adaptation experiments, in which variation occurs along only one acoustic or phonetic dimension (e.g., amplitude rise time or voicing). Creation of the trumpet–piano continuum involved simultaneous variation of multiple dimensions (spectral and dynamic). Use of this procedure was prompted by a difference in research goals. Adaptation studies have focused largely on identifying perceptual features that produce adaptation, requiring precise control over variation in other acoustic dimensions. The question of importance here is, at which levels of processing can timbre adaptation be found? Whether the continuum was created by varying one or more dimensions is irrelevant.

Apparatus

The stimuli were downsampled to a rate of 10 kHz and then transferred to a microcomputer that controlled the experiment. The subjects were tested in groups of 4 or fewer, with each subject in a sound-damped cubicle. Stimuli were presented over headphones at approximately 72 dBA SPL (lowpass filtered at 4.8 kHz). The subjects responded by pressing one of two buttons on a response board; the left was labeled “trumpet” and the right was labeled “piano.” The subjects pressed each button with the corresponding index finger.

Procedure

Day 1. The experiment began with a brief familiarization session to expose the subjects to the endpoint tones. This was followed by a 28-block identification session to establish a baseline labeling function. The first 13 blocks were considered practice and were not included in the data analysis. During the first 2 practice blocks, only the three steps at each endpoint were presented (in a randomly permuted order), and the subjects were provided with feedback on the accuracy of their responses. In the remaining 26 blocks, all continuum steps were presented and no feedback was given. After each tone was played there was a 5-sec time-out, during which the subjects had to label the instrument as either a piano or a trumpet. An 1,800-msec pause occurred between the subjects' responses and the start of the next trial. A rest break was provided after Block 15.

The next phase of the experiment consisted of 18 alternating adaptation and identification blocks, the first three of which were counted as practice. In the adaptation blocks, the adaptor (trumpet or piano) was presented 120 times in the first block and 60 times in each subsequent block. A silent period of 650 msec separated each presentation of the adaptor. Identification blocks were the same as those in the identification session that began each day of testing: The subjects were presented with a random ordering of the continuum steps for identification. Between the adaptation and identification blocks, there was a 3-sec pause. A 5-sec pause sep-

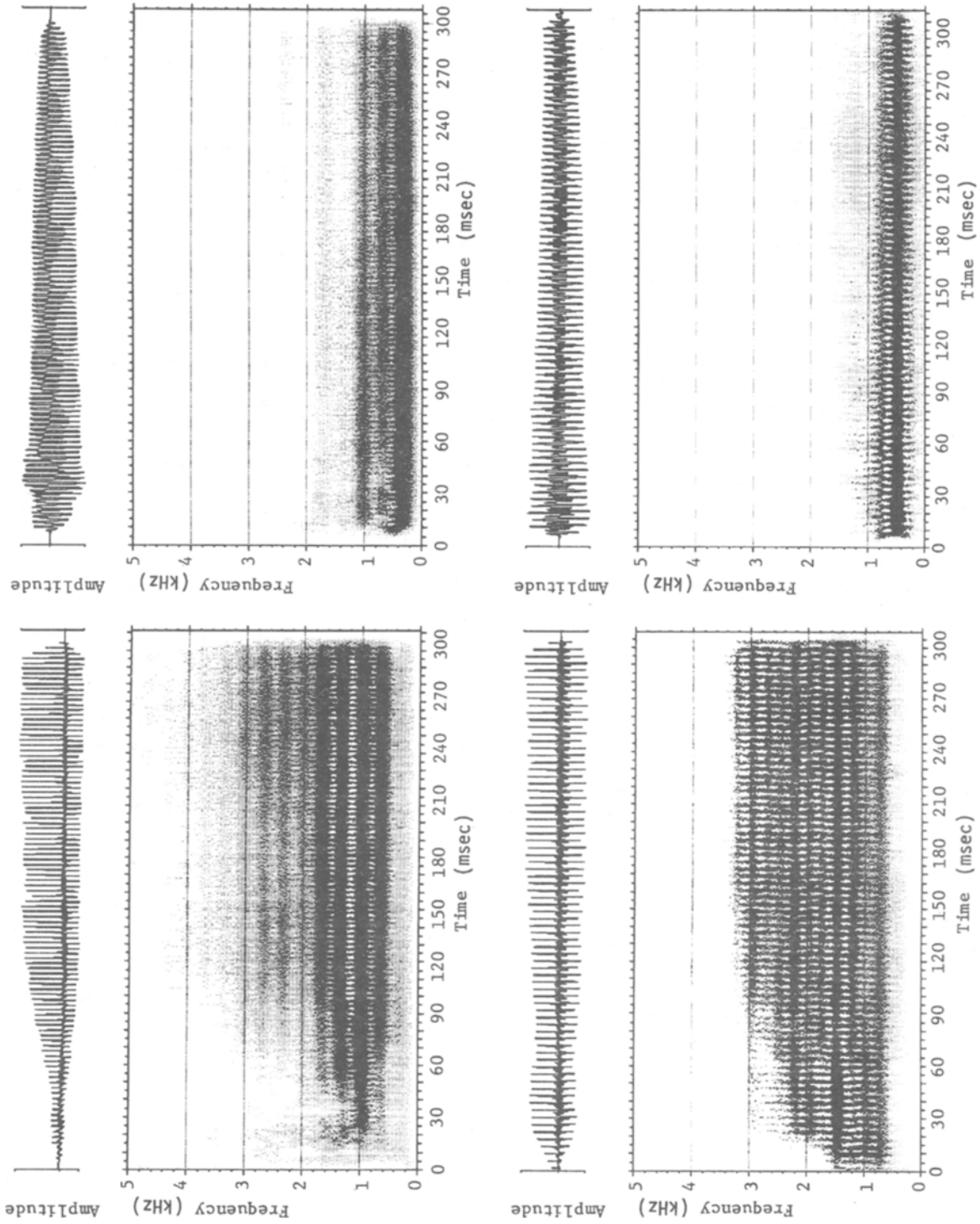


Figure 1. Waveforms and spectrograms of the trumpet-piano continuum (top panels) and the acoustically different trumpet and piano tokens (bottom panels). Trumpet tokens are on the left and piano tokens are on the right. Amplitude in the spectrogram is represented in a gray scale, with darker regions representing frequencies of higher energy.

arated each identification block from the next adaptation block. The subjects were given a rest break after Blocks 5 and 10.

