/ 1 , 06

6.8

Received 27 July 1970

Effect of Material on Flute Tone Quality

JOHN W. COLTMAN

3319 Scathelocke Road, Pittsburgh, Pennsylvania 15235

Three keyless flutes of identical internal dimensions were made of thin silver, heavy copper, and wood, respectively. They were played out of sight of musically experienced observers, who were asked to determine only whether tones were alike or different. No statistically significant correlation between the listeners' scores and the material of the instrument body was found. Flutists who played the flutes, using an arrangement to eliminate visual or tactile clues, were unable to identify again a previously selected instrument.

INTRODUCTION

The role that the wall material plays in determining the tone quality of wind instruments has long been a subject of argument. Laboratory measurements of sustained tones in artificially blown instruments[1,2] generally show no evidence that the wall material has an appreciable effect. In spite of these experiments, instrument makers, players, and listeners continue to insist that the nature of the wall material does have an effect on the instruments' sound. Metal clarinets are considered suitable only for use in school bands. The silver flute is the accepted standard in most countries today, though symphony players insist on retaining wood for piccolos. Not only the nature of the material, but the thickness of the wall is considered important. Many makers of quality flutes strive to make the wall as thin as is consistent with mechanical requirements. At the same time, high density is considered desirable. Varese wrote a composition entitled "Density 21.05" to celebrate the platinum flute played by America's most famous flutist, Georges Barrere.

Thus, there is a marked discrepancy between the results of laboratory measurement and the opinions of users. It has become apparent[3] that harmonic content of steady tones is not sufficient to characterize an instrument and that listeners generally depend on transitory effects for identification. Moreover, from the player's standpoint, the "responsiveness" of the instrument, that is, the ease with which sounds can be produced and modulated, is an important aspect that may be independent of the character of the resultant sound. This also brings into question the pertinence of laboratory measurement of steady tones.

In all of these discussions (which, as Backus[4] points out, probably started in early Stone Age circles with assertions that a flute made from a human thigh bone had a much better tone than one made from a stick of bamboo) there is an almost complete absence of reports of controlled objective experiments directed toward determining the degree to which listeners and players could, in fact, discriminate between instruments made of various materials. It therefore seemed worthwhile to carry out the two experiments described below, which attempt to elicit evidence toward this question, both for listeners and players.

I. CONSTRUCTION OF THE FLUTES

Three keyless flutes were constructed of silver, copper, and wood. These were all of internal diameter 1.90 cm. The silver tube was a piece of stock from which professional silver flutes are made and had a wall thickness of 0.036 cm. The copper-tube wall thickness was 0.153 cm, making it more massive than the silver tube by a factor of 4.0. The third tube was of grenadilla, a wood frequently employed for woodwinds. The wall thickness was 0.41 cm, a typical dimension for wooden flutes. It weighed 1.7 times as much as the silver tube. Each tube was 32.7 cm long and had affixed to it a short head of Delrin plastic extending another 5.1 cm to the center of the mouthhole. The three plastic heads were shaped to the same internal dimensions by a specially made tapered reamer. The diameter at the mouthhole was 1.75 cm. The mouthhole itself was shaped by using a paraffin impression of the mouthhole of a professional model flute and casting epoxy resin around this plug to fill in a larger hole previously bored

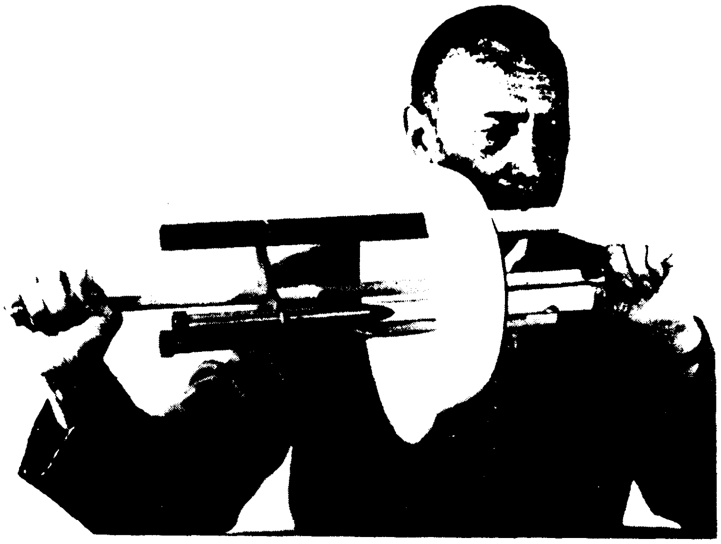

FIG. 1. Flutes of silver, copper, and wood as supported for playing.

in the head. The far ends of the silver and copper tubes were enlarged in thickness by a short outer cylinder to make the outer diameter of the ends equal to that of the wooden tube. The acoustically important dimensions of all three instruments were thus identical within 0.01 cm. The instruments' first mode sounded 398 Hz, approximately G_4, the middle of the orchestral flute's scale. The ease of sounding and the power and tone were judged excellent by flutists who tried them.

