Aural Phase Distortion Detection

Presented by Daisuke Koya
In Fulfillment of the Master's of Science Thesis Requirement

This work is dedicated to my mother, father, and dearest brother Eisuke, along with everyone who helped me in my pursuit of knowledge of audio.

Introduction

Previous research has proven that phase distortion in audio signals is audible. The real question then is not the existence, but the significance of the audibility. A psychoacoustic experiment is proposed to ascertain permissible levels of phase distortion in audio signals. This thesis research is not meant to be exhaustive by any means, but to ascertain some permissible levels based on careful experimental design and analysis. The *Kwalwasser-Dykema Music Tests* format will be used in the implementation of the thesis research with other considerations. These permissible levels may be beneficial in the design and application of audio equipment, especially in the area of transducer and loudspeaker system engineering.

Chapter 1 introduces the mathematical foundation of this thesis research. Phase distortion and group delay is defined in mathematical terms. Frequency- and time-domain characteristics of a minimum-phase second-order low-pass and first-order all-pass systems are investigated. Finally, an example of phase-intercept distortion is presented.

Chapter 2 introduces the audibility of phase distortion in audio signals. The physiology of the human ear is examined. The time-based processes of sound perception, such as timbral sensation in the human auditory process, are discussed. The phase-locking property of the auditory process is presented. Finally, previous research investigating the audibility of phase distortion is presented.

Chapter 3 will introduce the all-pass filter and audio test signals used in the listening test. The steps in formulating the all-pass filter are outlined. This is followed up by a formulation of a tunable second-order all-pass filter. The bilinear transform is

introduced to perform the analog to digital transformation for the all-pass filter. Finally, test signals used in the listening test are introduced and discussed.

Chapter 4 introduces the listening test formulation process. A study is shown which compares the validity of various psychoacoustic listening test implementations. Test equipment used in this thesis research is presented. Finally, the listening test implementation used in the thesis research is outlined.

Chapter 5 presents the results and discusses them. The statistical analysis method of equalizing type 1 and 2 error is presented and implemented. Average correct responses of all test subjects for both the headphone and loudspeaker listening tests are presented. Finally, the results are discussed and implications are brought forward.

Chapter 6 finalizes this thesis research by presentation of the conclusion.

Table of Contents

List of Figures

1. Mathematical definition of phase distortion [1]

The mathematical foundation of this thesis research will now be developed. The introduction of terms relevant to this research and mathematical definition of them will aid to better understand and identify the nature of phase distortion present in audio signals.

The impulse response $h(t)$ and complex frequency response $H(\omega)$, which characterize a causal, linear, and time-invariant system are interrelated by the Fourier transform pair

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt \tag{1.1}$$

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} d\omega \tag{1.2}$$

The complex frequency response in Eqs. (1.1) and (1.2) can be expressed in polar form as

$$H(\omega) = |H(\omega)| e^{j\phi(\omega)} \tag{1.3}$$

where $|H(\omega)|$ is the magnitude response (or gain for sinusoidal, steady-state operation) and $\phi(\omega)$ is the phase response (or steady-state phase shift of output relative to input). The requirement in the time domain for distortionless signal processing (preservation of waveform shape) is for $h(t)$ to have the form

$$h(t) = K\delta(t - T) \tag{1.4}$$

where δ is the unit impulse and constants $K > 0$ and $T \geq 0$. Eq. (1.4) and the convolution theorem together imply that a distortionless system scale the input signal by a constant K and delays the signal by T seconds. The output is therefore a delayed version of the input. Substituting Eq. (1.4) into Eq. (1.1) gives the corresponding restrictions on the frequency response

$$H(\omega) = Ke^{-j\omega T} \tag{1.5}$$

Comparing Eqs. (1.5) and (1.3) indicates that there is a twofold frequency-domain requirement for distortionless processing; constant magnitude response $|H(\omega)|=K$ and phase response proportional to frequency $\phi(\omega) = -\omega T$. Distortion of the waveform or *linear distortion* is caused by deviations of $|H(\omega)|$ from a constant value K. *Phase distortion* denotes departures of $\phi(\omega)$ from the linearly decreasing characteristic $-\omega T$.