Day 2. Testing differed from that of Day 1 in two ways. The baseline identification session consisted of only 18 blocks (3 practice and 15 test), and a different adaptor was used in the adaptation phase. Order of adaptor presentation (trumpet → piano, piano → trumpet) was counterbalanced across subjects over days. The experiment lasted for 1 h each day.

Results

For each subject, the proportion of trumpet responses at each continuum step was calculated for the preadapt and postadapt conditions for each adaptor. Subject data were then averaged in each condition. Preadapt and postadapt identification curves are shown as a function of continuum step in the top left panel of Figure 2.

The size of the adaptation effect was measured by determining the difference between the areas underneath the preadapt and postadapt functions over the boundary region of the continuum, defined as Steps 3–9 for all experiments in this series. Among the advantages of this method over computing ogives is that it reflects the extent of adaptation over most of the continuum (see Pitt & Samuel, 1993; Samuel, 1986). For each subject, the mean proportion of trumpet responses in the boundary region in the preadapt and postadapt conditions were calculated. Shift size was the mean of the difference between conditions over subjects.

Both adaptors produced large and reliable adaptation shifts in the appropriate direction. The piano adaptor increased trumpet responses by 17% after adaptation. An analysis of variance (ANOVA) across subjects yielded a statistically significant result [$F(1,7) = 8.47, p < .02$]. Trumpet responses dropped by 15% after adaptation to the trumpet. This effect was also reliable [$F(1,7) = 16.08, p < .005$].

Of interest in this study is how adaptation magnitude changes across manipulations. To capture the full extent of adaptation, its magnitude was measured as the area between the two postadapt curves (Samuel, 1986; Samuel & Kat, 1994). This measure reflects the total effect of the two adaptors and circumvents difficulties in trying to interpret and factor in baseline (preadapt) data. Note in Figure 2 that the preadapt functions drift apart in the boundary region of the continuum. This can cause over- or underestimation of the total amount of adaptation. By this new index, the shift size was 29% [$F(1,7) = 43.80, p < .0003$]. This index of adaptation magnitude was used in all subsequent analyses.

Following Samuel (1986), normalized reaction times were used to measure adaptation effects in the response time data. For each subject, mean RT was calculated for each continuum step in each condition. A grand mean was then calculated for each condition and subtracted from the corresponding step means, centering RTs around zero. Subject data were then averaged. The RT data are graphed as a function of condition and continuum step in the top right panel of Figure 2.

Adaptation should cause a slowdown in RTs to steps near an endpoint after adaptation to that endpoint.

Graphically, the functions (e.g., trumpet preadapt and postadapt) should be similar in shape, but the postadapt function should be raised relative to the preadapt function at the endpoint from which the adaptor originated (the adapted endpoint). This is what was found. Postadapt trumpet RTs were slower at the trumpet endpoint. The reverse was true for the two piano functions.

Adaptation magnitude was measured as the mean difference between the preadapt and postadapt RT functions at the adapted endpoint (Samuel, 1986). For each subject, a mean RT was calculated over Steps 1–3 for the pre- and postadapt trumpet functions. The same was done over Steps 9–11 for the two piano functions. Subject means were then submitted to ANOVAs. There was a 34-msec effect for the trumpet adaptor [$F(1,7) = 8.76, p < .02$] and a 29-msec effect for the piano adaptor that was marginally reliable [$F(1,7) = 4.94, p < .06$].

As mentioned above, comparison of the postadapt functions provides a measure of adaptation magnitude that reflects the combined effects of both adaptors. For the RT data, this measure was the mean RT difference across Steps 1–3 and 9–11 between the trumpet and piano postadapt functions. For each subject, a mean RT was calculated from the RTs at the adapted endpoints on the postadapt functions (Steps 1–3 on the trumpet continuum and Steps 9–11 on the piano continuum) and from the RTs at the unadapted endpoints on the postadapt functions (Steps 9–11 on the trumpet continuum and Steps 1–3 on the piano continuum). An ANOVA on these data yielded a reliable 46-msec shift [$F(1,7) = 5.70, p < .05$].

Discussion

The results clearly show that adaptation can be found on the trumpet–piano continuum when the endpoints serve as adaptors. A 29% shift was obtained in the identification analysis and a 46-msec shift emerged in the RT analysis. Direct comparison of the identification results with those from previous timbre adaptation studies is not possible because different measures of adaptation magnitude were used. However, visual comparison of identification functions suggests that the trumpet–piano continuum produced a larger shift than the pluck–bow continuum of Cutting et al. (1976). RT analyses were not carried out by Cutting et al.

EXPERIMENT 2

Binaural Adaptation Using Acoustically Different Adaptors

Experiment 2 was a first step in exploring the locus of the adaptation found in Experiment 1. Does the adaptation shift reflect processing changes at a peripheral locus, or are central-level processes contributing to the shift as well? To address this question, I replaced the endpoint adaptors with piano and trumpet tokens that differed acoustically (but not in perceptual identity) from those on the continuum. Reliable adaptation can be interpreted as preliminary evidence for a central locus of

