On the Perception of Phase Distortion*

HIDEO SUZUKI, SHIGERU MORITA, AND TAKEO SHINDO

Consumer Products Research Laboratory, Mitsubishi Electric Corporation, Kamakura, Kanagawa, Japan

Differences in the sound qualities of linear- and nonlinear-phase loudspeakers are evaluated by simulating the phase responses of loudspeakers using all-pass filters. Test results indicate that the perception of phase distortion is highly dependent on individual ability, and that it is easier to detect phase distortion by headphone listening rather than by loudspeaker listening. Instantaneous frequency and instantaneous envelope are tentatively introduced as representations of a transient signal to give some indication of how transient signals sound.

0 INTRODUCTION

For many years, phase perception has been the subject of a controversy among acoustic engineers. Some argue that phase shift is essentially immaterial, while others believe that phase response is as important as amplitude response. It seems to us that those who neglect the importance of phase response are referring to the type of phase changes that occur gradually over a wide frequency region. In contrast, most of the phase-oriented people, in testing, use synthetic signals that undergo large phase changes within a relatively narrow frequency band. We think that both opinions are correct, and the important thing is the degree to which the sound quality is affected by the phase response. This should be investigated quantitatively and systematically for many kinds of phase responses and sound sources. Quite a few studies have been done on the perception of phase changes [1]–[7]; so we would like to lay stress on the effect of the phase response of a loudspeaker.

Actual loudspeakers, even the so-called linear-phase loudspeakers, have nonlinear phase responses. The object of our study is to determine whether these phase distortions have any effect, and if so, to what extent do they degrade high-fidelity reproduction. The variety of signals and phase responses make our problem more difficult to deal with quantitatively. Using all-pass filters is one way to meet this problem, because all-pass filters have constant amplitude responses and can simulate the phase responses of loudspeakers.

* Manuscript received 1979 March; revised 1980 February.

1 PHASE AND GROUP DELAY OF SINGLE-POLE FILTERS

The all-pass filter, one of the nonminimum-phase networks, is a very convenient device for studying the effect of phase response on sound quality because the phase response can be changed independent of the amplitude response. Generally the transfer function $H(\omega)$, with amplitude response $A(\omega)$ and phase response $\phi(\omega)$, is represented by the following equation:

$$H(\omega) = A(\omega) \cdot e^{j\phi(\omega)}. \tag{1}$$

For all-pass filters $A(\omega)$ is a constant such that $A(\omega) = A_0$. The group delay t_g of Eq. (1) is defined as [8]:

$$t_g = -\frac{\partial\phi(\omega)}{\partial\omega}. \tag{2}$$

Single-pole all-pass filters composed of active elements are shown in Fig. 1. The transfer functions $H_A(\omega)$ and $H_B(\omega)$ for types A and B are given, respectively, by

$$H_A(\omega) = 1 \cdot e^{j\phi_A(\omega)} = e^{j\{-2\tan^{-1}(f/f_0)\}} \tag{3}$$

$$H_B(\omega) = 1 \cdot e^{j\phi_B(\omega)} = e^{j\{\pi-2\tan^{-1}(f/f_0)\}} \tag{4}$$

where f_0 is the 90-degree phase shift frequency given by the equation

$$f_0 = 1/2\pi R_0 C_0. \tag{5}$$

Eqs. (3) and (4) indicate the phase difference of π. The group delays of $H_A(\omega)$ and $H_B(\omega)$ are given by the same equation:

$$t_{g\,A,B} = 1/\pi f_0\{1 + (f/f_0)^2\}. \qquad (6)$$

For small $f\,(<<f_0)$, t_g is constant with $1/\pi f_0$ and gradually decreases to zero as the frequency increases. The phase responses $\phi_A(\omega)$ and $\phi_B(\omega)$ and the group delay t_g are shown in Fig. 2.