Due to causality, there is a minimum amount of phase shift, or minimum phase $\phi_m(\omega)$ necessarily associated with a given magnitude response $|H(\omega)|$. This is described by the Hilbert transform

$$\phi_m(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\ln|H(\omega')|}{\omega' - \omega}\, d\omega' \tag{1.6}$$

If $\phi(\omega) = \phi_m(\omega)$, as given in Eq. (1.6), then $h(t)$ will be zero for negative values of time. Due to sources described next, additional, or *excess phase* can exist as well, therefore in general the total phase shift is

$$\phi(\omega) = \phi_m(\omega) + \phi_x(\omega) . \tag{1.7}$$

A practical definition for excess phase is

$$\phi_x(\omega) = \theta_a(\omega) - \omega T + \theta_0 \tag{1.8}$$

where θ_0 is a constant and $\theta_a(0) = 0$. In Eq. (1.8), $\theta_a(\omega)$ is the frequency-dependent phase shift of a (nontrivial) all-pass network, $-\omega T$ represents pure time delay as in Eq. (1.5), and θ_0 is a frequency-independent phase shift caused by polarity reversal between input and output of a Hilbert transformer, for example. Comparing the substitution of Eqs. (1.8) with (1.7) and Eq. (1.3) to Eq. (1.5), it can be seen that the terms $\phi_m(\omega)$, $\theta_a(\omega)$, and θ_0 contribute to the phase distortion of the system.

Two frequency-domain measures related to the phase response are phase delay $\tau_p(\omega)$ and group delay $\tau_g(\omega)$ defined as

$$\tau_p(\omega) = -\frac{\phi(\omega)}{\omega} \tag{1.9}$$

$$\tau_g(\omega) = -\frac{d\phi(\omega)}{d\omega} \tag{1.10}$$

For an amplitude-modulated carrier wave, the difference between phase delay and group delay is displayed in Fig. 1.1.

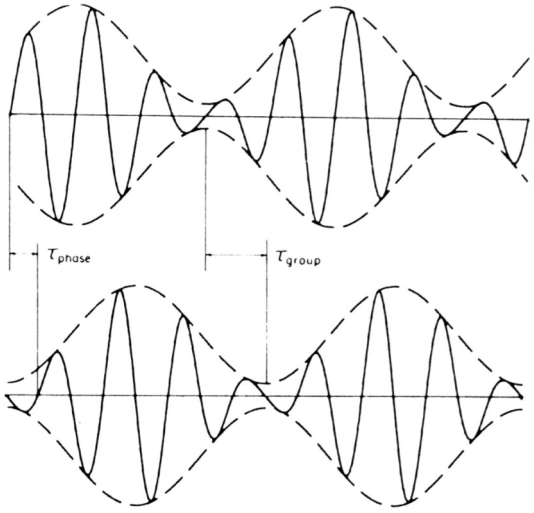


Fig. 1.1. Phase delay and group delay for amplitude-modulated wave. [D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review," *J. Audio Eng. Soc.,* vol. 30, pp. 774-794 (1982 Nov.), Fig. 1.]

The lower waveform here has a positive $\tau_p(\omega)$ and $\tau_g(\omega)$ with respect to the upper waveform at carrier frequency ω. From Eqs. (1.3) and (1.5), the absence of phase distortion requires that the phase and group delays in Eqs. (1.9) and (1.10) each equal the overall time delay $T \geq 0$, or

$$\tau_p(\omega) = \tau_g(\omega) = T \tag{1.11}$$

Therefore, deviations of either $\tau_p(\omega)$ or $\tau_g(\omega)$ from the constant value T indicate the presence of phase distortion. The substitution of Eqs. (1.7) and (1.8) into Eq. (1.10) shows the three terms that contribute to the total group delay

$$\tau_g(\omega) = T - \frac{d\phi_m(\omega)}{d\omega} - \frac{d\theta_a(\omega)}{d\omega} \tag{1.12a}$$

$$= T + \tau_{gm}(\omega) + \tau_{ga}(\omega) \tag{1.12b}$$

Comparing Eqs. (1.11) and (1.12a), group-delay *distortion* is defined as $\Delta\tau_g(\omega) = \tau_g(\omega) - T$ or by using Eq. (1.12b),

$$\Delta\tau_g(\omega) = \tau_{gm}(\omega) + \tau_{ga}(\omega) \tag{1.13}$$

Eq. (1.13) implies that $\Delta\tau_g(\omega) = 0$ is a necessary condition for no phase distortion and peak-to-peak excursions of $\Delta\tau_g(\omega)$ is a useful quantitative measure of phase distortion. Although the all-pass group delay $\tau_{ga}(\omega) \geq 0$, the minimum-phase group delay $\tau_{gm}(\omega)$ can be negative or positive. Therefore, $\Delta\tau_g(\omega)$ from Eq. (1.13) can be of negative or positive value. When $\tau_g(\omega)$ is calculated from $\phi(\omega)$ using Eq. (1.10), only the phase-slope information is preserved. The phase intercept information, $\phi(0) = \phi_m(0) + \theta_0$ is lost through differentiation. This result implies that when $\Delta\tau_g(\omega) = 0$ in Eq. (1.13), some phase distortion is possible, if for example, $\phi(\omega) = \phi_m(\omega) = -\pi/2$ [$H(\omega)$ is an ideal integrator] or $\phi(0) = \theta_0 = \pi/2$ [$H(\omega)$ contains a Hilbert transformer]. Thus $\Delta\tau_g(\omega) = 0$ and $\phi(0) \neq 0$ (or a multiple of π) implies no group-delay distortion but a different form of phase distortion known as *phase-intercept distortion*. In general, the total phase distortion produced by a linear system consists of both group-delay and phase-intercept distortion. As indicated in Eq. (1.13), the two sources of

group-delay distortion $\Delta\tau_g(\omega)$ are the minimum-phase response and the frequency-dependent all-pass portion of the excess phase response.