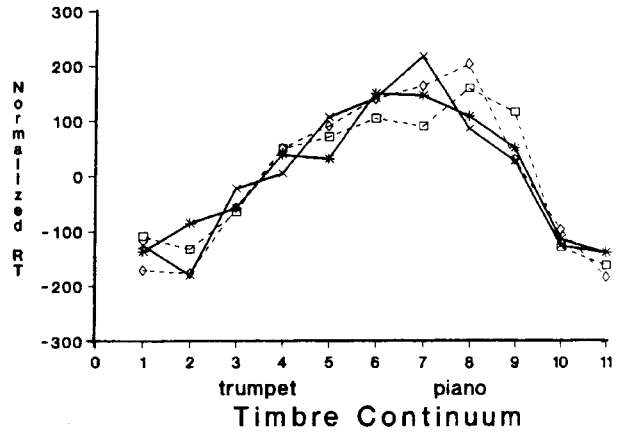
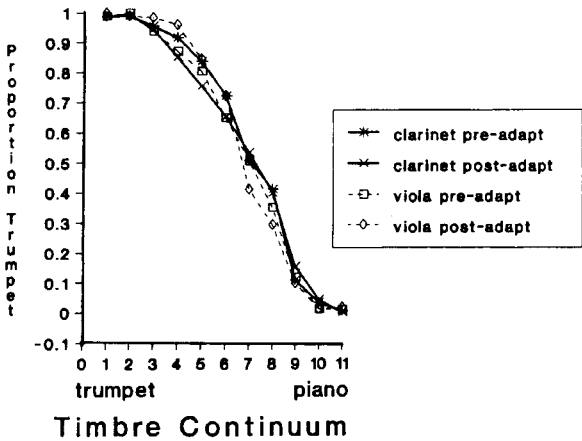
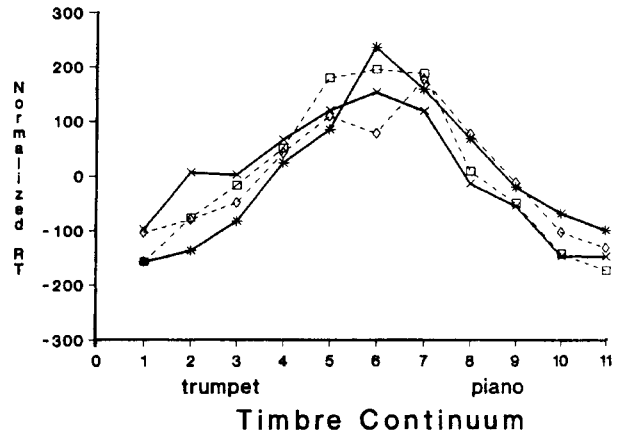
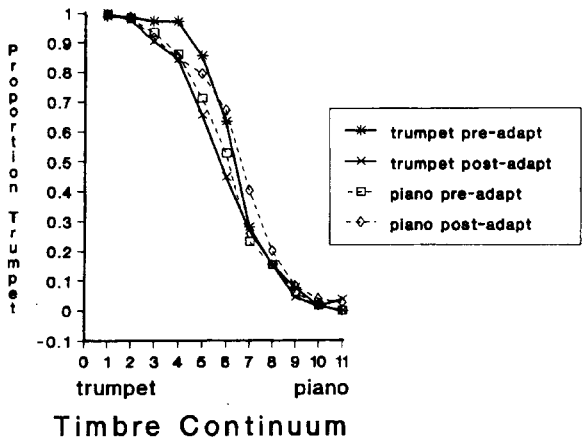
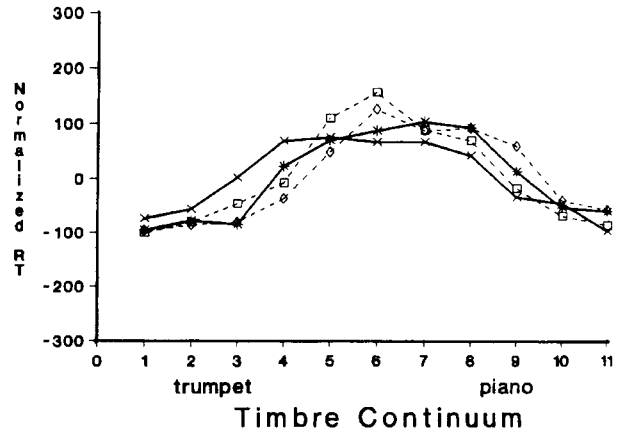
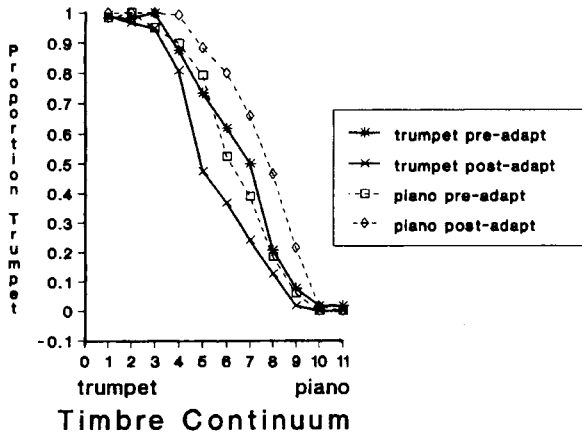


Figure 2. Identification and normalized reaction time (RT) data for Experiments 1–3. Shown in the graphs in the left column are mean proportion trumpet responses at each continuum step as a function of adaptor before and after adaptation. Mean normalized RT is plotted similarly on the right. Data from Experiment 1 are displayed in the top two graphs. The data of Experiment 2, in which the adaptors differed acoustically from the endpoint tokens, are shown in the middle panels. The data of Experiment 3, in which clarinet and viola tokens were used as adaptors, are shown in the bottom graphs.

timbre processing because it would suggest that more than simple acoustic overlap is the cause of the shift.

This adaptor manipulation is analogous to that used in phonetic adaptation studies (Samuel, 1988; Sawusch, 1977), in which shift size was 48% of that found with endpoint adaptors when averaged over experiments. A similar reduction in shift size is evident in the pluck-bow adaptation data of Cutting et al. (1976). Their continuum was constructed with a sawtooth waveform, and adaptors consisted of the endpoints and sinusoidal (i.e., different timbre) tokens of the endpoints. Compared with the sawtooth adaptors, the sinusoids produced a 50% effect when averaged over conditions. The present manipulation differs from that of Cutting et al. in that instrument (i.e., timbre) identity was held constant between adaptor and test items. Nevertheless, a reduction in adaptation magnitude was expected across Experiments 1 and 2.

Method

Subjects

Twelve new undergraduates, enrolled in an introductory psychology course, participated in exchange for course credit.

Stimuli

The methodology differed from that of Experiment 1 only in the use of adaptors. New piano and trumpet tones were recorded from the source used in Experiment 1. Their waveforms and spectrograms are shown in the bottom half of Figure 1. Examination of the figure shows how these adaptors differed from those of Experiment 1. For the trumpet tokens, amplitude onset is more abrupt in the bottom tone than in the top tone. The spectral envelopes of the tones are similar, except that the upper harmonics of the bottom tone are more intense than those of the top tone. For the two piano tones, amplitude onset is slightly more abrupt in the top tone. The upper harmonics of the top tone are more intense than those of the bottom tone.

