APU 331. 3. 95.

©1995 BY THE REGENTS OF THE
UNIVERSITY OF CALIFORNIA

Music Perception
Fall 1995,
Vol. 13,No. 1,1—15

The Duration Required To Identify the Instrument, the Octave, or the Pitch Chroma of a Musical Note

KEN ROBINSON
MRC Institute of Hearing Research, Glasgow

ROY D. PATTERSON
MRC Applied Psychology Unit, Cambridge[1]

This paper investigates the role of pitch in the extraction of timbre information by measuring listeners' ability to identify the timbre, the octave, and the pitch chroma of musical notes, as a function of the duration of the notes. The stimuli were produced by one of four instrument types (brass, flute, harpsichord, or strings) in one of four octaves (centered at C1, C2, C3, or C4) on one of four notes (C, D, E, or F). The stimulus duration ranged from 1 to 64 cycles of the note. In any given session of the experiment, listeners were played all 64 notes associated with one duration in a random order and asked to identify the instrument, or the octave, or the note of each stimulus as it occurred. The results show that the timbre of the notes can be identified when the durations are too short to support pitch-chroma judgments, and so it is unlikely that pitch plays a key role in timbre identification at short durations. At these same durations, octave identification was better than pitch-chroma identification but worse than instrument identification.

Introduction

In many models of auditory perception, it is assumed that pitch plays a key role in the extraction of the timbre of complex periodic sounds like musical notes and vowels. For example, in an article on virtual pitch and music perception, Terhardt (1987, p.279) concludes that pitch is "more than just a by-product.. .[rather] it gains the position of an essential tool in the auditory extraction of phonetic information." In the same volume, Patterson (1987, p. 179) concludes a paper on the pulse ribbon model with "...pitch is not just one of many speech features. Rather, it is a key feature that makes it possible to stabilize the timbre of the voiced parts of speech and so extract the speech features more effectively."

BACKGROUND

The fact that the identification of concurrent vowels improves when the vowels have different pitches is commonly assumed to show that pitch plays a key role in timbre extraction (Assmann & Summerfield, 1990; Meddis & Hewitt, 1992; Scheffers, 1983). Specifically, Scheffers (1983) hypothesized that the auditory system extracts the pitch of a sound on a moment-to- moment basis and uses the pitch value to direct voice

[1] Requests for reprints may be sent to Ken Robinson, MRC institute of Hearing Research, Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER, United Kingdom or Roy D. Patterson, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom.

segregation. He generated pairs of concurrent vowels with the same or different pitches and demonstrated that listeners' performance on a vowel identification task was better when the vowels had differences in fundamental frequency (f_0) of a semitone or more. Then, using the harmonic sieve model of Duifhuis, Willems, and Sluyter (1982), he attempted to show that vowel identification systems based on spectral template matching could be improved if pairs of pitches were derived from the vowel pair on a frame-by-frame basis and used to scale the vowel templates before matching. Subsequently, Assmann and Summerfield (1990) replicated Scheffers' experimental findings, and both they and Meddis and Hewitt (1992) fitted the data with a multichannel, autocorrelation model of pitch perception (Licklider, 1951/ 1979). Both groups derived f_0 estimates from individual autocorrelogram frames and used them to restrict the components of the autocorrelogram that are passed to the recognition system. They demonstrated that recognition systems with 10 information achieved better performance than those same systems without f information.

The computational models in these papers are presented as auditory models, and in each case, they include the explicit, or implicit, assumption that the auditory system can extract an accurate estimate of the absolute value of the f of a vowel from a single frame of the internal representation of the sound, be it a spectral frame or an autocorrelogram frame. The accuracy required by these computational models is 2—3% of f_0 for stimuli as short as 20 ms, which is the frame size used in the models of Scheffers and of Meddis and Hewitt. The auditory system is extremely sensitive to changes in the pitch of both sinusoidal and complex tones, provided they are about 10 dB above masked threshold (Henning, 1967; Scheffers, 1983). But this does not mean that the auditory system extracts the absolute value of the pitch, and it certainly does not mean that it can extract an absolute f value to the level of accuracy required by the computational segregation models.

There do not appear to be any studies comparing the extraction of pitch and timbre information from short-duration, complex sounds. Nor are there studies on pitch identification as a function of duration. There are studies, however, that have examined the effect of duration on vowel identification (Gray, 1942; Suen & Beddoes, 1972) and studies that have examined the effect of duration on pitch perception (Patterson, Peters, & Milroy, 1983; Whitfield, 1979). Gray varied the duration of 11 vowels from 3 to 520 ms and found that all listeners were able to identify the vowels at better than chance levels given one cycle of the waveform. For vowels with low f0s, performance is better than 75% correct; for vowels with high f0s, performance improved from about 30% to 65% correct as the number of cycles increased from one to four. Suen and Beddoes studied the identification of five vowels at durations of 10, 20, and 30 ms and found that identification was possible at 10 ms. These studies suggest that one to four cycles of a vowel are required to extract their timbre.