Fig. 1.2 shows the frequency and time-domain responses of a minimum-phase low-pass second-order resonant system whose transfer function is

$$T(s) = \frac{\beta}{(s+\alpha)^2 + \beta^2} \tag{1.14}$$

where s is the complex frequency variable ($s = j\omega$) and $\beta > \alpha > 0$.

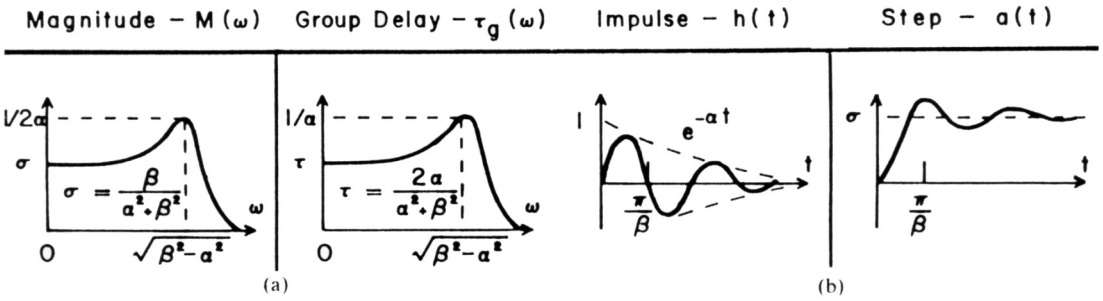


Fig. 1.2. Responses of minimum-phase low-pass resonant system. (a) Frequency domain. (b) Time domain. [D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review," *J. Audio Eng. Soc.,* vol. 30, pp. 774-794 (1982 Nov.), Fig. 2.]

As the peaking in the magnitude response increases by the reduction of the damping parameter value α, the group delay response peaks near resonance. In the time domain, the impulse response $h(t) = e(-\alpha t)\sin(\beta t)$ oscillates at frequency $f = \beta/2\pi$ Hz and is exponentially damped with time constant $1/\alpha$. The pronounced ringing is due to the frequencies in the vicinity of $\beta/2\pi$ Hz are strongly emphasized. The lack of symmetry in the impulse response is due to group-delay distortion. For this case, from Eq. (1.12b), $\tau_g(\omega) = \tau_{gm}(\omega)$.

Fig. 1.3 shows the frequency and time-domain responses of a first-order all-pass network transfer function that has the form

$$A(s) = \frac{1}{\alpha}\frac{\alpha - s}{\alpha + s}. \tag{1.15}$$

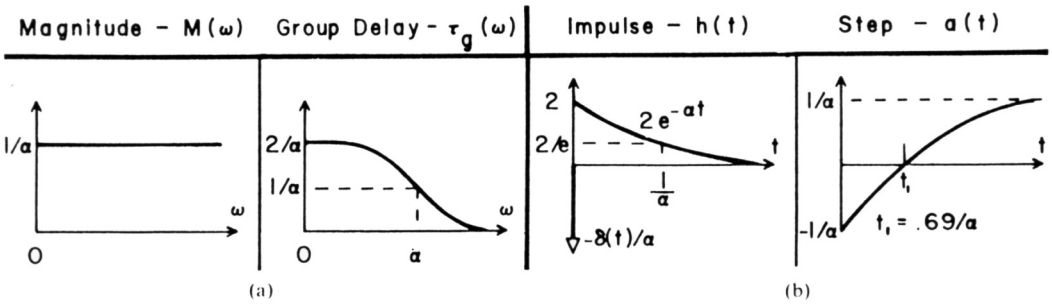


Fig. 1.3. Responses of first-order, all-pass system. (a) Frequency domain. (b) Time domain. [D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review," *J. Audio Eng. Soc.,* vol. 30, pp. 774-794 (1982 Nov.), Fig. 3.]

The magnitude $M(\omega)$ of the complex frequency response is constant. The group delay $\tau_g(\omega)$, which is always positive in value, shows a decrease as frequencies increase. For this case, from Eq. (1.12b), $\tau_g(\omega) = \tau_{ga}(\omega)$.

Fig. 1.4 illustrates phase distortion caused by a frequency-independent phase shift or phase-intercept distortion.