The fundamental frequency of the new tones was 243 Hz (B_3), 90 Hz less than those used in Experiment 1. Although it could be argued that fundamental frequency should have been held constant while only acoustic properties that are associated with timbre varied (e.g., spectral and amplitude envelope), these properties covary naturally with frequency across an instrument's range. Indeed, holding them constant can result in a change in timbre (Charbonneau, 1981; see also Grey & Moorer, 1977). I expected the frequency difference to have a negligible effect on the outcome, given that past work (Cutting et al., 1976) has obtained adaptation under frequency variation.

Results

Data analysis followed the procedures described in Experiment 1. Mean identification and normalized RT functions are shown in the middle panels of Figure 2. As discussed in Experiment 1, adaptation was measured as the size of the shift between the two postadapt curves.

In the identification data, adaptation with the acoustically different adaptors produced a 10% shift that was reliable [$F(1,11) = 7.93, p < .02$]; it was also statistically different from the 29% shift obtained with the endpoint adaptors in Experiment 1 [$F(1,18) = 11.47, p < .003$]. In the RT data, the functions diverged from each other in the expected direction in all six endpoint steps (1-3 and 9-11), with postadapt trumpet RTs being slow-

est at the trumpet endpoint and fastest at the piano endpoint. The 41-msec shift was not reliable [$F(1,11) = 2.55, p < .14$], nor did it differ from the 46-msec shift in Experiment 1 ($F < 1$).

Discussion

The results provide some initial evidence that timbre is processed at a central level in the perceptual system. In the identification results, not only was reliable adaptation obtained, but it shrank to 34%, an amount that approximates the 48%-50% reduction that has been found in past work (Cutting et al., 1976; Sawusch, 1977). Shifts in RT functions have been shown to be nonsignificant when tapping central levels of analysis (Samuel, 1986). The failure to find a reliable shift is in accord with this interpretation. However, given that its magnitude is similar to that found in Experiment 1 (46 msec), conclusions about this effect should remain tentative for the time being.

The data of Experiments 1 and 2 can also be accounted for without invoking central-level processing, however. For example, the reduction in adaptation found across Experiments 1 and 2 could come about if adaptation were a function of acoustic (e.g., spectral) overlap at a peripheral stage of processing (Roberts & Summerfield, 1981; Sawusch & Jusczyk, 1981). Note in Figure 1 that the spectral envelope of the acoustically different adaptors differs somewhat from those of the endpoint adaptors. Although spectral overlap is not a necessary condition for adaptation (Samuel, 1986), this explanation cannot be discounted by the present results. One method of doing so is to rerun Experiments 1 and 2 with the cross-ear manipulation, which is designed to bypass the peripheral level of processing. Reliable contralateral adaptation would provide stronger evidence for a central locus of timbre processing. Before doing this, I addressed another issue.

EXPERIMENT 3

Binaural Adaptation With Different-Instrument Adaptors

Experiment 3 was a control experiment that was intended to demonstrate that the adaptation found in the preceding two experiments was selective to properties of the timbres used. That the piano and trumpet adaptors produced RT and identification shifts in the appropriate directions in Experiments 1 and 2 suggests that the adaptation observed was not random. Nevertheless, one could argue that any adaptor would shift the labeling function in some direction. This concern was addressed by rerunning Experiment 1 using clarinet and viola tones as adaptors. If any timbre can be adapted by any other timbre, then adaptation should emerge.

Method

Subjects

The participants were 12 Ohio State University undergraduates who had not participated in the preceding experiments.

Stimuli

The methodology was identical to that of Experiment 1, except that a clarinet tone and a viola tone (E_4 , 333 Hz, 311 msec in duration) were used as adaptors. The two timbres were chosen because of their dissimilarity to piano and trumpet timbres (Iverson & Krumhansl, 1993).

Results and Discussion

The data were analyzed using the procedures outlined in Experiment 1. Identification and RT functions before and after adaptation are shown in the bottom portion of Figure 2.

Both graphs tell the same story: Neither adaptor had much of an effect. Preadapt and postadapt functions overlap each other fairly closely in the identification and RT graphs. Not one statistical analysis reached significance ($F < 1$, in all cases). The results suggest that adaptation found in the preceding experiments was selective to specific perceptual dimensions of the stimuli that are not present in the clarinet or viola tones. What those dimensions are is an important question, but it is beyond the focus of this study (see Grey, 1977; Handel, 1989; Hall & Pastore, 1993).

EXPERIMENT 4

Cross-Ear Adaptation With Endpoint Adaptors

Recall that the purpose of this investigation was to test the claim that abstract acoustic (i.e., nonphonetic) dimensions of stimuli are encoded at a central stage of processing. The combined results of Experiments 1 and 2 provide preliminary evidence that this is the case. Experiments 4 and 5 provide a more stringent test of this claim by measuring adaptation interaurally. Contralateral adaptation should be found if timbre is processed centrally. This should be true regardless of whether the adaptor is an endpoint step on the continuum (Experiment 4) or an acoustically different piano or trumpet token (Experiment 5). This is because an abstract representation of timbre is assumed to be encoded at the central level.

In the present experiment, continuum endpoints served as adaptors and adaptation was measured ipsilaterally and contralaterally to the adapted ear. Phonetic adaptation experiments (Sawusch, 1977) have found contralateral identification shifts to be half the size of ipsilateral shifts under these conditions. RT shifts exhibit the same trend across laterality conditions (Samuel, 1986). I expected to obtain similar results.

Method

Subjects

Twelve new Ohio State University students were paid to participate in the experiment. The data from 1 subject were dropped from the analysis because the person responded randomly during Day 2 of testing.

Design

The experimental design was the same as that of Experiment 1, except that it was expanded to include the laterality manipulation

(ipsilateral vs. contralateral). As before, there were 2 days of testing, with adaptor (piano, trumpet) varying over days. On each day, the subjects were adapted in a different ear with a different adaptor, but were tested ipsilaterally and contralaterally. Thus, for a single subject, ear of adaptation (left vs. right) covaried with adaptor over days. Across subjects, however, adaptor and ear of adaptation were counterbalanced.