2. SIMULATION OF LOUDSPEAKER PHASE RESPONSE BY ALL-PASS FILTERS

Now we will show how we can simulate the phase response of a loudspeaker using single-pole all-pass filters. The phase response of a two-way loudspeaker system is shown in Fig. 3, curve a. The fundamental resonance frequency of a woofer with an enclosure is 65 Hz, while the tweeter is connected in reversed polarity at a crossover frequency of 1.5 kHz. The phase response of a phase shifter consisting of two single-pole all-pass filters connected in series is shown in Fig. 3, curve b. Here f_0 of the phase-lead type is 70 Hz, and that of the phase-lag type is 2 kHz. The phase responses of an actual loudspeaker system and the phase shifter are in good agreement. This loudspeaker can be made phase linear by the conventional linear phase technique, resulting in the response shown in Fig. 3, curve c. This phase response is the one of a single-pole all-pass filter of the phase-lead in which f_0 = 70 Hz. It is clear, in this case, that the difference between the so-called linear- and nonlinear-phase loudspeaker systems exists only if we have a single-pole all-pass filter of the phase-lag type with f_0 = 2 kHz. Therefore the advantage of linear-phase response can be evaluated by checking the effect of this all-pass filter on sound quality.

3. TRANSIENT SIGNALS FOR HEARING TEST

Our preliminary tests indicated that single-pole all-pass filters do not affect the sound quality of such continuous signals as rectangular or sawtooth waves where the subjects and the setup are the same as for the experiment to be described in Section 4. Therefore we used the transient signals of short duration as shown in Fig. 4, where they are designated as S1, S2, R1, R2, and T1, T2. The time interval of one cycle T_0 for each signal was chosen such that $T_0 = 2/f_0$, where f_0 is the 90-degree phase shift frequency of the phase-lag type all-pass filter defined by Eq. (5).

What we actually hear is a sound, and sound signals converted from electrical signals by a loudspeaker or a headphone have different waveforms. These waveforms are shown in Fig. 5 for S2, R2, and T2, including the electrical output of an amplifier. The unfiltered waveform is at the top in each block, the filtered waveform is at bottom. The loudspeaker used for the hearing test was a bass-reflex two-way system with a 30-cm woofer and a 5-cm tweeter. Loudspeaker measurement was made in a listening room with a reverberation time of 0.3 second, and a dummy head was used for the headphone measurement.

4 EXPERIMENT

Fig. 6 shows a schematic diagram of the instrumentation. The transient signals generated by a function generator are phase shifted by the phase-lag type all-pass filter. Switch positions 1 and 2 give the unfiltered and filtered signals, respectively. Intervals of the transient signals were changed according to the frequencies of 300 Hz and 1 kHz selected for f_0 in our experiment (see Fig. 4). The unfiltered and filtered signals were given to the subjects through a loudspeaker or a headphone in either a listening room or an anechoic chamber. At their peaks, the levels of the transient signals measured by a B&K 2606 measuring amplifier were about 100 dB sound pressure level at the listening position. The loudness of the headphone test was about the same as that of the loudspeaker test. All the subjects participating in our experiment were members of the audio engineering group of our laboratory, with ages ranging from 25 to 35.

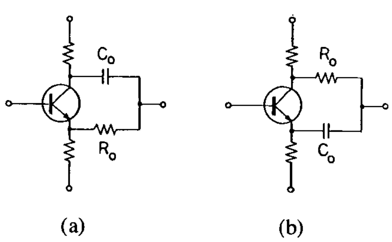

Fig. 1. Single-pole all-pass filters composed of active elements. (a) Phase lag. (b) Phase lead.

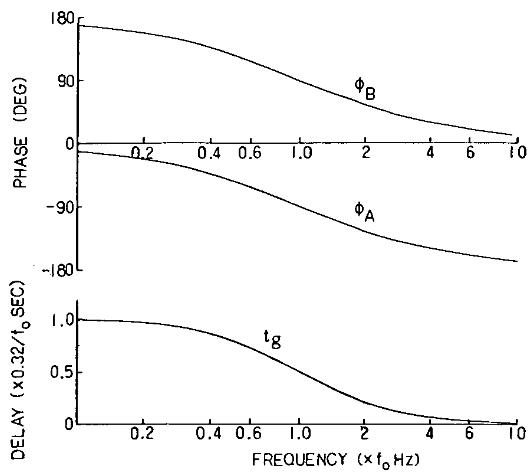

Fig. 2. Phase responses and group delays of single-pole all-pass filters.