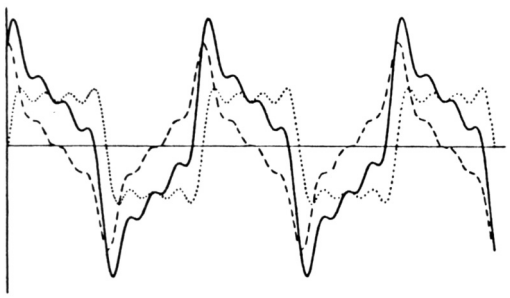


Fig. 1.4. Band-limited square wave (dotted), its Hilbert transform (dashed), and sum of dotted and dashed curves (solid). [D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review," *J. Audio Eng. Soc.,* vol. 30, pp. 774-794 (1982 Nov.), Fig. 4.]

The dotted curve represents a band-limited square wave (sum of the first four nonzero harmonics). The dashed curve is the Hilbert transform of the square wave obtained by the shifting the phase of each harmonic by π/2 radians, or 90°. This constant phase shift of each of the harmonics yields a linearly distorted waveform having greatly increased peak factors. The solid curve is the sum of the square wave and its Hilbert transform. Since the corresponding harmonics in this sum are of equal amplitude and in-phase quadrature, the solid curve could have been obtained by scaling the magnitude of the amplitude spectrum of the original square wave by $\sqrt{2}$ and rotating its phase spectrum by 45°. For this case, from Eq. (1.8), $\phi_x(\omega) = \theta_0 = \pi/4$ rad. Mathematically, the Hilbert transform of the time function $f(t)$ is itself a function of time which is defined by the convolution of $-1/\pi t$ with $f(t)$:

$$H\{f(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(x)}{x-t}\,dx\,. \tag{1.16}$$

The integral in Eq. (1.16) is understood in the sense of principal value. Assuming the Fourier transform of $f(t)$ to be $F(\omega)$ and noting that the Fourier transform of $-1/\pi t$ is $e(j\pi/2) = j = \sqrt{-1}$ for positive frequencies $\omega > 0$, spectrally, the Hilbert transform of $f(t)$ corresponds to a perfect π/2 rad (or 90°) shift in the positive-frequency phase spectrum of $F(\omega)$. Two successive Hilbert transformations of $f(t)$ yield $-f(t)$, which is the simple polarity reversal in the time domain or 180° phase shift of $F(\omega)$ in the frequency domain. The impulse response $-1/\pi t$ of a Hilbert transformer is not causal, meaning it is non-zero for negative values of time. The response is also of infinite duration in time, which implies that, in practice, an approximation is only possible with a Hilbert transformer.

This chapter introduced the mathematical foundation of the thesis research. Phase distortion and group delay were defined in mathematical terms. Frequency and time-domain characteristics of a minimum-phase second-order low-pass and first-order all-pass systems were investigated. Finally, an example of phase-intercept distortion was presented. These results will aid in the understanding and identification of phase distortion of audio signals in mathematical terms. The next chapter will investigate phase distortion from a human perception standpoint.

2. Audibility of Phase Distortion in Audio Signals

The physiological basis of how phase changes in audio signals may be perceived is now presented. Physiological foundation and previous research knowledge of the audibility of phase distortion may provide further insight in the human temporal auditory process. Furthermore, examination of previous research results provides assistance in the relevant design of further research.