Procedure

On each day of testing there were 32 blocks of baseline identification, in which block presentation alternated between ears; the first 2 blocks were practice. In the adaptation phase (30 blocks), the adaptor was always presented to one ear while identification alternated from ear to ear across blocks. This procedure enabled ipsilateral and contralateral adaptation to be measured in the same experimental session. Identification always began in the right ear. Rest breaks were provided after every 10 blocks. The procedure was identical to that of Experiment 1 in all other aspects. The experiment lasted 1.5 h each day.

Results

Data were collapsed over ear of adaptation (adaptation magnitude was very similar across ears) and analyzed as a function of laterality (ipsilateral, contralateral) and adaptor (trumpet, piano). Identification and normalized RT functions are shown in Figure 3. Ipsilateral data are at the top of the figure; contralateral data are at the bottom.

Ipsilateral adaptation produced a large (32%) shift in the functions [$F(1,10) = 33.11, p < .0002$]. Contralateral adaptation, by contrast, yielded a much smaller (9%) shift [$F(1,10) = 13.16, p < .005$]. The difference between the size of the shifts was significant [$F(1,10) = 7.13, p < .02$], with the contralateral shift being 28% the size of that found ipsilaterally. This is about half the amount of interaural transfer found in previous work.

The RT data show a similar reduction in adaptation magnitude from the ipsilateral to the contralateral condition. Note that Steps 1–3 and 9–11 of the postadapt trumpet and piano functions are closer together in the contralateral case. In the ipsilateral condition, there was a 73-msec adaptation shift [$F(1,10) = 6.01, p < .03$]. The same comparison contralaterally yielded a 26-msec effect in the same direction that failed to reach significance [$F(1,10) = 1.61, p < .23$]. The contralateral shift was 36% the size of that found ipsilaterally. The difference between the shifts was not reliable [$F(1,10) = 2.95, p < .12$].

Discussion

In the identification data, the ipsilateral findings replicated those of Experiment 1: Adaptation was large and similar in magnitude (32% vs. 29%). Reliable adaptation was obtained in the contralateral condition, indicating that central-level processing was affected by the adaptor. Finally, the contralateral shift is similar in size to that found in Experiment 2 (9%), in which acoustically different adaptors were used, suggesting that those data also reflect central-level processing.

Trends present in the RT data tell a similar story. Contralateral adaptation was smaller (though not reliably)

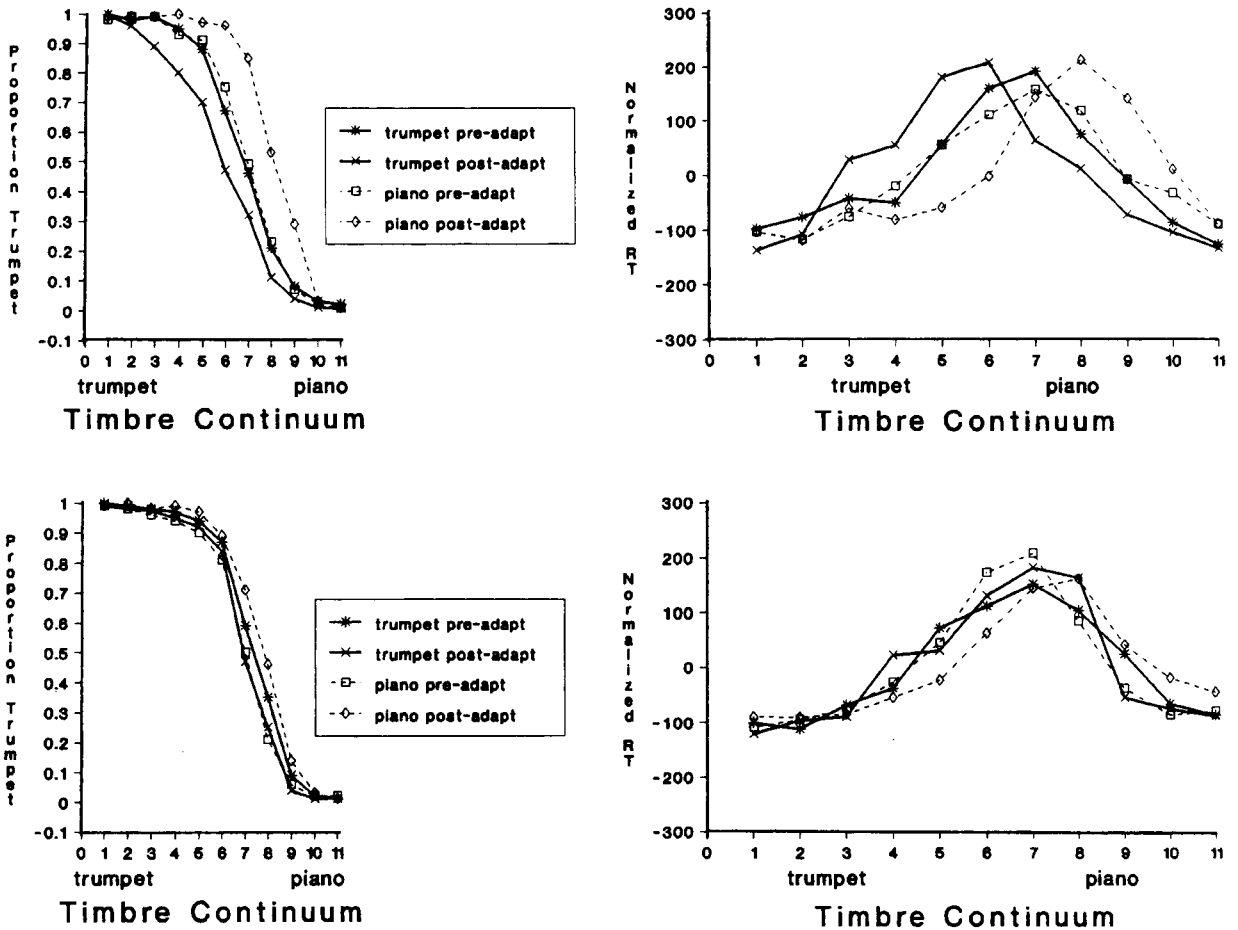


Figure 3. Identification and reaction time (RT) data in the ipsilateral (top graphs) and contralateral (bottom graphs) conditions of Experiment 4. Continuum endpoints served as adaptors.

than ipsilateral adaptation, and the drop in shift size to 36% in the contralateral condition is comparable to what was found in the identification results. Such a reduction in adaptation is expected across laterality conditions when a central stage is being tapped.

