Bulletin of the Psychonomic Society
1978, Vol. 11 (5), 327-329

Production of white tone from white noise and voiced speech from whisper

RICHARD M. WARREN and JAMES A. BASHFORD
University of Wisconsin, Milwaukee, Wisconsin 53201

A new class of tonal sounds can be generated by repeating brief sections of noise over and over without intervening silence. When the repeated waveform is white noise, a "white tone" with a rich distinctive timbre and no noise-like quality is heard over a considerable range of repetition rates. If the noise is a whispered vowel rather than white noise, repetition of a sample equal in duration to a single glottal pulse during voicing can generate a "whisper tone" sounding like a voiced version of the vowel. Whispered discourse can be converted to an intelligible voiced monotone by repetition of regularly spaced samples drawn from the whispered speech.

We are reporting a procedure for transforming noises into tones.

All tones are periodic, consisting of some particular pattern of sound-pressure changes repeated at a rate corresponding to the fundamental frequency. These iterated waveforms can be analyzed by human hearing (and by appropriate instruments) into a harmonic spectrum of sinusoidal tones. The spectral distribution of energy is responsible for the quality of a tone, helping us to discriminate between complex tones having the same fundamental frequency and pitch (as with different vowels and different musical instruments).

Noise lacks periodicity, and an oscilloscope tracing shows random variations in sound pressure. However, it has been known for some time that it is possible to derive sounds with some degree of pitch-like character from the hiss and rumble of broad-band noise: *narrow-band pitch* can be generated by filtering off the noise frequencies above and below the limits of the band (Rainbolt & Schubert, 1968; Small & Daniloff, 1967); *interruption pitch* can be generated by periodically turning off or modulating the amplitude of noise (Burns, 1976; Miller & Taylor, 1948); *echo pitch* can be generated by mixing an ongoing noise with a single restatement of itself at a fixed time delay (Bilsen, 1970; Fourcin, 1965). These pitches based on noise have random variations in short-term spectra and waveforms, and generally sound quite noisy or hiss-like. But it should be possible to derive sounds with a completely tonal quality from noise.

By excising a section from noise and repeating this fixed waveform over and over without pause, a sound can be generated which is completely periodic and completely tonal. As we shall see, such recycled noise tones (RNTs) have characteristics reflecting their origins in noise.

The research was supported by NSF Grant BMS73-06787, NIH Grant HDO7855, and by the University of Wisconsin-Milwaukee Graduate School. We thank Mr. Joseph Nyland for his valuable technical help.

METHOD

All observations reported are based upon responses by five subjects with training in psychoacoustics (graduate students and staff). Listening was diotic through matched TDH-39 headphones while seated in an audiometric room. Steady stimuli were presented at 80 dB SPL, and stimuli with fluctuating levels were presented with peak levels at 80 dB. Special training was not necessary to perceive the effects to be described, and similar reports were made informally by untrained listeners hearing stimuli played over a loudspeaker in a classroom and in an auditorium.

A digital delay line (built by Physical Data, Inc. to our specifications) was used to repeat sections excerpted from noise. The auditory information was stepped in digital form by a stable clock through shift registers until the desired delay was reached. Upon switching to recycling mode, further input was rejected and the stored information was repeated indefinitely at a rate determined by the delay setting.

The spectra (shown in Figures 1-4) of some of the stimuli used in this study were derived from an X-Y plot produced by a Rockland Model 512S spectrum analyzer. These figures also show waveforms which are X-Y tracings produced by the spectrum analyzer operating in waveform-storage mode. All figures shown are based on signals which were low-pass filtered at 8,000 Hz to correspond to the frequency-response limits of the headphones.