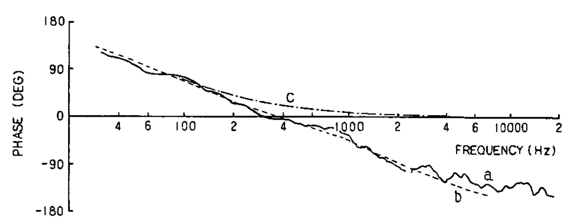

Fig. 3. Phase responses of an actual loudspeaker and all-pass filters. a—Two-way loudspeaker system, tweeter is reversed in polarity; b—combination of phase-lead and phase-lag type all-pass filters; c—phase-lead type all-pass filters.

They were not tested for normal hearing, but all of them had enough experience to check the sound qualities of audio equipments.

There are four combinations of unfiltered signals (position 1) and filtered signals (position 2): 1 – 1, 1 – 2, 2 – 1, and 2 – 2. For each combination the first signal, A, was given to the subjects five times with 1-second intervals. After this, with the same intervals of 1 second, the second signal, B, was given five times. This process was repeated five times. We thus have a row of 50 transient signals which is represented as A·A·A·A·A·B·B·B·B·B·A·A·A·A·A ··· B·B·B·B·B, where A and B denote the first and second signals, respectively, and · stands for the 1-second pause between signals. It was necessary to present the same signal several times consecutively, because only one presentation of a transient signal with very short duration was not enough for the subjects to grasp and keep in mind its sound quality. Also a number of comparisons were necessary to make a decision for one specific combination because the difference of the sound qualities between filtered and unfiltered signals was so small.

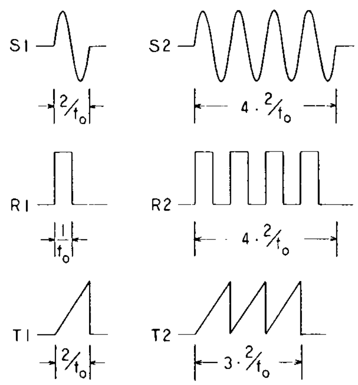

Fig. 4. Transient signals used for hearing test.

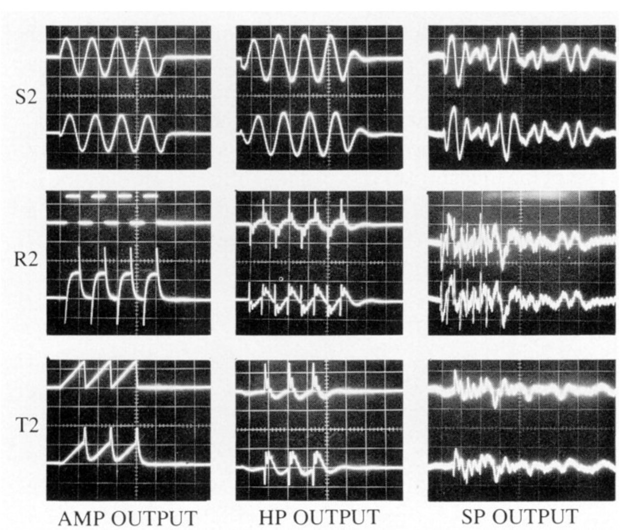

Fig. 5. Waveforms of amplifier (AMP), headphone (HP), and loudspeaker (SP) output for S2, R2, and T2. Units of the horizontal axes are 4.4 ms for amplifier and headphone and 8.9 ms for loudspeaker.

Then each subject was asked whether the first and second signals sounded the same or different. If the listener judged sequences 1 – 1 and 2 – 2 as the same and sequences 1 – 2 and 2 – 1 as different, the score was counted correct. From 50 judgments the percentage of correct answers was obtained for each subject. If a subject could detect the effect of phase shift perfectly, the score would be 100 percent. On the contrary, if the listener could not distinguish any difference, the score is about 50 percent. Our impression from the experiment is that a percentage of correctness in the low 70 percent is a reasonable line for separating subjects who can distinguish phase shift from those who cannot. (If this experiment had been conducted by the normal A – B – A paradigm, the results would have been much worse.)

5 RESULTS AND DISCUSSIONS

The percentages of correct answers for each subject are shown in Fig. 7, where $f_0 = 300$ Hz, and a loudspeaker in a listening room was used. Subject A scored more than 75 percent for every signal tested. Other figures above 75 percent were scored by subject B for S2 and T2, by subject C for T1 and by subject H for S2. The scores of the other subjects are lower than 75 percent for all signals. Widely distributed figures show the important fact that the sensitivity to the change of phase response is highly individual. To judge from the mean values, S2 seems to be the easiest signal when detecting changes in sound quality.