Studies of the pitch of complex tones suggest that 8—10 cycles of the stimulus are required for a stable pitch perception. Whitfield (1979) constructed stimuli with alternating segments from two multiharmonic stimuli with f0s of 214 and 187 Hz. He varied the number of cycles in the segment to determine when two separate pitches could be heard. So, if A and B are single-cycle segments from the two multiharmonic stimuli, then a compound one-cycle stimulus had the form ABABABABABAB..., a compound two-cycle stimulus had the form AABBAABBAABB..., a three-cycle stimulus had the

form AAABBBAAABBB..., and so on. He found that with one- cycle and two-cycle segments, listeners heard a steady tone with a pitch between 214 and 187 Hz, and with four to six cycles per segment, listeners heard a fluttering sound with an unstable pitch. Two alternating pitches were not heard until the stimuli had about 10 cycles per segment. Patterson et al. (1983) have reported a melodic pitch experiment using sinusoids. A random four-note melody was presented and then repeated, and one of the four notes in the second version was transposed up or down one note of the diatonic scale. The number of cycles was varied to determine when the listeners could identify the position of the transposed note in the second version of the melody. For sinusoids ranging in frequency from 100 to 900 Hz, listeners required about 10 cycles to perform the task.
if it is correct that 1—4 cycles of a vowel is sufficient to identify the quality, and 10 cycles is required to define pitch chroma to the point where it will support a melodic pitch judgment, it seems unlikely that pitch chroma plays a key role in the extraction of timbre information for short-duration sounds.

THE CURRENT EXPERIMENT

This paper presents an experimental test of Scheffers' (1983) hypothesis that the auditory system extracts the pitch of a sound on a moment-to-moment basis and uses the pitch value to assist timbre identification. Specifically, the experiment was designed to determine how many cycles of a periodic sound are required to support identification of the timbre, the octave, or the pitch chroma of the sound. Synthesized instrument sounds spanning a four-octave range of fundamental frequencies were presented to listeners, who were required to identify the instrument, the octave, or the note of the stimulus. The primary independent variable was the number of cycles of sound in the stimulus. If pitch were essential for timbre perception, we might expect the psychometric functions for octave and note identification to rise above chance either before or at the same point as those for instrument identification.

A note-identification task (C, D, E, or F) was chosen in preference to a pitch - discrimination task for two reasons: First, it is more comparable to the instrument-identification task. Second, the computational models that use pitch to improve vowel identification require an absolute f_0 value, which is more analogous to a pitch-identification task than to a pitch-discrimination task. Identifying which of four chroma categories a note belongs to should be a relatively easy task if the auditory system extracts the absolute value of f_0, as suggested in the computational models. The models require f_0 estimates with 2—3% accuracy. The chroma steps between C, D, and E of the diatonic scale are 12% off0 that from E to F is 6%. The just-noticeable difference for the 10 of a musical note would typically be less than 05% of f_0, and so the width of the chroma categories should not restrict performance.

The octave identification task was included as an alternative, rather easier, pitch-categorization task. Experiments on octave identification have been described by Patterson (1990) and Patterson, Milroy, and Allerhand (1993). They presented listeners with multiharmonic synthetic notes in which the octave of the stimuli ranged from C1 to C6. Listeners had to identify the octave of the note either on an integer scale from 1 to 6 (Patterson, 1990) or on a decimal scale from 1.0 to 6.0 (Patterson et al., 1993). The average rating was very close to the physical octave throughout the six-octave range,

indicating that, at least for long-duration stimuli, listeners can perform an octave identification task with a high degree of accuracy.

Method

STIMULI AND EQUIPMENT

For purposes of this experiment, timbre was specified simply as the class of the instrument: brass, flute, harpsichord, or string. The listeners readily recognized instrument sounds labeled this way, and the labels clearly refer to an aspect of the sound that is distinct from its pitch and loudness. The tone height, or octave, of the instrument was specified as 1, 2, 3, or 4 in standard keyboard notation where "middle C" is C4, and A4 is 440 Hz. The fundamentals of the notes C1, C2, C3, and C are just under 33, 66, 131, and 262 Hz, respectively. The tone chroma, or note value, was, D, E, or F on the equal-temperament scale, that is, each D is 11.9% above the corresponding C, each E is 11.9% above the corresponding D, and each F is 5.9% above the corresponding E. There were a total of 64 instrument conditions in the experiment, and for each, stimuli were generated with 1—64 cycles.

The brass, flute, harpsichord, and string sounds were produced as analog waveforms by a Yamaha DX-9 synthesizer. The instruments were all taken from the "Master Group" of instruments supplied by the manufacturer. The notes C , C2, C3, and C4 were digitized at a sample rate of 16,384 Hz with a 12-bit analog-to-digital converter. The C notes used in the experiment were then made by excising the cycle that had the highest amplitude and playing that cycle repeatedly. In the experiment, the excised cycle was played 1, 2, 4, 8, 16, 32, or 64 times. The beginning and end of the cycle were both at positive-going zero crossings. There was no attack, decay, or other temporal cue as there would be with normal musical instruments, and so the instrument judgment is somewhat more difficult than it would normally be. For convenience, the D, E, and F notes were produced by playing the corresponding C note with a sample rate that was increased from 16,384 to 18,350, 20,550 and 21,785 Hz, respectively.