2.1 Physiological Foundations

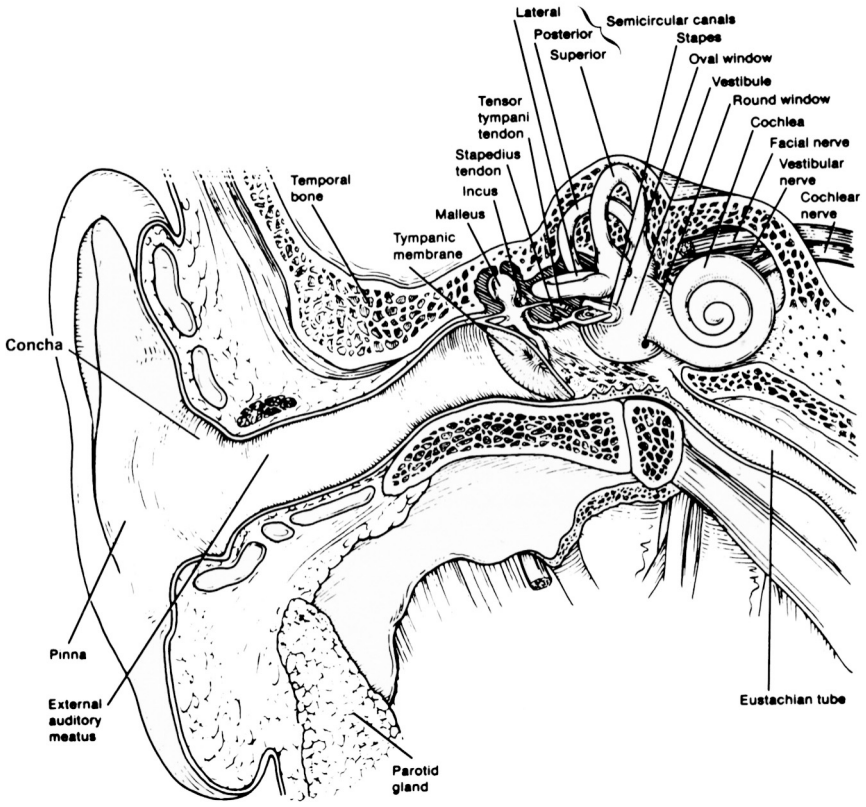


Fig. 2.1. The external, middle, and inner ears in man. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 11, Fig. 2.1]

Fig. 2.1 shows the outer, middle, and inner ear in humans. The external portion of the ear is the pinna. The “hole in the head” is the external auditory meatus. The tympanic membrane then separates the external and middle ear. For humans, in the range of frequencies of 2 ~ 7kHz, resonances of the external ear increases the sound pressure at the tympanic membrane, as shown in Fig. 2.2.

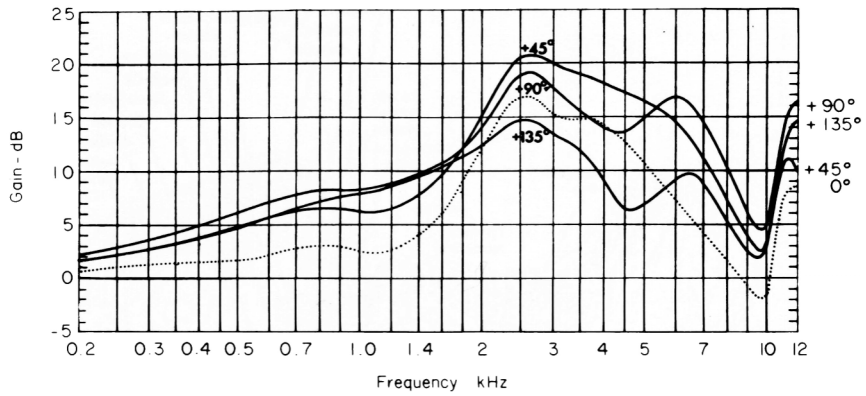


Fig. 2.2. The average pressure gain of the external ear. The gain in pressure at the eardrum over that in the free fired is plotted as a function of frequency, for different orientations of the source in the horizontal plane ipsilateral to the ear. Zero degrees is straight ahead. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 12, Fig. 2.2]

The Eustachian tube assists in the equalization of eardrum movement by maintaining equal air pressure on both sides of the tympanic membrane. Three small bones called malleus, incus and stapedius tendon connect the tympanic membrane to the oval window, which separates the middle and inner ear.

The oval window’s motions create wavelike movements in the fluid called the perilymph contained in the cochlea. The basilar membrane vibrates depending on the vibrations in the fluid. Where, when, and to what degree the basilar membrane is excited provides the foundation for aural sensation.

2.2 Temporal Processes

Timbre [2] is a multidimensionally perceived tonal attribute that differentiates tones of identical pitch, loudness, and duration. It is influenced by steady state waveforms, transient characteristics (the onset especially), and slower spectral changes over a series of tones. For example, a piano and a trumpet can play the note A440 of identical frequency, sound pressure, and duration but have clearly audible differences. Although it was once believed that the human ear is "phase deaf," in accordance to Ohm's acoustical law [2], more recent research has shown that relative phase has subtle effects on timbre, in particular when changing phase relationships occur within a continuously sounding tone.

For timbral sensation, the *onset* portion and other transient characteristics within the dynamic waveform are especially important. The onset is the opening portion of a tone, where the energy supplied exceeds the energy expended. Tones produced by continuous excitation of the vibrating source, such as a blown reed or mouthpiece, or a bowed string, have an onset that is followed by a steady state section, where the energy supplied and expended are in balance for the most part. Tones produced by impulsive excitation of the vibrating source, such as plucked string and piano tones, do not possess a steady state. The *offset* or decay, where the energy expended exceeds any supplied, concludes a tone.

A steady-state vibration pattern from a vibrating system cannot be attained instantaneously. Onset times of various instruments vary. The trumpet has an onset time of about 20 msec while the flute requires 200 to 300 msec. Even this short (in absolute terms) onset time of 300 msec is significant in perception of timbre. This

significance was demonstrated in an experiment [4] where the initial portion of a tone was removed. It was revealed that even experienced musicians had considerable difficulty in discerning common orchestral instruments.