EXPERIMENT 5 Cross-Ear Adaptation Using Acoustically Different Adaptors

Experiment 5 was a final test of the question under study. If the central stage of processing encodes abstract acoustic features of instrument timbre, then adaptation at this level should be insensitive to the cross-ear manipulation. By this logic, if the shift obtained with the acoustically different adaptors (Experiment 2) reflects primarily central-level adaptation, then contralateral adaptation with these same adaptors should yield a shift similar in size.

This prediction was tested by rerunning Experiment 4 with the acoustically different adaptors of Experiment 2. Because the central level should be tapped in both later-

ality conditions, small but reliable identification shifts should be found in both cases. For the same reason, the RT shifts should be small and similar in magnitude across conditions.

Method

Subjects

Eleven new introductory psychology students served as subjects.

Stimuli

The methodology was the same as that in Experiment 4, except that the acoustically different adaptors used in Experiment 2 were used in place of the endpoint adaptors.

Results

Data analysis paralleled that in Experiment 4. Adaptation was again similar in magnitude across ears. Identification and RT functions are shown in Figure 4; the ipsilateral data are displayed above the contralateral data.

As can be seen in the identification data, ipsilateral adaptation and contralateral adaptation were small and similar in magnitude. A 17% shift was obtained in the former condition [$F(1,10) = 18.43, p < .002$], and a 10%

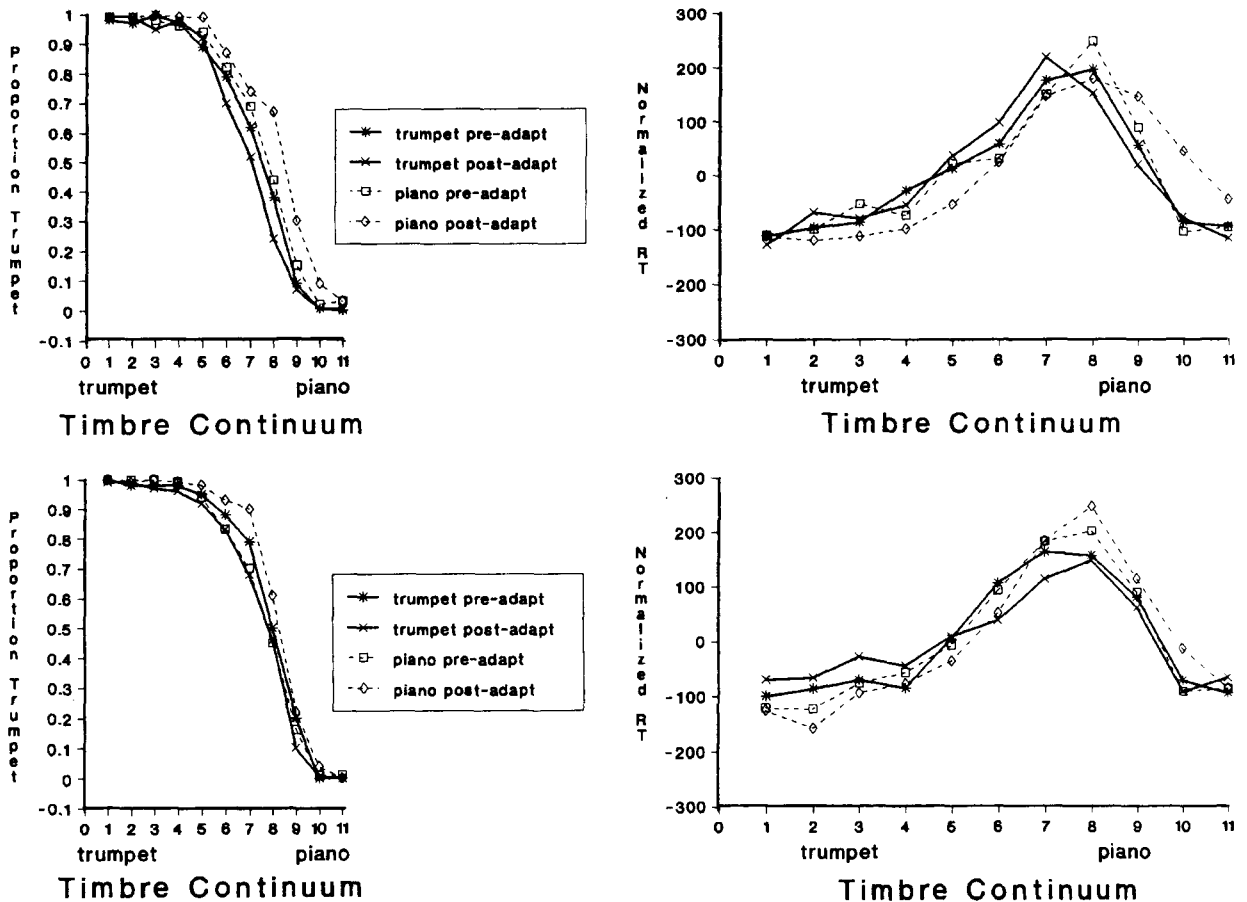


Figure 4. Identification and reaction time (RT) data in the ipsilateral (top graphs) and contralateral (bottom graphs) conditions of Experiment 5. Adaptors were piano and trumpet tokens that differed acoustically from the endpoints.

shift was obtained in the latter, which was marginally significant [$F(1,10) = 4.40, p < .06$]. The two shifts were not statistically different from each other ($F < 1$).

Adaptation was also similar in magnitude across laterality conditions in the RT data. Surprisingly, the shifts were large in both cases. There was a 65-msec shift in the ipsilateral condition [$F(1,10) = 15.33, p < .003$] and a slightly smaller (54 msec) shift in the contralateral condition [$F(1,10) = 11.75, p < .007$]. The two shifts did not differ reliably ($F < 1$).