The results conducted in an anechoic chamber are shown in Fig. 8, where $f_0 = 300$ Hz, and signals S2 and R2 were used. There seems to be no meaningful difference between the results of Fig. 7 and those of Fig. 8, indicating that the reverberation of the listening room does not have much

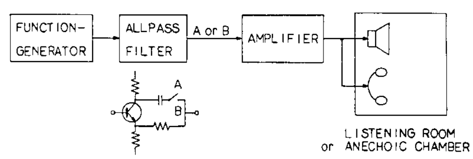

Fig. 6. Schematic diagram of instrumentation.

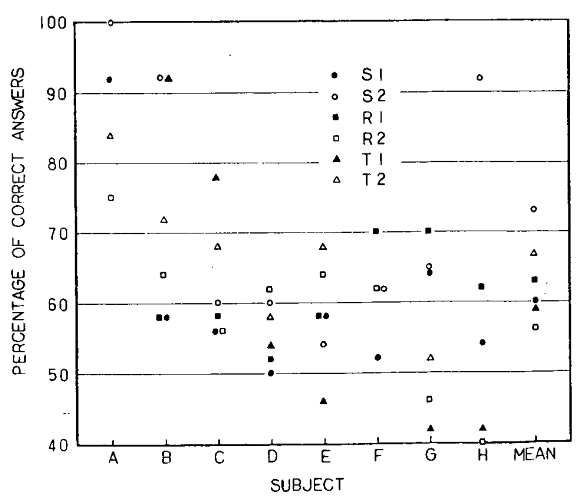

Fig. 7. Percentages of correct answers of loudspeaker listening for various signals in a listening room, where $f_0 = 300$ Hz.

effect on phase perception. The effect of reverberation may also be discussed comparing the results of loudspeaker and headphone listening. Fig. 9 illustrates the results of headphone listening for $f_0 = 300$ Hz. Headphone listening shows much greater sensitivity than loudspeaker listening. The reason for this may be that the headphone sound has greater detail, which is lost in loudspeaker reproduction because the difference of the frequency response rather than because of the room reverberation.

Further results of headphone listening are shown in Fig. 10, where the frequency f_0 was raised to 1 kHz and, accordingly, the intervals of signals were shortened to 3/10. It is evident that even for headphone listening the phase shift of Fig. 2 when $f_0 = 1$ kHz is not discernible. The effect of the all-pass filter is to give a frequency-dependent time delay to the system, but the resolution of the ear in the time domain may not be enough for the perception of this delay. For $f_0 = 1$ kHz the group delay is 0.32 ms in the low-frequency region.

Our major concern is to find a permissible level of phase

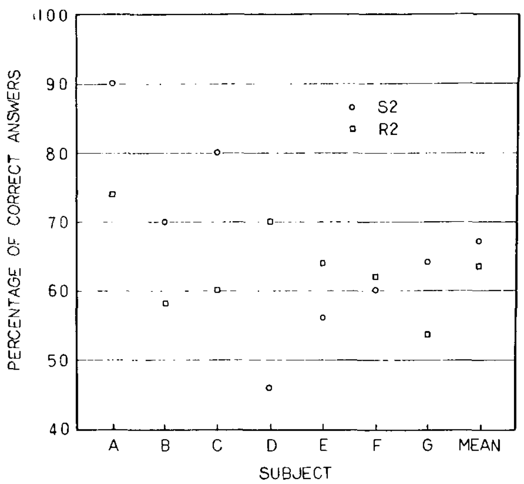

Fig. 8. Percentages of correct answers of loudspeaker listening for S2 and R2 in an anechoic chamber, where $f_0 = 300$ Hz.

distortion in loudspeaker reproduction, and from this aspect we can interpret the results of Figs. 7–10 in two ways. First, some people certainly can hear the phase distortion in the low frequencies of the type shown in Fig. 2 when highly artificial signals are used. In this sense, phase distortion is not permissible for high-fidelity reproduction. A nonlinear-phase multiway loudspeaker system that has a relatively low crossover frequency, say at several hundred hertz, may have some coloration due to the phase distortion. Second, however, as the average values indicate, most subjects were much less sensitive to the phase distortion. In fact, no one found even the slightest change in sound quality by the phase shift when popular music from several commercial disks was used for a qualitative loudspeaker listening test.