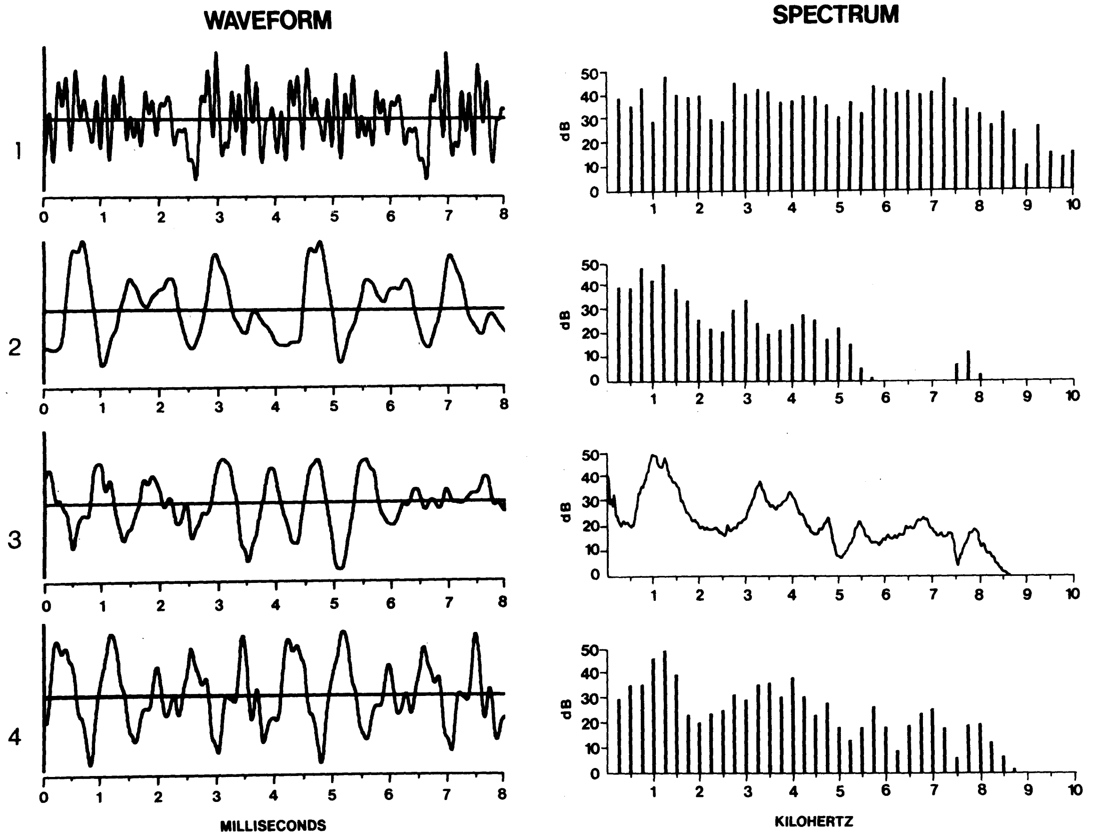
RESULTS AND DISCUSSION

For frequencies below 100 Hz (periods above 10 msec), a noisy or hiss-like character was heard by all listeners, but from about 100 Hz through 10,000 Hz (the upper response limit of our standard audiometric headphones), recycled noises seemed completely tonal. Pitches of these RNTs were equivalent to those of other types of periodic sounds of the same frequency. Thus, when a pulse train was adjusted to the same apparent pitch as a 200-Hz RNT by our five listeners, who each made five judgments, the mean pulse train frequency selected was 214.3 Hz, with a standard error of the mean of 10.6 Hz. (Separate noise excerpts were used to generate a different RNT for each of the 25 judgments.) All but 2 of the 25 adjustments were within 4 Hz of the RNT frequency. Both aberrant judgments were one octave too high (such octave errors are common in pitch-matching experiments).

The repeated waveform and corresponding line spectrum of a single 250-Hz RNT produced from white noise are shown in Figure 1. Each spectral component of this RNT has a randomly determined amplitude and phase, with no frequency-dependent change (up to the low-pass filter cutoff at 8 kHz), so that the RNT can be considered as a "white tone." White tones are exemplars of a general or nonspecific periodic sound, so they may be useful in psychoacoustic experiments when it is desirable to avoid special waveforms and restrictions in relative spectral amplitudes and phases.

Separate captures of 4-msec samples of white noise produced tones of the same pitch, each with its own rich and distinct timbre corresponding to the particular randomly determined harmonic structure. By changing the stepping rate of the delay line shift registers recycling one of these noise sections, it was possible to expand or contract the waveform and produce a family of RNT frequencies with different pitches but similar qualities. Use of frequencies corresponding to the musical scale generated tones sounding like notes produced by a novel mellifluous instrument. As frequencies of white tones approached the upper limit of pitch for orchestral instruments (about 4,000 Hz), timbre distinctions between different repeated noise sections gradually decreased; above 4,000 Hz, each white tone resembled a sinusoidal tone of the same frequency, no doubt as a consequence of the elimination of audible harmonics.

White tones are only one type of RNT. The number of possible types of RNTs is as large as the number of types of noises. Because of the importance of speech, we decided to turn next to the voice-related noises of whisper. Whisper is a noise with its spectral distribution shaped by the resonances of the vocal tract. It contains formants (peaks in spectral energy distribution) resembling those of voice. Recycling a section of the noise of a whispered vowel should produce a whisper tone with a line spectrum following the formant structure of the whisper and a pitch corresponding to the repetition rate.



Figures 1-4. Each figure corresponds to a different sound, with an 8-msec sample of its waveform appearing on the left and the spectrum showing the relative intensity of component frequencies on the right. Figure 1: Repeated 4-msec section of white noise. Figure 2: Voiced vowel /a/. Figure 3: Whispered vowel /a/. Figure 4: Repeated 4-msec section of whispered vowel /a/.