Discussion

The results fell pretty much as predicted. In the identification data, small shifts were obtained in the two laterality conditions, although that in the contralateral condition was 41% smaller than the shift in the ipsilateral condition. This difference in shift size across laterality conditions may indicate that some of the adaptation in the ipsilateral condition occurred at a peripheral level. Alternatively, it could reflect variability across test situations. It is interesting to note that the contralateral shift is identical in size to that obtained in the equivalent binaural condition of Experiment 2 (10%). Furthermore, it

is close to that in the contralateral condition of Experiment 4 (9%), suggesting that perceptual commonalities between adaptor tokens are the source of the contralateral shift.

The RT results also fell as predicted, in that adaptation was of similar magnitude across laterality conditions. The unexpected finding was that large shifts were found in both cases. Adaptation as measured by RTs is typically small and statistically unreliable when tapping the central level. A comparison of these data with those from Experiments 2 and 4 may shed some light on their meaning. Adaptation in Experiment 2 was of a magnitude comparable (41 msec) to that in the present experiments, but failed to reach significance. Only in the contralateral condition of Experiment 4 was the shift small (26 msec) and nonsignificant. Given that three of the four conditions yielded large ($M = 53$ msec) shifts, the effects may well be real.

Samuel and Kat (1994) have recently proposed a three-stage model of information processing that can accommodate the outcomes in the present identification and RT analyses. The first stage is peripheral and encodes simple acoustic properties of stimuli. The second is located centrally and encodes integrated combinations

of acoustic features from the peripheral stage. At the third stage, which is also central, only phonetic properties of stimuli are represented. Adaptation is thought to be due to a fatigue-like mechanism in the second level, but not in the third.

Evidence for the existence of the two central stages comes from experiments that used a /ba/-/da/ continuum and two types of adaptors: /pa/ and /ta/ syllables and adaptors that comprised the second and third formants of the /ba/ and /da/ endpoint syllables. The measure that distinguishes processing at the two levels is the presence of contralateral RT shifts, which are assumed to reveal a processing slowdown that is indicative of a fatigue mechanism. The adaptors /pa/ and /ta/ produced reliable contralateral identification shifts and unreliable RT shifts, suggesting that the effect occurred centrally but was *not* due to fatigue. The two-formant adaptors produced reliable contralateral adaptation on both measures, suggesting that the shift affected an abstract representation that could be fatigued. Follow-up experiments showed that the latter effect was insensitive to peripheral-level manipulations (e.g., changes in intensity), further suggesting that adaptation occurred at a central level. The data in the present experiment more closely resemble those obtained with the two-formant adaptors, so they may also reflect processing at the second level.

GENERAL DISCUSSION

The aim of this study was to undertake another test of the proposal that abstract acoustic information is encoded at a central level in the information processing system (Samuel, 1986; Sawusch, 1986). Until now, this proposal has been based largely on results obtained with phonetic continua. The present study provides additional support for it by replicating results in the phonetic adaptation literature with a nonphonetic continuum.

The most compelling evidence suggesting that instrument timbre is processed at a central level of analysis comes from the identification results. Contralateral adaptation was obtained not only with endpoint adaptors, but also with acoustically different endpoint tokens. Furthermore, adaptation magnitude was very similar in both cases (10% and 9%, respectively), and it approximated the ipsilateral effect when acoustically different adaptors were used (17%).

The reaction time results also provide evidence for a central locus of processing. However, within Samuel and Kat's (1994) multistage model of information processing, the data are not in complete agreement about which central level is being tapped. The contralateral data of Experiment 4 are characteristic of what Samuel and Kat have referred to as a phonetic level of analysis: Contralateral RT shifts are small. The large RT shifts found in Experiments 2 and 5 are indicative of their abstract acoustic level. Further work is needed to resolve this issue.

The similarity of the current findings to those obtained with phonetic stimuli can be interpreted as indi-

cating that timbre and phonetic information are processed through some common stages. In the identification data, adaptation with the continuum endpoints produced reliable cross-ear effects that were 31% the size of those found ipsilaterally, approximating the 50% effect found with speech. Adaptation with acoustically different trumpet and piano tokens yielded contralateral adaptation that was comparable to that obtained binaurally (Samuel, 1986, 1988; Sawusch, 1977). In the RT results, adaptation was found near the continuum endpoints, mirroring speech findings (Samuel, 1986). Shifts that were a result of peripheral-level manipulations (same-ear presentation of endpoint adaptors) were in some instances larger than shifts due to central-level manipulations (different-ear presentation of endpoint and acoustically different adaptors). Even the unexpectedly large RT shifts in Experiment 5 have been found with phonetic stimuli (Samuel & Kat, 1994).

The current findings are not unexpected when vowels are thought of as timbres. Instrument timbres and vowels are similar acoustically in that both are periodic sounds with distinguishable formants. They also exhibit perceptual similarities. Slawson (1967) found that subjects judged the dissimilarity of pairs of stimuli the same, regardless of whether they were told the stimuli were vowels or instrument tones. (The stimuli were acoustically very similar in both conditions.) The one experiment that examined cross-ear adaptation with vowels yielded results similar to those of the present Experiment 4. Samuel (1986) used an /i/-/I/ continuum with endpoint adaptors and found that ipsilateral and contralateral identification shifts averaged 31% and 12%, respectively. Corresponding values for the trumpet-piano continuum were 32% and 9%. Samuel found larger ipsilateral than contralateral RT shifts, which was also true in Experiment 4.

Of course, obtaining analogous results across stimulus classes (phonetic, nonphonetic) does not require that a single processing pathway be advanced as the sole explanation for both sets of data. Separate pathways could handle the two types of stimuli and exhibit commonalities in processing. However, a primary reason to propose separate pathways is that differences in processing have been found. This was not the case in the present study.

A logical extension of the current work is to test whether phonetic adaptors can adapt nonphonetic stimuli. Remez et al. (1980) did just this and the reverse, using pluck-bow and /za/-/ja/ continua and adaptors, but found no reliable cross-series adaptation. Kat and Samuel (1984; Samuel, 1988) followed up this work and showed that plucks and bows can adapt phonetic continua (/ba/-/wa/ and /ča/-/ša/) when the adaptor and test items match in periodicity (i.e., spectral quality). They did not test the reverse, however.