Most music consists of superpositions of complex tones. An upper-level auditory neural system performs the timbre discrimination of two (or more) simultaneously occurring complex tones. Fig. 2.3 shows the hypothetical superposition of complex tones, a monaural signal comprised of instrument 1 playing *exactly* the note A_4, and instrument 2 playing exactly A_5, one octave higher, at similar intensity levels. The length of the vertical bars represents the total intensity of each harmonic actually reaching the ear.

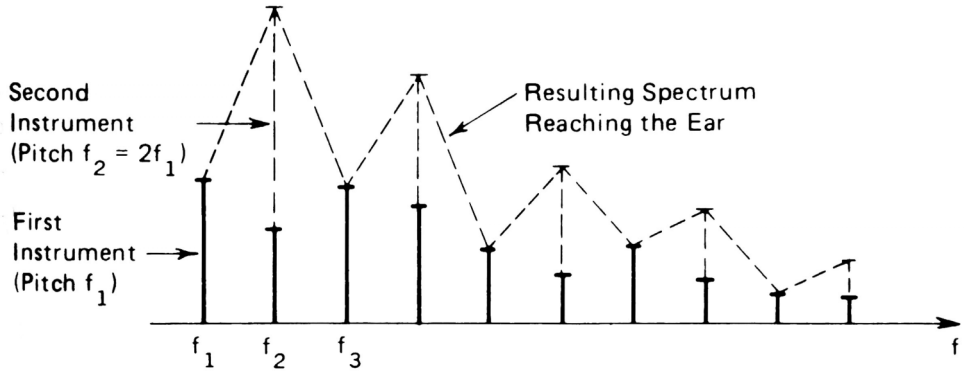


Fig. 2.3. Resulting spectrum of two complex tones of different timbre (spectrum) an octave apart. [J. G. Roederer, *The Physics and Psychophysics of Music : An Introduction* (Springer-Verlag, New York, 1995), pp. 163, Fig. 5.2]

Although the human timbre discrimination mechanism is not yet well established, a *time element* does seem to play a key role. Neither the attack nor the tone buildup of two supposedly simultaneous tones is ever exactly synchronized, especially if both tones have disparate sources (stereo effect). During the transient period of an onset of

a tone, the processing mechanism in the human brain seems to be able to lock on particular characteristic features of each instrument's vibration pattern and keep track of them, despite being blurred by the other instrument. Periodic vibrations in pitch (vibrato) may also provide important cues for discriminating tone quality.

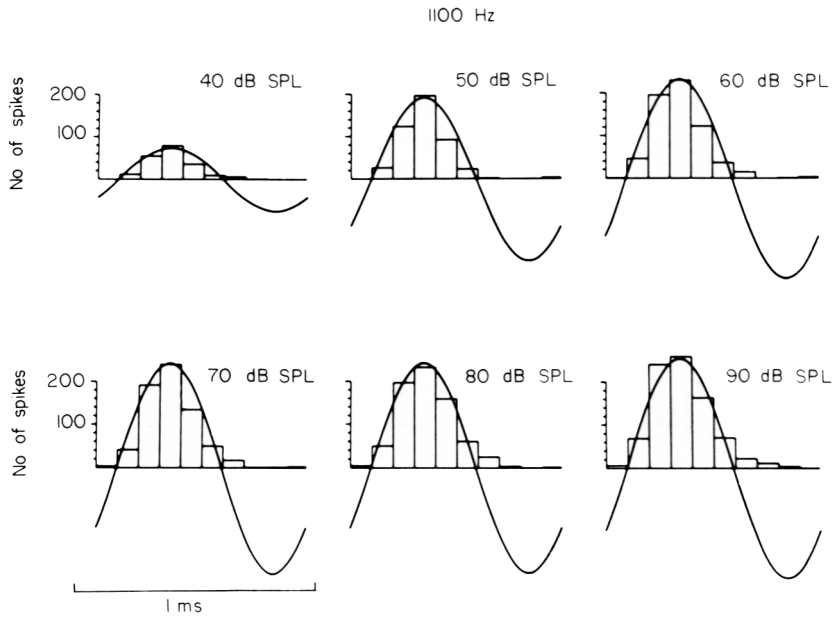


Fig. 2.4. Period histograms of a fiber activated by a low frequency tone indicated that spikes are evoked only one half of the cycle. The histograms have been fitted with a sinusoid of the best fitting sinusoid of the best fitting amplitude but fixed phase. Note that although the number of spikes increases little above 70 dB SPL, meaning that the firing is saturated, the histogram still follows the sinusoid without any tendency to square. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 83, Fig. 4.8]

Auditory nerve fibers [3] provide a direct synaptic connection between the hair cells of the cochlea and the cochlear nucleus. In humans, there are 30,000 of these fibers. Above 5 kHz, these nerve fibers fire with equal probability in every part of the cycle. However, at lower frequencies, it is apparent that the spike discharges are locked to one phase of the stimulating waveform. This phase locking can be shown by the means of a period histogram, Fig 2.4. In the construction of the period histogram,

the occurrence of each spike discharge is plotted in time. However, the time axis appears to reset in every cycle at a constant point on the stimulus waveform, most likely at the positive zero crossings, as shown in Fig. 2.4. It appears that the period histogram follows a half-wave rectified version of the stimulus waveform. If hair cell activation is linked directly to the mechanical events, it is reasonable to assume that this corresponds to deflection of the cochlear partition in the effective direction. Deflection in the opposite direction reduces the spontaneous activity of the nerve fiber.