A portion of a sustained vowel /a/ produced by a female speaker at her normal fundamental frequency of 250 Hz is represented in Figure 2: The characteristic waveform of her glottal pulses is shown alongside the spectrum (based on a 500-msec voice sample) showing the harmonic components and their relative intensities. Figure 3 portrays a whispered production of /a/ by the same speaker: A segment of the continuously varying noisy waveform is presented, as well as the spectrum (based on a 500-msec whisper sample) showing lower formant frequencies which resemble those of the voiced vowel. Another section of this whispered /a/ equal to the duration of the speaker's glottal pulses during voicing (4 msec) was recycled to produce the whisper tone shown in Figure 4: The line spectra of the lower formants reponsible for vowel identification resemble those of the parent whisper in Figure 3 and the naturally voiced /a/ in Figure 2. Despite the gross waveform differences evident in Figures 2 and 4, the whisper tone was identified by all as a woman's voiced /a/.

The pitch of the artificially voiced whisper could be manipulated in two ways: (1) By using recycled whispered excerpts of different durations, the fundamental frequency was changed while the formant frequencies remained fixed at the values determined by the nonrecycled whisper; (2) by varying the delay line clock rate after the start of recycling, the frequencies of the fundamental as well as the formants were changed.

We reasoned that by using the synthetic voicing of whisper tones, we might be able to convert whispered discourse into a comprehensible voiced monotone through iteration of regularly spaced samples. By repeating a segment a few times, a brief whisper tone would be created with a pitch determined by the repetition frequency and a quality determined by the formant structure common to the whispered and voiced version of the speech sound. Sampling the successive iterated sections of whisper at sufficiently short intervals might preserve the phonemic information necessary for intelligibility.

We did find it possible to convert a whisper into intelligible voiced speech. A digital delay line was used together with programming equipment: A 4-msec segment from whispered discourse was repeated an additional three times to produce four statements of identical waveforms lasting a total of 16 msec; then a second 4-msec segment in the original whispered discourse was recycled to produce another set of four statements, etc.[1] The resulting stimulus was quite understandable, although it should be noted that there was a nonselective voicing of all phonemes, including those unvoiced in normal speech.[2] We were able also to introduce changes in pitch to avoid a monotone: Intonation could be introduced by varying the duration of the successive iterated captures.[3]

This study has been limited to the perception of repeated random waveforms in the tonal range above 100 Hz. However, there is evidence that recycled noises form a perceptual continuum, permitting periodicity detection to a lower atonal limit of .5 Hz. While there are characteristic qualities for particular ranges of repetition frequencies, some common rules seem to apply throughout this continuum of detectable periodicities (Warren & Bashford, 1977).

REFERENCES

Bilsen, F. A. Repetition pitch: Its implication for hearing theory and room acoustics. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: A. W. Sijthoff, 1970.

Burns, E. M., & Viemeister, N. F. Nonspectral pitch. *Journal of the Acoustical Society of America*, 1976, **60**, 863-869.

Fourcin, A. J. The pitch of noise with periodic spectral peaks. *Rapports 5e Congrès International d'Acoustique*, 1965, **Ia**, B52.

Miller, G. A., & Taylor, W. G. The perception of repeated bursts of noise. *Journal of the Acoustical Society of America*, 1948, **20**, 171-182.

Rainbolt, H. R., & Schubert, E. D. Use of noise bands to establish noise pitch. *Journal of the Acoustical Society of America*, 1968, **43**, 316-323.

Small, A. M., & Daniloff, R. G. Pitch of noise bands. *Journal of the Acoustical Society of America*, 1967, **41**, 506-512.

Warren, R. M., & Bashford, J. A. Infrapitch periodicity: Perception of missing fundamental and component periodicities for mixed iterated random waveforms below 20 Hz. *Journal of the Acoustical Society of America*, 1977, **61**, S51. (Abstract)

NOTES

1. Increasing repetition above three added little to the voiced quality. As for fewer repetitions, one restatement imparted only a slight pitch (resembling closely the classical echo pitch produced by mixing an ongoing sound with a delayed version of itself), and two restatements still retained a slight whisper-like character.

2. Of course, the parent whispered passage was also intelligible in spite of the absence of all voicing.

3. We also converted speech whispered by a male to an intelligible voice with a fundamental pitch of 125 Hz by generating 8-msec segments, each repeated four times, with fresh 8-msec samples taken every 24 msec. Despite the discard of 24-msec chunks from the whisper (twice the discard duration for synthesizing the higher pitched female voice), there was no difficulty in understanding the man's speech.

(Received for publication January 28, 1978.)