When we listen to S2 via a loudspeaker or headphone, two pitches can be discerned, a louder low pitch and a quieter high pitch. It is difficult to tell which comes first, but when we listen to a filtered signal, the higher pitch is much more distinct than that of the unfiltered signal. This may be explained as follows. When S2 is presented without the all-pass filter, the low-frequency and high-frequency components of the waveform are perceived almost simultaneously, and the low-frequency content partially masks the high-frequency components. When S2 is presented through the all-pass filter, the low-frequency components are delayed relative to the high-frequency components. This allows the high-frequency components to be perceived before the low-frequency masking starts, giving the sensation of a high-frequency tone just at the beginning of the filtered pulse. This could be further explained quantitatively by using the filter bank model [9].

Another way, we think, to successfully explain the perception of signal S2 is to use the instantaneous frequency and the instantaneous envelope. These are defined as follows. When we represent an arbitrary time function $f(t)$ by

$$f(t) = A(t) \cdot \cos \phi(t) \tag{7}$$

where $A(t)$ is called the instantaneous envelope and $\phi(t)$ the instantaneous phase, $A(t)$ and $\phi(t)$ are given, respectively,

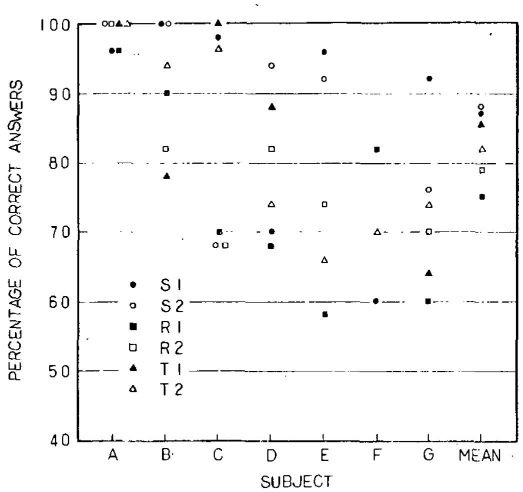

Fig. 9. Percentages of correct answers of headphone listening for various signals, where $f_0 = 300$ Hz.

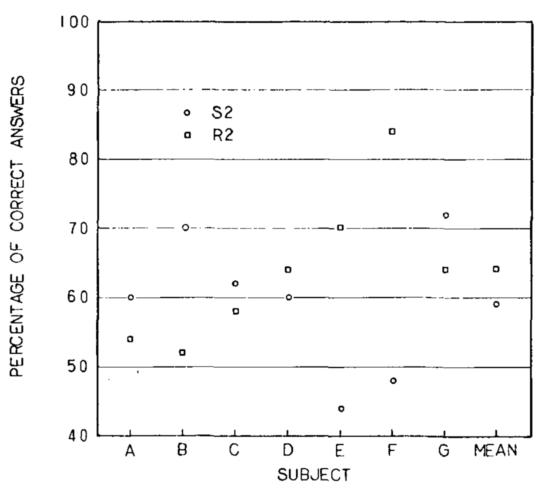

Fig. 10. Percentages of correct answers of headphone listening for S2 and R2, where $f_0 = 1$ kHz.

by

$$A(t) = \frac{f(t)}{\cos[\tan^{-1}\{-f_1(t)/f(t)\}]} \quad (8)$$

$$\phi(t) = \tan^{-1}\{-f_1(t)/f(t)\}. \quad (9)$$

Instantaneous frequency is obtained by differentiating Eq. (9) with respect to time:

$$\omega_i(t) = \frac{\{f_1(t)\cdot f'(t) - f(t)\cdot f_1'(t)\}}{\{f^2(t) + f_1^2(t)\}} \quad (10)$$

where the time function $f_1(t)$ is the Hilbert transform of $f(t)$, which is given by

$$f_1(t) = -f(t) * (1/\pi t). \quad (11)$$

where * denotes convolution.