Phase-locking is a sensitive indicator of the activation of a nerve fiber by a low frequency tone. At low stimulus intensities, even though the mean firing rate is not increased, a tone can produce significant phase-locking. Although for the above reason they could be more sensitive by about 20 dB, tuning curves that are based on a criterion of phase-locking are similar to those based on an increasing firing rate. As can be seen in Fig. 2.4, phase-locking is preserved as the intensity is raised. Although above 70 dB SPL the total number of spikes evoked does not increase, meaning that the firing rate is saturated, the period histogram still follows the waveform of the stimulus, and does not display any sign of squaring. This may be due to that fact that the hair cell's AC response is still sinusoidal. It is also possible that there exists a feedback mechanism that maintains the mean firing rate constant in saturation.

A click, which lasts a short time, but spreads spectral energy over a wide frequency range, may be thought of as the spectral complement of a tone, which has a long time duration but also has narrow frequency spread. Fig. 2.5 shows the poststimulus-time histograms of the auditory nerve fibers to click stimuli.

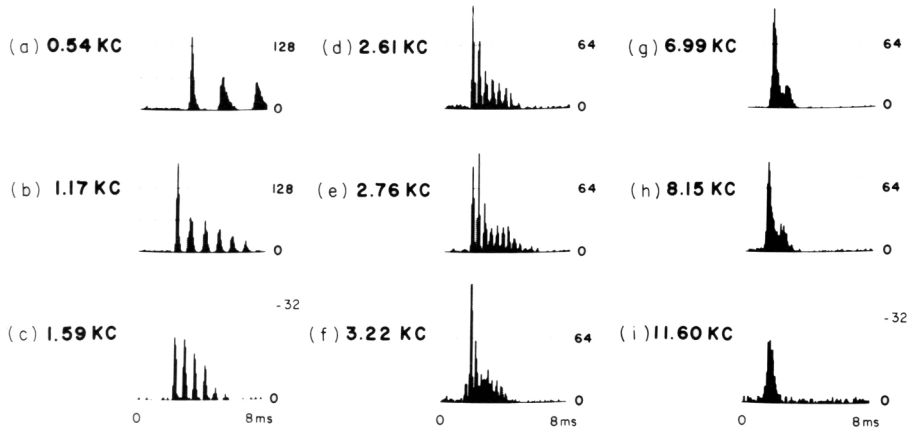


Fig. 2.5.The form of the poststimulus-time histograms to clicks depends on the characteristic frequency of the fiber. Low frequency fibers show ringing (a ~ f), high frequency fibers do not (g ~ i). High frequency fibers also show a later phase of activations (f ~ h). [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 84, Fig. 4.9]

Each auditory fiber possesses a 'characteristic' frequency that appears to behave as a band-pass filter. Histograms of low-frequency fibers display several decaying peaks that appear to be produced by a decaying oscillation (as if the cochlear transducer rings in response to a stimulus). The frequency of this ringing is equivalent to the characteristic frequency of the cell. This ringing at the characteristic frequency is exactly what is expected if the tuning of the auditory nerve fibers were produced by an approximately linear filter. It is also expected that the rate of decay of the ringing to be inversely proportional to the bandwidth of the tuning curve, so that a sharply tuned fiber would ring for a long time. However, this is true only to a certain extent. There exist some practical difficulties in making an exact comparison, because the number of spikes in the early peaks tends to limit, similar to the response to tone saturation at high intensities.