The absolute duration of the stimuli varied from 2.9 ms, which is I cycle of the F in the highest octave, to 1,952 ms, which is 64 cycles of the C in the lowest octave. Within a run, the range of durations was restricted to a factor of 8; that is, the largest number of cycles was eight times the smallest. The variation in duration contributes to the perceptual variability of the set in any given run, and this probably increases the difficulty of the task. Nevertheless, as the perceptual load is the same for all three tasks, the procedure does not make the pitch-chroma task inherently more difficult than the timbre task or the octave task.

The spectrum of each instrument playing the note C4 is presented in Figure 1. The ordinate shows relative level in decibels; the abscissa is frequency. The cutoff frequencies of the anti-aliasing filters were set at 4.2 kHz, and so the higher harmonics shown in the figure were not presented to the listeners. After intensity equalization, the flute notes were clearly less loud than the notes of the other instruments. To reduce these and other loudness differences, the harpsichord stimuli were increased in level by 1.5 dB, the string stimuli by 3 dB, and the flute stimuli by 7.5 dB, all relative to the brass stimuli. For each doubling in the number of cycles, the stimuli were reduced in power by 3 dB. The

presentation level for the 64-cycle brass stimuli was 65 dB SPL total power, when played continuously.

The stimuli were presented via a 12-bit digital-to-analog converter, two low-pass filters connected in series (96 dR/octave attenuation above the cutoff frequency), a programmable attenuator, and a Quad 303 amplifier, to a pair of Sennheiser HDS4OR headphones. The fidelity of the system was measured at the input to the headphones by presenting a 1-kHz sinusoid at the level (90.5 dB SPL) of the notes with the greatest amplitude (the one-cycle flute notes). The harmonics of the sinusoid and the noise floor were both more than 60 dB down from the level of the sinusoid. The stimuli were presented binaurally to the listener in an IAC sound-attenuated booth.

PROCEDURE AND LISTENERS

On each trial, the listener was presented a single note representing one of the 64 combinations of four instruments, four octaves, and four pitch chromas. A trial consisted of a 200-ms ready Light, followed by a 300-ms silence, and then a single presentation of the stimulus. Listeners had 5 s in which to respond, and they were given feedback on every trial. There were seven durations 1, 2, 4, 8, 16, 32, and 64 cycles, and in each run of the experiment, all combinations associated with four adjacent durations were presented, for a total of 256 trials. The shortest of these four durations was varied between runs to measure the psychometric function and devote most trials to the steepest part of the psychometric function. Stimulus presentation was randomized within the dimension of interest, and either blocked or randomized across the other two dimensions. For example, when the task was instrument identification in the blocked condition, both instrument and number of cycles were randomly varied between trials, whereas octave and note were blocked. In the randomized condition, instrument, octave, note, and number of cycles were randomized for every trial. The randomization was performed without replacement.

On any given day, the type of response required of the listener was fixed, and there were four response alternatives: For instrument identification, the response buttons were labeled “Brass,”“Flute,”“Harpsichord,” or “String.” For octave identification, the response choices were “1,”“2,”“3,” or “4,” and for note identification the available responses were “C”“D,”“E,” or “F.” Listeners were asked to identify instruments on Day 1, octaves on Day 2, and notes on Day 3, and then the task order was reversed for Days 4—6. All listeners completed Days 1—6; two of the listeners were available for further testing (KR and BS), and for them, the experiment was extended for a further 6 days. Each day began with a demonstration of the stimuli to remind listeners of the full range. They were then given 32 practice trials on the response task of that day in which no data were collected. During each run of the experiment, four demonstration trials were presented with the correct response between every set of 16 trials. Four listeners participated in each version of the experiment; three participated in both versions. The listeners ranged in age from 24 to 42 years, and all had normal binaural hearing thresholds as tested by pure-tone audiometry at frequencies of 0.25, 0.5, 1, 2, and 4kHz.