II. EXPERIMENT WITH LISTENERS

The first experiment was directed toward finding out whether listeners could discriminate among the instruments when they were played by the same performer.

Twenty-seven observers participated simultaneously in the listening trials. Of these, 20 claimed to be professional or skilled amateur musicians, 13 of them flutists. Seven claimed little training or experience. The trials were conducted with the author playing the flutes behind a small screen made of aluminized mylar 750 μm thick, whose sound transmission extends well beyond 15 kHz. The tests were conducted in a classroom that was acoustically treated and often used for chamber music performances.

In each trial three identical musical phrases were played. Two of these were played on the same instrument. The observers were asked to mark the position where the different instrument was used. This technique was chosen because it avoided any reference to "better" or "worse" tone quality, and did not rely on any presumptions as to how a wood or silver flute should sound.

Thirty-six such trials were made. In the first six trials the phrase consisted simply of a single tongued and sustained note, the first mode of the instrument. Only the silver and wood flutes were used, the order of playing in a single trial being varied randomly. In three cases the wood flute was unique, in three cases the silver. A similar set of six trials compared the silver and copper flutes. The observers were told which pair was being compared in each set of six trials. The third and fourth groups were similar except that the second mode (approximately 800 Hz) was sounded, while the phrase for the fifth and sixth groups consisted of a legato transition from G_4 to the octave above, and return. Additional information collected from the observer was his musical background as a flutist, other musician, or listener and his own rating of himself as a professional, skilled amateur or student, or person with little training or experience.

The results obtained may be examined statistically in many ways. For completeness, the results are given in two tables. Table I lists the individual scores (number of trials out of the entire 36 in which the observer identified correctly the unique instrument) in two groups, one for those who claimed to be skilled or trained muscially and one for those of little training or experience.

On a purely random hypothesis, one expects success on one-third of the trials, so the expected score is 12.

TABLE I. Scores of 27 observers for the entire set of 36 trials.

		Mean score
Musically skilled observers	15, 16, 10, 11, 12, 11, 14 15, 13, 18, 13, 7, 14, 14 12, 10, 10, 14, 13, 16	12.9±0.5
Unskilled observers	17, 18, 13, 14, 15, 14, 8	14.1±1.1
Total of 27 observers		13.1±0.5

TABLE II. Distribution of votes on 36 trials. W, S, and C indicate the position of the flute played once during a trial and whether it was wood, silver, or copper.

	Silver vs wood			Silver vs copper		
Fundamental	3	S9	15	C8	8	11
	W8	13	6	10	S10	7
	S7	9	11	6	C13	8
	4	5	W18	5	4	S18
	9	W6	12	S10	10	7
	12	9	S6	5	12	C10
Octave	6	S8	13	C7	11	9
	8	10	W9	7	13	S7
	S11	7	9	7	10	C10
	2	W11	14	9	S12	6
	W10	7	10	S5	9	13
	5	14	S8	12	C9	6
Combination	7	7	W13	S15	9	3
	9	S8	10	C7	12	8
	10	7	S10	5	S15	7
	W10	4	13	2	6	C19
	S5	12	10	9	C7	11
	11	W7	9	9	7	S11

The standard deviation of a single score should be 2.8 and the mean for 27 observations 12 ± 0.5. We see from Table I that the unskilled observers scored slightly (but insignificantly) higher than the skilled observers. The over-all mean score has the expected standard deviation, implying that no outstandingly successful observers were present, while the mean score of 13.1 is slightly higher than the expected value of 12.

The nature of the trials, however, was such that statistical independence of the 972 observations is not assured. Each of the 27 observers was listening to the same set of trials. If the performer, for example, were accidentally to play one note poorly, all 27 observers might be expected to score this trial the same. The statistical expectation that a playing variation accidentally corresponds with the unique flute is again one-third, but now the standard deviation is much higher, since there are only 36 rather than 972 trials. Thus we need to look in more detail at the correlations between markings before deciding whether the slight departure from 12 is significant.

Table II displays the results of the tests. Each row of three numbers represents a trial. The letter preceding a number designates the position of the unique flute for this trial and its material. The numbers are the number of persons in the 27 observers voting for the corresponding position.

If we total the columns, we find that observers cast 275 votes for the first instrument played, 330 for the second, and 367 for the third. The spread in these values is more than five standard deviations from the expected 324. Since the unique flute appeared an equal number of times in each column, this is not a result of the material. It strongly points to the fact often observed that the order of playing is important in determining an observer's choice; Saunders,[5] for example, found in listener tests of preference for Stradivarius violins that the jury voted most often for the second instrument played, regardless of its type.