Another useful next step—one that merges the current approach with the phonetic adaptation work—is to explore whether voice identification proceeds through the same stages as phonetic identification. A talker's voice has a timbre, and talker information and phonetic infor-

mation are conveyed simultaneously in the speech signal. A growing body of work suggests that phonetic information is not processed independently of the source of the message (Nearey, 1989; Nygaard, Sommers, & Pisoni, 1994). Talker information has been shown to influence the perceptual encoding of speech (Johnson, 1990; Mullenix & Pisoni, 1990) and word recognition (Mullenix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993). The processing of talker and phonetic information may therefore be carried simultaneously at each stage of processing. Work is under way to explore this possibility.

REFERENCES

- ADES, A. E. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, **16**, 61-66.
- BACKUS, J. (1977). *The acoustical foundations of music*. New York: Norton.
- BENADE, A. H. (1976). *Fundamentals of musical acoustics*. London: Oxford University Press.
- BERGER, K. W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, **36**, 1888-1891.
- BLAKEMORE, C., & SUTTON, P. (1969). Size adaptation: A new after-effect. *Science*, **166**, 245-257.
- BRICKER, P., & PRUZANSKY, S. (1976). Speaker recognition. In N. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295-326). New York: Academic Press.
- CHARBONNEAU, G. R. (1981). Timbre and the perceptual effects of three types of data reduction. *Computer Music Journal*, **5**, 10-19.
- CUTTING, J. E., ROSNER, B. S., & FOARD, C. F. (1976). Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, **28**, 361-378.
- DIEHL, R. (1976). Feature analyzers for the phonetic dimension *stop vs. continuant*. *Perception & Psychophysics*, **19**, 267-272.
- DIEHL, R. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, **89**, 1-18.
- EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. (1973). Some properties of linguistic feature detectors. *Perception & Psychophysics*, **13**, 247-252.
- EIMAS, P. D., & CORBIT, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, **4**, 99-109.
- EIMAS, P. D., & MILLER, J. L. (1978). Effects of selective adaptation of speech and visual patterns: Evidence for feature detectors. In H. L. Pick & R. D. Walk (Eds.), *Perception and experience* (pp. 307-345). New York: Plenum.
- GANONG, W. E., III (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics*, **24**, 71-83.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, **61**, 1270-1277.
- GREY, J. M. (1978). Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America*, **64**, 467-472.
- GREY, J. M., & MOORE, J. A. (1977). Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, **62**, 454-462.
- HALL, M. D., & PASTORE, R. E. (1993, November). *An auditory analogue to feature integration*. Paper presented at the meeting of the Psychonomic Society, Washington, DC.
- HANDEL, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.
- HECKER, M. (1971). Speaker recognition. *American Speech and Hearing Association Monographs*, **16**, 24-49.
- IVERSON, P., & KRUMHANS, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, **95**, 2595-2603.
- JAMIESON, D. G., & CHEESMAN, M. F. (1986). The locus of selective adaptation in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 286-294.
- JOHNSON, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, **88**, 642-654.
- KAT, D., & SAMUEL, A. G. (1984). More adaptation of speech by non-speech. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 512-525.
- KENDALL, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception*, **4**, 185-214.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- MOORE, F. R. (1990). *Elements of computer music*. Englewood Cliffs, NJ: Prentice-Hall.
- MOORE, F. R. (1993). *The cmusic sound synthesis program* [Computer program]. La Jolla, CA: Computer Audio Research Laboratory, Center for Music Experimentation, University of California at San Diego.
- MULLENNIX, J. W., & PISONI, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, **47**, 379-390.
- MULLENNIX, J. W., PISONI, D. B., & MARTIN, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365-378.
- NEAREY, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.
- NYGAARD, L. C., SOMMERS, M. S., & PISONI, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.
- OPOLKO, F., & WAPNICK, J. (Producers). (1987). *McGill University master samples* [Compact disk]. Montreal, Quebec: McGill University.
- 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. (in press). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance*.
- PITT, M. A., & SAMUEL, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 699-725.
- PORTNOFF, M. R. (1976). Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Transactions on Acoustics, Speech & Signal Processing*, **ASSP-24**, 243-248.
- REMEZ, R. E. (1979). Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, **11**, 38-57.
- REMEZ, R. E. (1980). Susceptibility of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception. *Perception & Psychophysics*, **27**, 17-23.
- REMEZ, R. E. (1987). Neural models of speech perception: A case history. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 199-225). Cambridge: Cambridge University Press.
- REMEZ, R. E., CUTTING, J. E., & STUDDERT-KENNEDY, M. (1980). Cross-series adaptation using song and string. *Perception & Psychophysics*, **27**, 524-530.
- ROBERTS, M., & SUMMERFIELD, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, **30**, 309-314.
- SALDANHA, E. L., & CORSO, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, **36**, 2021-2026.
- SAMUEL, A. G. (1986). Red herring detectors and speech perception:

- In defense of selective adaptation. *Cognitive Psychology*, **18**, 452-499.
- SAMUEL, A. G. (1988). Central and peripheral representation of whispered and voiced speech. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 379-388.
- SAMUEL, A. G., & KAT, D. (1994). *Levels of analysis of speech and other complex sounds*. Manuscript submitted for publication.
- SAMUEL, A. G., & NEWPORT, E. L. (1979). Adaptation of speech by nonspeech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 563-578.
- SAWUSCH, J. (1977). Peripheral and central processes in selective adaptation of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **62**, 738-750.
- SAWUSCH, J. R. (1986). Auditory and phonetic coding of speech. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines* (pp. 51-88). New York: Academic Press.
- SAWUSCH, J. R., & JUSCZYK, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 408-421.
- SLAWSON, A. W. (1967). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America*, **43**, 87-101.
- SUSSMAN, J. E. (1993). Focused attention during selective adaptation along a place of articulation continuum. *Journal of the Acoustical Society of America*, **93**, 488-498.
- 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.

(Manuscript received March 7, 1994;
revision accepted for publication August 1, 1994.)