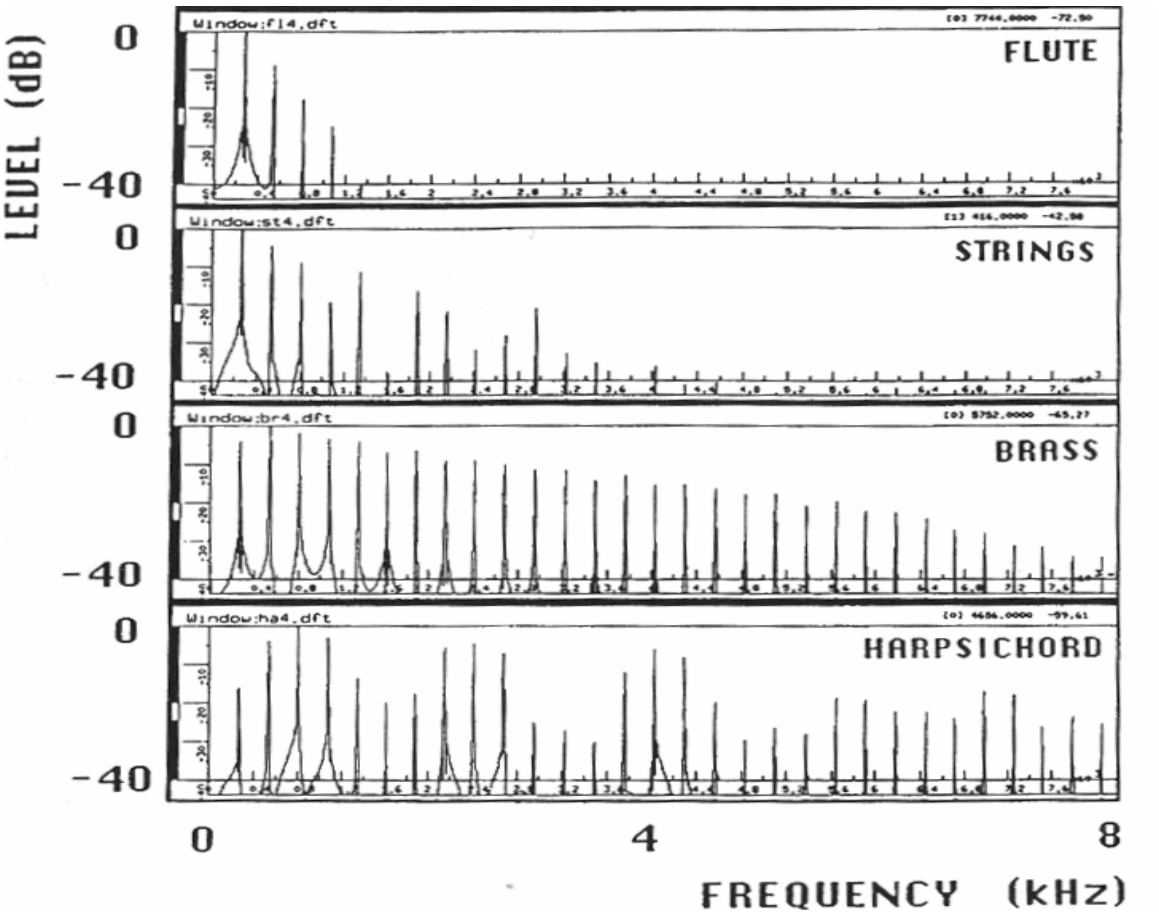

Fig. 1. Amplitude spectra for the brass, harpsichord, flute, and string stimuli. The fundamental frequency is 262 Hz (C4).

In musical terms, the entire experiment was performed within the key of C major. This meant that listeners, who understood the concept of the tonic, either explicitly or implicitly, did not need to extract the absolute value of (0 to perform the task. They could use an interval judgment to perform the four-way categorization. In psychoacoustic terms, these listeners would be making a four-way discrimination.

Results

PSYCHOMETRIC FUNCTIONS

Psychometric functions were prepared for instrument, octave, and note identification for each listener separately and for the blocked and random conditions separately. All of these individual psychometric functions rise above chance performance levels by 16 cycles at the latest, indicating that every listener could perform all three tasks in both the blocked and random conditions. Performance was lower in the random conditions than the blocked conditions, as would be expected, and it was differentially lower for the random pitch-chroma task. But the pattern of results was similar in the blocked and random conditions for all listeners, and so the blocked and random data were averaged in each case. These average psychometric functions revealed two patterns of

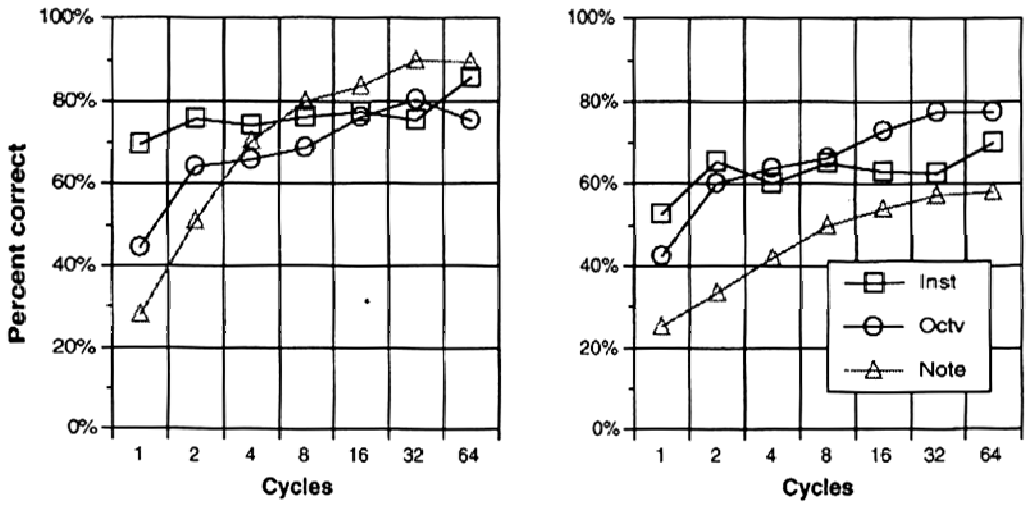

Fig. 2. Psychometric functions for instrument (square symbols), octave (circular symbols), and note identification (triangular symbols). The left and right panels show average data for the musical and nonmusical listeners, respectively.