By applying a chi-squared test to the entire set of data, using a value of nine as the expected value, and assigning 2 degrees of freedom (df) to each trial, a strong rejection ($p<0.0005$) of the random hypothesis is obtained. Thus the votes are grouped more strongly than one could expect from pure chance. However, if we take the majority vote at each trial as the observer's selection (counting ties as one-half), we find the observers as a jury voted correctly in just 12.5 out of 36 trials. Thus the grouping is not a result of the material. It is likely that, in addition to the effect of order, perceptible variations in playing contribute to the grouping. This is supported by a comparison of pairs of trials in which the position of the unique flute was the same. All observers questioned admitted to great difficulty in making a selection in almost all cases, so that these playing variations were not large. They prevent us, however, from applying a simple Bernoulli test on the over-all scores.

An analysis of variations carried out with respect to order of playing, phrase used, copper versus wood, and their interactions showed none of these to be significant, at the 5% level, in determining whether an observer correctly identified the unique instrument.

III. EXPERIMENT WITH PLAYERS

In order to obtain an objective measure of a player's ability to discriminate between instruments, it is necessary to remove any clues to its identity other than those associated with the sound produced. This was accomplished by the arrangement shown in Fig. 1. The three flutes described in Sec. I were mounted symmetrically with their axes parallel to a central rod and their heads projecting through a plastic shield affixed to the same rod. The player grasped the rod rather than the flutes, and when his head was in the normal blowing position, he could not see the tubes forming the bodies of the instruments. A counterweight inside the shield restored the imbalance due to the differing weights of the instruments, so that in all respects visual and tactile clues to identity were removed.

In making a test, the subject was permitted to blow the instruments in any order, rotating the fixture to change flutes, but otherwise not changing position. He was asked to select an instrument whose tone he preferred (or thought he could identify again), and the experimenter noted the selection. The subject was then told to spin the fixture so as to lose the identity, and then, by blowing at will on various instruments, find his original choice. When he was satisfied he had found it, he notified the experimenter, who scored one point for success or zero for failure.

A test consisted of five attempts to find the originally selected instrument, so that test scores can vary from zero to five. In the first test, the flutist was restricted to sounding the fundamental, and in the second the octave, but he could tongue or attack the note at will.

Nine such tests were carried out with four different flutists, all reasonably skilled performers. The four flutists scored, as follows:

4, 0, 1; 0, 2; 0, 2; 2, 1.

On a random hypothesis, the expected score is $5\times\frac{1}{3}=1.7$, and the mean of nine such scores would have a standard deviation of 0.35. The experimental results give a mean score of 1.3, barely one standard deviation less than the expected mean. We can expect a score of four or higher to occur in one test out of 22; its occurrence in a set of nine is not surprising. The flutist who scored this four got scores of zero and one on the next two tests.

If we assume that not only the first selection, but also the flutes named as "identical" constitute "preferred" flutes, we find silver named 21 times, wood 19, and copper 14. On a random hypothesis, we expect a value of 18 in each category. A chi-square test on the observed numbers produces a value for chi-square of 1.44, well below the 5.99 required for significance with 2 df. There is thus no evidence here of anything but random selection.

IV. CONCLUSION

No evidence has been found that experienced listeners or trained players can distinguish between flutes of like mouthpiece material whose only difference is the nature and thickness of the wall material of the body, even when the variations in the material and thickness are very marked. Of course, it is possible that individuals exist whose discriminatory senses are keen enough to find a distinction, but if so, they are certainly not common. Moreover, the results suggest that even careful attempts to produce identical sounds on the same instrument produce variations that are more perceptible than any that might be associated with the material.

One player did, correctly, point out that one of the three instruments appeared at first to be slightly flat. This effect is due to the high thermal mass of the heavy copper tube, which causes it to warm up more slowly than the others. This is an example of a reason to prefer certain materials for flute construction, and there are many others. Tone quality or ease of response are not, however, among them.

ACKNOWLEDGMENT

Thanks are due to Robert Hooke of the Westinghouse Research Laboratories for assistance in design of the experiments and in the statistical analysis of the results.

[1] J. Backus, J. Acoust. Soc. Amer. **36**, 1881–1887 (1964).
[2] J. Backus and T. C. Hundley, J. Acoust. Soc. Amer. **39**, 936–945 (1966).
[3] J.-C. Risset and M. Mathews, Phys. Today **22**, 23–33 (1969).
[4] J. Backus, *The Acoustical Foundations of Music* (Norton, New York, 1969), p. 208.
[5] F. A. Saunders, Sound **1**, 7–15 (1962).