A METHOD FOR THE COMPARATIVE EVALUATION OF VISUAL AND AUDITORY DISPLAYS

P. W. Davall and J. M. Gill*

Abstract: In reading and mobility aids for the blind it is desirable to design the display to optimize the man-machine interface. This paper describes a possible method for comparing various displays.

Five subjects used compensatory tracking with random input signals (five different cutoff frequencies) with five one-dimensional displays where the error is represented by: 1) Visual: deflection, 2) Auditory: mark-space ratio, 3) Auditory: matching frequency, 4) Auditory: beat frequency, 5) Auditory: amplitude matching.

The significance of the coherency between the input signal and the subjects' output is discussed. The measured closed-loop frequency response is compared to that obtained from a modified form of cross-over model. Finally, two performance parameters are proposed for assessing the various displays.

INTRODUCTION

Research on the design of auditory displays has usually been motivated by the desire to provide non-visual displays for those who cannot use vision or to supplement a visual display particularly in aeronautical applications.

Pollack and Ficks (1954) studied multi-dimensional auditory displays in which each variable had only two states. They found that, in general, multiple stimulus encoding is a satisfactory procedure for increasing the information transmission rate associated with such displays. Mudd (1965) studied the relative effectiveness of frequency, intensity, duration and interaural difference dimensions of a pure-tone binary display. Frequency proved to be the most effective dimension for the purposes of cueing; intensity was the least effective.

Roffler and Butler (1968a) found that listeners could locate auditory stimuli accurately in the vertical plane when the stimulus was complex and included frequencies above 7kHz. Roffler and Butler (1968b) then found that subjects tended to place the audio stimuli on a vertical scale in accordance with their respective pitch. Higher-pitched sounds were perceived as originating above lower-pitched sounds.

Vinje and Pitkin (1971) studied human operator dynamics for aural-compensatory tracking using pitch of the tone to represent the magnitude of the tracking error. Error polarity was indicated in the two-ear display by switching the tone between ears as a function of error sign. For the one-ear display, error polarity was indicated by using modulated and unmodulated tones. The describing function and remnant data indicated that humans can control as well with aural cues as with visual cues for input signals with cutoff frequencies of 0.27 to 0.56 Hz.

A variety of two-dimensional auditory displays have been built. For instance, Black (1968) developed

*Department of Engineering, University of Warwick, Coventry CV4 7AL, England.

a display where the horizontal coordinate was represented by time delay and amplitude of the signal and the vertical coordinate by frequency (100-400 Hz). Fish and Beschle (1973) also used frequency (200-7000 Hz) for representing the vertical position of the scan but interaural intensity differences (up to 40 dB) for the horizontal.

Phillips and Seligman (1974) developed a multi-dimensional auditory display in which frequency, amplitude and timbre are all utilized. Robinson (Gill, 1974) also developed a two-dimensional display, as a non-visual equivalent to the cathode-ray oscilloscope, where the frequency of the signal depends on the vertical coordinate and time delay for the horizontal. However, there has been little systematic comparison of the various types of auditory displays, even though Kramer (1962) mentioned the need for this over a decade ago.

This paper describes a method for comparing various displays and applies the technique to one visual and four auditory one-dimensional displays. The technique can be extended to multi-dimensional displays.

EXPERIMENTAL METHOD AND ANALYSIS

The human operator was presented with an error signal $e(t)$, which was the difference between the input, $x(t)$, and the output, $y(t)$, (Fig. 1). The output was measured from two strain gauges (making two arms of a bridge section) strapped across a stiff joystick. The choice of a force (as opposed to a position) joystick considerably reduced the response time delay introduced by the muscle motor action of the arm. The input, $x(t)$, was band-limited gaussian white noise. Five bandwidths of 0.3, 0.5, 0.7, 1.0, and 1.5 Hz were used for each form of error-signal presentation. The operators' task was to minimize the error.

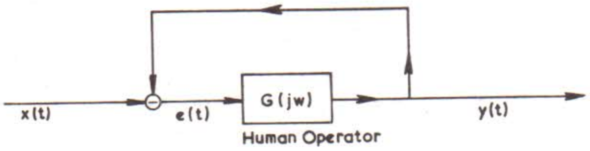

Figure 1. Closed-loop Configuration for Human Operator Tracking Tests.

Error Signal Presentation

Visual feedback. A horizontal line on a cathode-ray tube display moved vertically about a center zero, the deflection being a function of the error signal. By pushing on the joystick the line was moved up the screen, while pulling brought the line down. Full scale deflection was ±10 cm in response to an error signal of ±10 volts (0.5 kg-per-volt).

Mark-space audio feedback. The error signal was presented as short bursts of a single tone. Pulling the joystick reduced the duration of the pulses and increased the silent periods, while pushing reversed the effect. The fundamental clocking period for repeating the pulses was 0.15 seconds (Fig. 2). The zero error state was defined as unity mark-space ratio between noise and silence.