results, one for listeners KR and EM and another for listeners JA, BS, and AT. The average psychometric functions for these two groups of listeners are presented in Figure 2. We will refer to the two groups as musical listeners (Figure 2, left) and nonmusical listeners (Figure 2, right), inasmuch as listeners KR and EM professed to using tonic reference to perform the note-identification task and the other listeners did not. In fact, KR was the only listener with musical training.

The average data show that both groups of listeners achieve higher performance on the instrument-identification task than on the note-identification task with the briefer stimuli, and approximately the same level of performance on the two tasks with the longer stimuli. Instrument-identification performance is essentially independent of the number of cycles for both groups of listeners, but the level of performance is about 10% higher for the musical listeners. Note identification is at chance with one-cycle stimuli for both groups of listeners. Thereafter, as the number of cycles increases, the performance of the musical listeners rises rapidly to the level of their instrument identification, whereas for the nonmusical listeners, note-identification performance rises slowly, eventually asymptoting near the level of instrument identification. Finally, performance on the octave-identification task rises gradually from just over 40% to just under 80% as the number of cycles increases from 1 to 64, and the form of the function is the same for the two groups of listeners.

The data of all five listeners were combined to perform analyses of variance, separately, for the three response tasks. The analyses confirm that there was no significant effect of cycles on instrument identification [$F(6,21) < 1$, n.s.J, whereas there was an effect of cycles on octave identification [$F(5,30) = 22.63$, $p < .0011$, and on note identification [$F(5,30) = 43.35$, $p < .001$]. The data of listeners AT and BS for instrument identification at one and two cycles in the randomized condition were irretrievably lost in a computer disk failure. As a result, the mean performance for listeners EM and KR in the randomized condition was used to estimate group instrument identification at one and two cycles in the analysis of variance.

Despite the individual differences, then, it appears that sufficient timbre information can be extracted from one- and two-cycle stimuli to support instrument

identification at its asymptotic level, whereas eight or more cycles are required to support asymptotic pitch-chroma performance.

INTERACTIONS AND CONFUSION MATRICES

Instrument Identification

The data for instrument identification were separately analyzed to assess the interaction of instrument with octave and note value. The analysis of variance revealed that instruments were harder to identify in Octave 4 than in the lower octaves; there was a main effect of octave in both versions of the experiment [Blocked: $F(3,63) = 37.75$, $p < .001$; Randomized: $F(3,36) = 41.95$, $p < .001$]. Mean performance in Octave 4 was 54%, whereas performances in Octaves 1, 2, and 3 were 77%, 78%, and 74%, respectively. Post-hoc comparisons also showed that the higher the note of instrument in Octave 4, the more difficult it was to identify. As the note increased from C4 to D4, performance decreased from 57% to 50% correct.

Performance for instrument identification in the blocked version of the experiment is presented in Figure 3 as a function of the number of cycles. The pattern is the same for the random version of the experiment. The figure shows that instrument identification was more difficult in Octave 4, and asymptotic performance requires more cycles in Octave 4. The effect of duration on instrument identification at high fundamental frequencies was confirmed by an interaction between octave, note, and cycles [F(54,189)1.68, p < .0 1]. The effect of increasing fundamental frequency on instrument identification might be explained by the relatively low cutoff frequency the anti-aliasing filters (4.2 kHz).

Octave Identification

The timbre of the sound affected octave identification, and this effect was significant for both versions of the experiment [$F(3,19) = 15.19$, p .001]. Octaves were better identified for brass, harpsichord, and string stimuli than for flute stimuli. An examination of the confusion matrix showed that octave errors for the harpsichord were more likely to be at a higher octave, whereas octave errors for the flute were more likely to be at a lower octave. This is presumably because there is relatively more high-frequency energy in the harpsichord and relatively more low-frequency energy in the flute (Figure 1). This explanation is supported by the observations that the octave of the lowest flute stimulus (C1) and the octave of the highest harpsichord stimulus (F4) were identified best. These observations contributed to the significant instrument by octave by

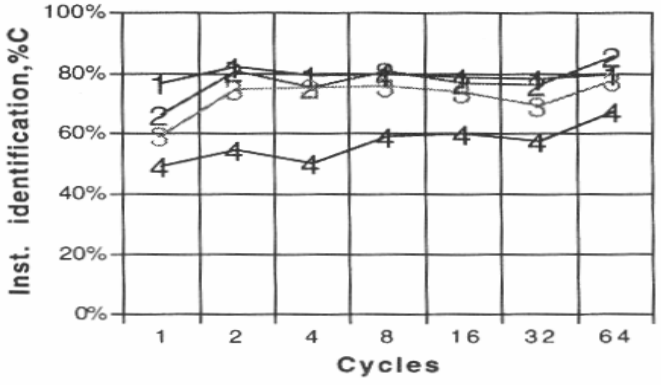

Fig. 3. Psychometric functions for instrument identification as a function of the number of cycles in Octaves 1–4 for blocked stimulus presentation.