The instantaneous frequency and the instantaneous envelope of S2 for headphone reproduction are shown in Fig. 11. The waveform was digitalized by an analog-to-digital converter with a sampling time of 50 μs and a sampling number of 1024. The instantaneous envelope is shown by relative quantity, and the negative portions of its curve were reversed in sign by program manipulation. Fig. 11 shows that the instantaneous frequency of the unfiltered S2 is about 340 Hz at the start and about 150 Hz during the remaining portion. These frequencies seem to correspond to the two pitches of the sound. When S2 is filtered, the peak of the instantaneous frequency becomes much higher at the start than for the unfiltered S2. It is reasonable to think in this case that the higher pitch will be enhanced when we listen to this filtered signal. The small value of instantaneous envelope at the time of maximum instantaneous frequency for both filtered and unfiltered S2 shows that the higher pitch will sound quieter than the lower pitch.

The above discussion may show that the instantaneous frequency and the instantaneous envelope are sometimes successful in explaining the perception of phase distortion of transient sounds.

6 CONCLUSION

For a long time too much of our attention has been directed to amplitude response, perhaps because measurements were so easy. Another reason might be that amplitude response has such an extremely large influence on the sound quality that we tend to forget the other characteristic, phase response.

Many high-grade loudspeaker systems have been designed without any considerations of the phase response. But once we have started to investigate phase response and have the tools for measurement, it is no longer easy to ignore the effect of phase response. No one can deny the necessity of linear phase response to reproduce a waveform with high accuracy. Still, we must be careful that reproduction of the waveform with high accuracy does not become the sole criterion for high-fidelity reproduction. The results obtained by our experiment show that the effect of phase change on the quality of transient sound is much smaller compared to the one we expect when we see the change of the waveform due to the phase distortion.

For music signals the quantity of phase distortions used in our experiment was not enough to be detected by any subject. It will be our future goal to find out the permissible level of phase distortion for music sources.

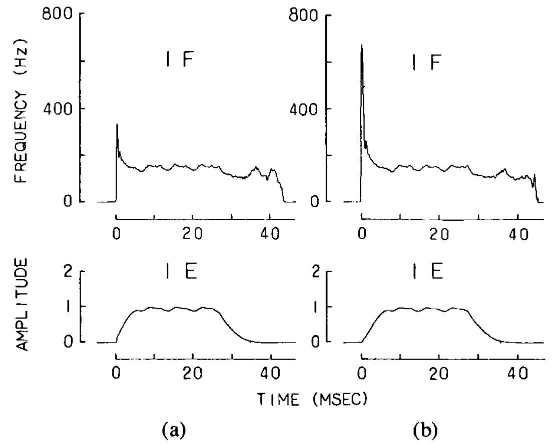

Fig. 11. Instantaneous frequencies (IF) and instantaneous envelopes (IE) for headphone listening. (a) unfiltered S2. (b) filtered S2.

7 REFERENCES

[1] H. D. Harwood, "Audibility of Phase Effects in Loudspeakers," *Wireless World* (1976 Jan.).

[2] B. B. Bauer, "Audibility of phase distortion," *Wireless World* (1974 Mar.).

[3] R. C. Cabot, M. G. Mino, D. A. Dorans, I. S. Tackel, and H. E. Breed, "Detection of Phase Shifts in Harmonically Related Tones," *J. Audio Eng. Soc.*, vol. 24, pp. 568–571 (1976 Sept.).

[4] H. von Fleischer, "Gerade wahrnehmbare Phasenänderungen bei Drei-Ton-Komplexen," *Acustica*, vol. 32, p. 44 (1975).

[5] H. von Fleischer, "Über die Wahrnehmbarkeit von Phasenänderungen," *Acustica*, vol. 35, p. 202 (1976).

[6] J. Blauert, "Group Delay Distortions in Electroacoustical Systems," *J. Acoust. Soc. Am.*, vol. 63, p. 1478 (1978).

[7] R. Berkovitz and B. E. Edvardsen, "Phase Sensitivity in Music Reproduction," presented at the 58th Convention of the Audio Engineering Society, New York, 1977 November 4–7.

[8] R. C. Heyser, "Loudspeaker Phase Characteristics and Time Delay Distortion," *J. Audio Eng. Soc.*, vol. 17, p. 130 (1969).

[9] C. L. Searle, J. Zachary Jacobson, and S. G. Rayment, "Stop Consonant Discrimination Based on Human Audition," *J. Acoust. Soc. Am.*, vol. 65, p. 799 (1979).

[10] J. B. Thomas, *Statistical Communication Theory* (Wiley, New York, 1969).

The authors' biographies appeared in the July/August issue.