Matching-frequency feedback. A 1-kHz square was fed to the left ear as a reference tone while a variable-frequency square wave of equal amplitude was fed to the right ear. The frequency of the variable signal was controlled by the joystick, and could vary from zero to 3 kHz for full-scale error deflection of -10 volts to +10 volts. Zero error was recorded when the two inputs were of matching frequency.

Beat-frequency feedback. The signal inputs were the same as those in "Matching-frequency feedback" (above), except that both inputs were presented to both ears simultaneously. The effect was different, however, in that signals of *three* frequencies could be heard; a beat frequency was present as well as the two fundamental frequencies. The beat frequency decreases as the two fundamental frequencies come closer

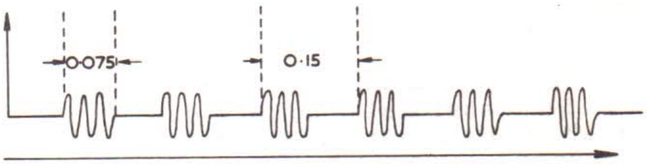

Figure 2. Zero Error State for Mark-Space Audio Feedback Error-Signal Presentation.

together. The advantage of this form of frequency matching is that it does not require the operator to be tone sensitive.

Amplitude-matching feedback. A 1-kHz square wave was presented to both ears. The amplitude presented to each ear as a function of the error signal is shown in Fig. 3. The loudness of the tone presented in each ear was approximately a linear function of the error (the ear being a logarithmic device to a first approximation). Zero error was recorded when the amplitudes were matched.

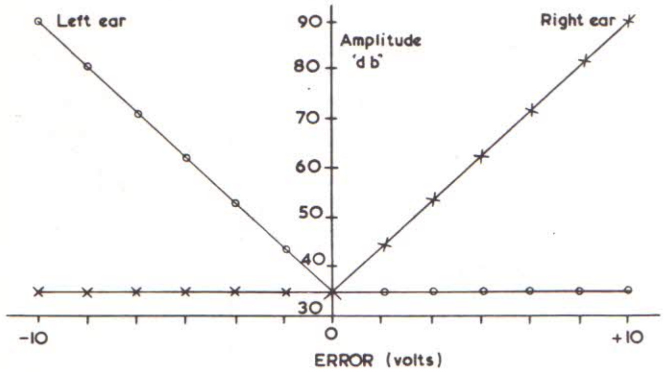

Figure 3. Amplitude V-Error Function for Left and Right Ears for the Amplitude-Matching Audio Test.

Analysis

$x(t)$: signal input

$y(t)$: system output

$e(t)$: error signal, $x(t)-y(t)$

$\hat{x}_i(j\omega)$: ith short-term Fourier estimator of $x(j\omega)$ at frequency ω.

$\bar{H}(j\omega)$: frequency-response estimate relating $x(j\omega)$ to $y(j\omega)$ at frequency ω.

$\bar{G}(j\omega)$: open-loop frequency-response estimate relating $e(j\omega)$ to $y(j\omega)$ at frequency ω.

$\bar{\phi}_{xx}(\omega)$: estimate of the auto-spectrum of $x(t)$ at frequency ω.

$\bar{\phi}_{xy}(j\omega)$: estimate of the cross-spectrum between x and y at frequency ω.

$\overline{C^2}_{xy}(\omega)$: squared-coherency estimate between x and y at frequency ω.

The signals $x(t)$, $y(t)$ and $e(t)$ were sampled five-times-per-second. Approximately ten-minute records were taken for each type of input signal (0.3, 0.5, 0.7, 1.0 and 1.5 Hz-cut-off frequencies.

The sample records were divided into blocks of N points (128 in this case) for Fourier analysis using a radix '2' fast Fourier algorithm (Blackman and Tukey, 1958; Wellstead, 1971) producing N/2 spectral coefficients over the frequency range 0 - 2.5 Hz. Approximately 30 short-term estimates of $e(j\omega)$, $x(j\omega)$ and $y(j\omega)$ were obtained for each input cutoff frequency.

$$\hat{x}_i(j\omega) = \sum_{k=0}^{N-1} x_t^i(k)\exp(-j\omega k) \quad (1)$$

where k denotes the kth time sample in the ith block and for $\omega = 0.039$. L $0<L<N/2$ where 0.039 is the frequency resolution in Hertz. Similar expressions are used to calculate $\hat{y}_i(j\omega)$ and $\hat{e}_i(j\omega)$.

The overall closed-loop frequency response as estimated between $x(t)$ and $y(t)$ was calculated as follows:

$$\bar{H}(j\omega) = \frac{\bar{\phi}_{xv}(j\omega}{\bar{\phi}_{xx}(j\omega)} \quad (2)$$

and the human-operator frequency response (open loop) relating $e(t)$ and $y(t)$ is given by:

$$\bar{G}(j\omega) = \frac{\bar{\phi}_{xy}(j\omega)}{\bar{\phi}_{xe}(j\omega)} \quad (3)$$

where

$$\bar{\phi}_{xx}(j\omega) = \sum_{i=1}^{30} \hat{x}_i(j\omega)*\hat{x}_i(j\omega) \quad (4)$$