note interaction [$F(27,162) = 1.80$, $p = .01$]. The group mean for octave identification for Flute C1 was 84% correct, whereas for Flute C2, C3 and C4 they were 51%, 54%, and 52% correct, respectively. The octave of Harpsichord F4 stimuli was identified better than the octave of Harpsichord F1, F2 and F3 stimuli. Group means were 96%, 59%, 66%, and 63% correct, respectively.

The influence of timbre on octave identification declined as the number of cycles increased; there was an instrument by cycle interaction [$F(15,90) = 2.49$, $p < .011$. The interaction was more apparent in the randomized version of the experiment, so these data were analyzed separately. The total number of "positive" and "negative" octave responses, for each instrument at each stimulus duration, is presented in Figure 4. The "+" symbols indicate that the perceived octave was higher than the stimulus octave, whereas the "—" symbols indicate that the perceived octave was lower than the stimulus octave. So, if an Octave 2 stimulus attracted an "Octave 1" response, it was coded as negative, whereas "Octave 3" and "Octave 4" responses were coded as positive. The abscissa is the number of cycles; the ordinate is the percentage of positive or negative responses for all listeners combined. The figure shows that the harpsichord, with relatively more high- frequency energy, led to more positive octave errors, and they were associated with the briefer stimuli. At the same time, the flute, with relatively more low-frequency energy, led to more negative octave errors, and they were associated with both shorter and longer durations.

Low-frequency auditory filters have narrower bandwidths than high- frequency auditory filters, and as a result, at the onset of a sound, low- frequency filters rise to their steady-state output level more slowly than high-frequency filters (for an illustration, see Fig. la of Patterson et al., 1992). This, combined with the fact that high-pass filtering a sound increases the perceived tone height of the sound (Patterson, 1990), suggests that it might be possible to explain the octave by duration interaction as follows: When the duration of a sound is short, the internal spectrum is not well defined, and in these circumstances, the center of gravity of the spectrum influences the perceived octave. As the duration of the sound increases, the resolution of the internal spectrum improves, the influence of the center of gravity decreases, and so the number of octave errors decreases. The effect extends to longer durations for the flute because longer durations are required to achieve the same degree of definition in the internal spectrum.

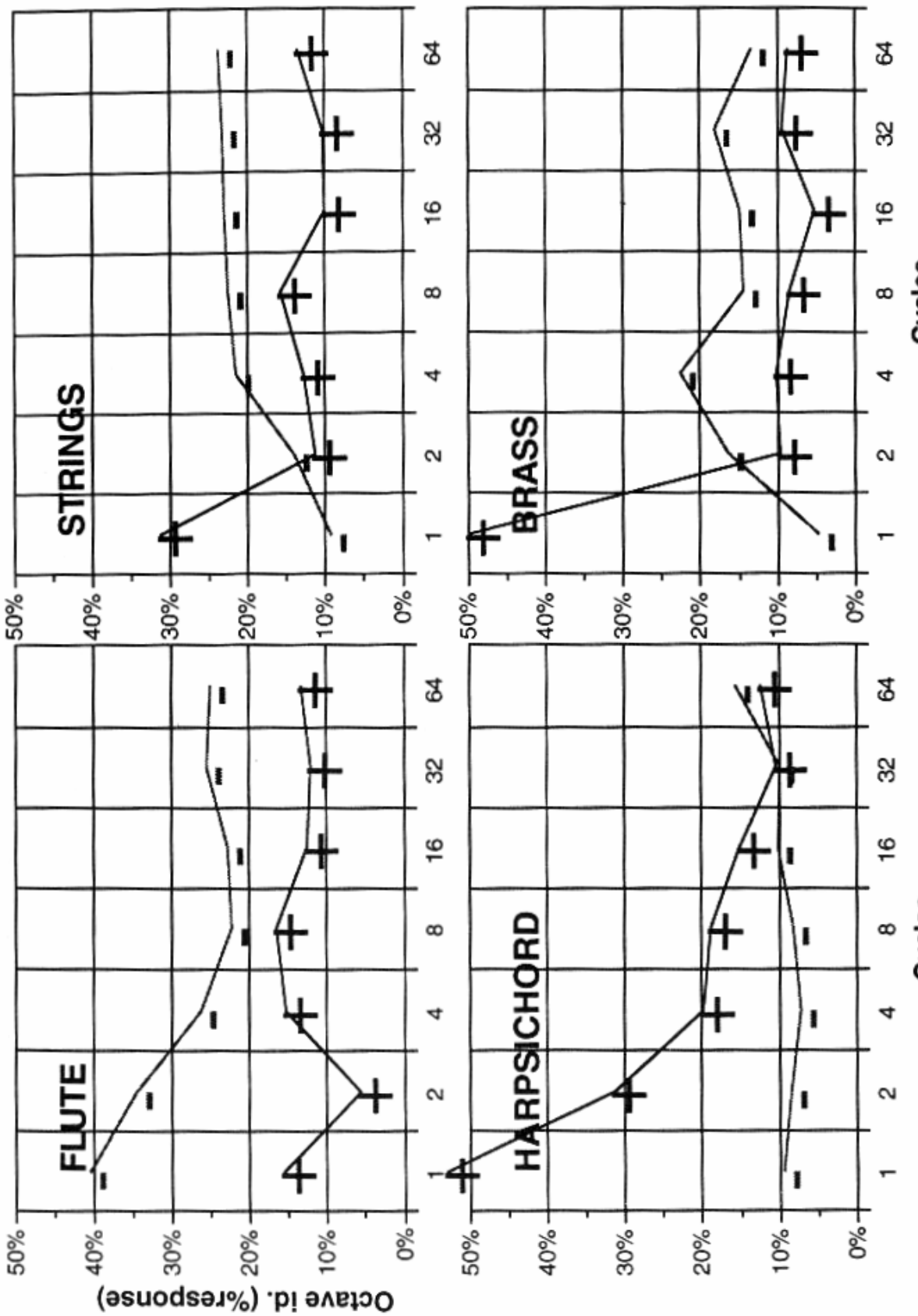

Fig. 4. Higher (+) and lower (–) octave responses as a function of the number of cycles for each instrument for random stimulus presentation.

Note Identification

The timbre of the sound was also found to interact with note identification, although the effect was not as strong as that for octave identification Notes were harder to identify when the instrument was a flute (60%) than when the instrument was a string (64%), brass (63%), or harpsichord (66%). This was confirmed in the analysis of variance as an effect of instrument on note identification [$F(3,18) = 5.58, p < .01$]. However, the confusion matrices did not reveal readily interpretable effects like those observed with octave identification

Notes presented at the lowest octave were not as well identified as notes in the higher octaves, and this was con firmed in the analysis of variance as a main effect of octave [$F(3,1\ 8) = 9.53$, $p < .001$]. Performance at Octave I was at 55% correct compared with 66%, 66%, and 67% at Octaves 2, 3, and 4. This interaction may arise because of the limit on auditory filter bandwidth at low center frequencies. The individual harmonics of the lower notes are not well resolved because the harmonic spacing is narrow relative to the auditory filter bandwidth. Finally, there was an octave by cycle interaction [$F(15,90)= 4.42$, $p <.001$], which occurred because note identification at the lowest octave did not improve as rapidly with duration as note identification in the higher octaves.

Discussion

The contrast between the experimental results at the longer and shorter durations would appear to cast considerable doubt on the concept of pitch and timbre extraction presented in recent computational models of complex sound segregation (Assmann & Summerfield 1990; Meddjs & Hewitt, 1992; Scheffers, 1983). These models assume that the auditory system can extract accurate, absolute pitch values from brief segments of multisource sounds and use these pitch values to direct the extraction of timbre information from the complex neural activity patterns flowing from the cochlea in response to these sounds.

The experimental results show that when the durations of musical notes are relatively long, listeners achieve roughly comparable performance on a four-category instrument-identification task and a four-category pitch chroma identification task. Nonmusical listeners perform somewhat less well on both the pitch and timbre tasks, and whereas the pitch-chroma performance of the musical listeners exceeds their instrument-identification performance by a little, the pitch-chroma performance of the nonmusical listeners falls just short of their instrument-identification performance. Moreover the nonmusical listeners require slightly longer stimuli to achieve asymptotic performance than the musical listeners. Nevertheless, at long durations, the differences between instrument and note identification are relatively small. As the number of cycles in the notes decreases below about eight cycles, performance on the pitch-chroma task falls off markedly, reaching chance performance with one-cycle notes, whereas there is no significant decrease in performance on the instrument-identification task even with one-cycle notes. On the basis of these results, it would seem reasonable to postulate that timbre information extracted from the initial segment of a sound might be used to assist pitch extraction, but it seems unreasonable to postulate the reverse within the relatively constrained frameworks of existing computational models of source segregation.

In all likelihood, auditory extraction of pitch and timbre information is much more complicated than it is portrayed in existing computational models of sound segregation. It may be that in the auditory system, the extraction of pitch and timbre information from brief stimuli involves different processes than those used to extract pitch and timbre from extended stimuli. This would certainly explain the differences between performance on the pitch and timbre tasks at shorter and longer durations. It could also be the case that the same pitch mechanism is used at short and long durations and that the values extracted from brief stimuli have sufficient accuracy to support good pitch-chroma performance, but there is some limitation in the more central, note-naming process that requires the pitch values to be available for longer than they are with brief stimuli. The results of the current experiment would be compatible with this interpretation as well. But in the absence of explicit models of these forms, the argument will not be pursued.

Yet another solution might be pursued within the framework of the more extensive virtual-pitch model proposed by Terhardt (1987). In this model, the processing of pitch and timbre begins with the extraction of individual spectral pitches, which are then combined to determine the virtual pitch and the timbre of the sound. The accuracy of spectral pitch values improves with duration at the start of a sound in any model, so it may well be that these initial estimates have sufficient accuracy to specify the timbre of the sound for the instrument-identification task, and at the same time, insufficient accuracy to specify the virtual pitch of the sound for the pitch chroma task. In this case, the influence of pitch on timbre extraction would be limited to the influence of spectral pitch values, since the virtual pitch is assumed to require the resolution associated with longer duration stimuli. So this would not solve the problem for the computational sound segregation models, because they use the analog of virtual pitch to guide vowel segregation. Nevertheless, the example illustrates how more complicated auditory models might deal with the limitation that the pitch-chroma data would appear to impose.

Conclusions

The experiments demonstrate that listeners can perform four-category instrument identification given one-cycle segments of musical notes and that performance is virtually independent of stimulus duration out to 64 cycles. In Contrast, four-category pitch-chroma identification is at chance levels with 1-cycle segments of the sounds, and 4—32 cycles are required to achieve asymptotic performance on the task, depending on the musicality of the listener. This suggests that the extraction of pitch information is not a prerequisite for the extraction of timbre information, at least not in the way assumed in existing auditory models of concurrent sound segregation (Assmann & Summerfield, 1990; Meddis & Hewitt, 1992; Scheffers, 1983). With the longer sound segments, average pitch-chroma identification is the same as average instrument identification, so it is not simply the case that the chroma identification is inherently more difficult than instrument-identification on.

Four-category octave identification with the same sounds is well above chance performance with I-cycle stimuli (40% correct) and rises gradually to an asymptotic level near 80% correct as the number of cycles increases to about 16 cycles. Unlike instrument and chroma identification, the psychometric functions for octave identification had the same form for both musical and nonmusjcal listeners.

With the shorter-duration stimuli, octave identification was affected by the timbre of the stimulus. The harpsichord, with relatively more high- frequency energy, was often perceived as higher in tone height than its presented octave, whereas the flute, with relatively more low-frequency energy, was often perceived as lower in tone height than its presented octave.'

References

Assmann, P. F., & Summerficid, A. Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. Journal of the Acoustical Society of America, 88: 680—97.

Duifhuis, H., Willems, L F., & Sluyrer, R. J. (1982). Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception. Journal of the Acoustical Society of America, 71, 1568—1580.

Gray, G. W. (1942). Phonemic microtomy: The minimum duration of perceptible speech sounds. Speech Monographs, 9:75—90.

Henning, G. B. (1967). Frequency discrimination in noise. Journal of the Acoustical Society of America, 41, 774—777.

1. The experiments were performed while Ken Robinson was a doctoral student at the MRC Applied Psychology Unit (Robinson, 1993). The work was supported by a grant from DRA Farnborough (project AAM HAP, 2239). The authors would like to thank E. Tcrhardt and Peter Assmann for helpful comments on an earlier draft of the paper.

Licklider, J. C. R. (1979). A duplex theory of pitch perception. In E. D. Schubert (Ed.), Psychological acoustics. Stroudsburg, PA: Dowden, Hutchinson, & Ross Inc. (Reprinted from Experientia, 1951, 7, 128—133)

Meddis, R., & Hewitt, M. J. (1992). Modeling the identification of concurrent vowels with different fundamental frequencies. Journal of the Acoustical Society of America, 91, 233—245.

Patterson, R. D. (1987). A pulse ribbon model of peripheral auditory processing. In W. A. Yost & C. S. Watson (Eds.), Auditory processing of complex sounds (pp. 167—179). Hilisdale, NJ: Lawrence Eribaum Associates.

Patterson, R. D. (1990). The tone height of multi-harmonic sounds. Music Perception, 8, 203—214.

Patterson, R. D., Milroy, R., & Allerhand, M. (1993). What is the octave of a harmonically rich note? In I. Cross (Ed.), Proceedings of the 2nd International Conference on Music and the Cognitive Sciences (pp. 69—8 1). London: Harwood.

Patterson, R. D., Peters, R. W., & Milroy, R. (1983). Threshold duration for melodic pitch. In R. Klinke & W. Hartmann (Eds.), Hearing: Physiological bases and psychophysics (pp. 321—325). Berlin: Springer.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). Complex sounds and auditory images. In Y. Cazals & L. Demany (his.), Auditory physiology and perception (pp. 429—446). Oxford: Pergamon.

Robinson, K. L. (1993). Studies in timbre and pitch. Unpublished doctoral dissertation, University of Cambridge.

Scheffers, M. T. M. (1983). Sifting vowels. Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Suen, C. Y., & Beddoes, M. P. (1972). Discrimination of vowel sounds of very short duration. Perception & Psychophysics, 11, 417—419.
Terhardt, E. (1987). Psychophysics of audio signal processing and the role of pitch in speech. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. Leiden: Martinus Nijhoff.
Whitfield, I. C. (1979). Periodicity, pulse interval and pitch. Audiology, 18, 507—512.