In accordance with the response to tone stimuli, it also appears that only one phase of the basilar membrane movement is effective. The histogram corresponds to the

upper half cycle of the decaying oscillation produced in the transducer. Since at the highest intensities, a rarefaction click (Fig. 2.6 (a)) produces the earliest response, it appears as though an upward motion of the basilar membrane is responsible for excitation. A rarefaction click will move the oval window outwards, which in turn moves the basilar membrane upwards, shown in Fig. 2.7 A(i). A condensation (Fig. 2.6 (b)), as opposed to a rarefaction click, reverses the positions of the peaks and troughs of the histogram, as though the basilar membrane were being driven in the opposite direction, shown in Fig. 2.6 A (ii). An approximate picture of the excitatory oscillation can be produced by inversion of the histogram for a condensation click below that for a rarefaction click, to produce what is called a compound histogram, as shown in Fig. 2.7 B. This resultant pattern can be compared with the basilar membrane impulse responses of Fig. 2.8. Histograms to clicks can also reveal that the suppression of activity during the less effective half cycle of the stimulating waveform is not due to refractoriness from previous activity, since the first sign of influence on a fiber can sometimes be a suppression of spontaneous activity produced by the less effective half cycle. It is not known at the moment if the nature of the decaying oscillation is only mechanical on the basilar membrane, mechanoelectrical, or purely electrical, as though the fiber were being driven by the decaying oscillation of an electrical filter.

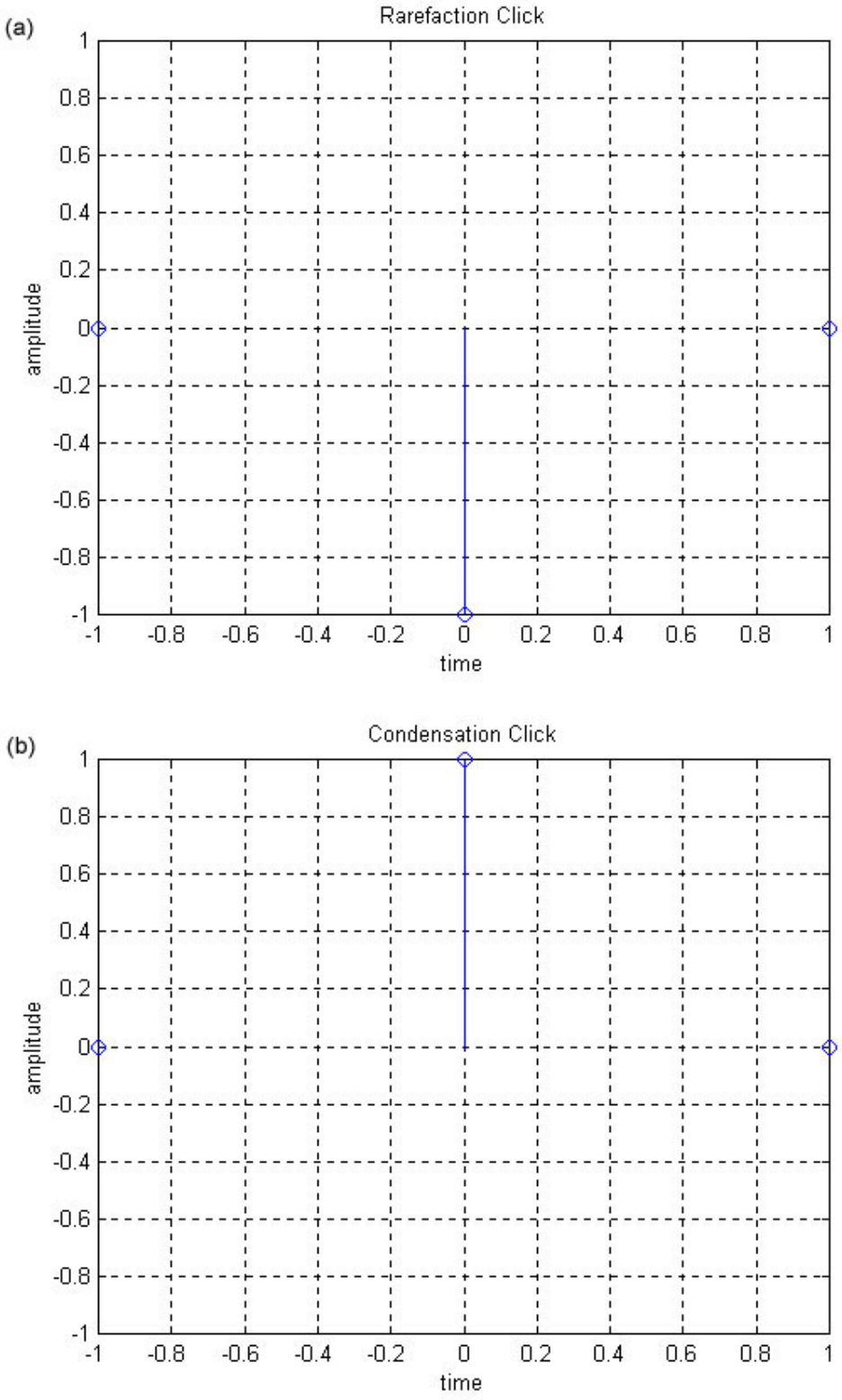


Fig. 2.6 (a). Rarefaction click (or impulse) which possesses negative amplitude.
(b). Condensation click (or impulse) which possesses positive amplitude.

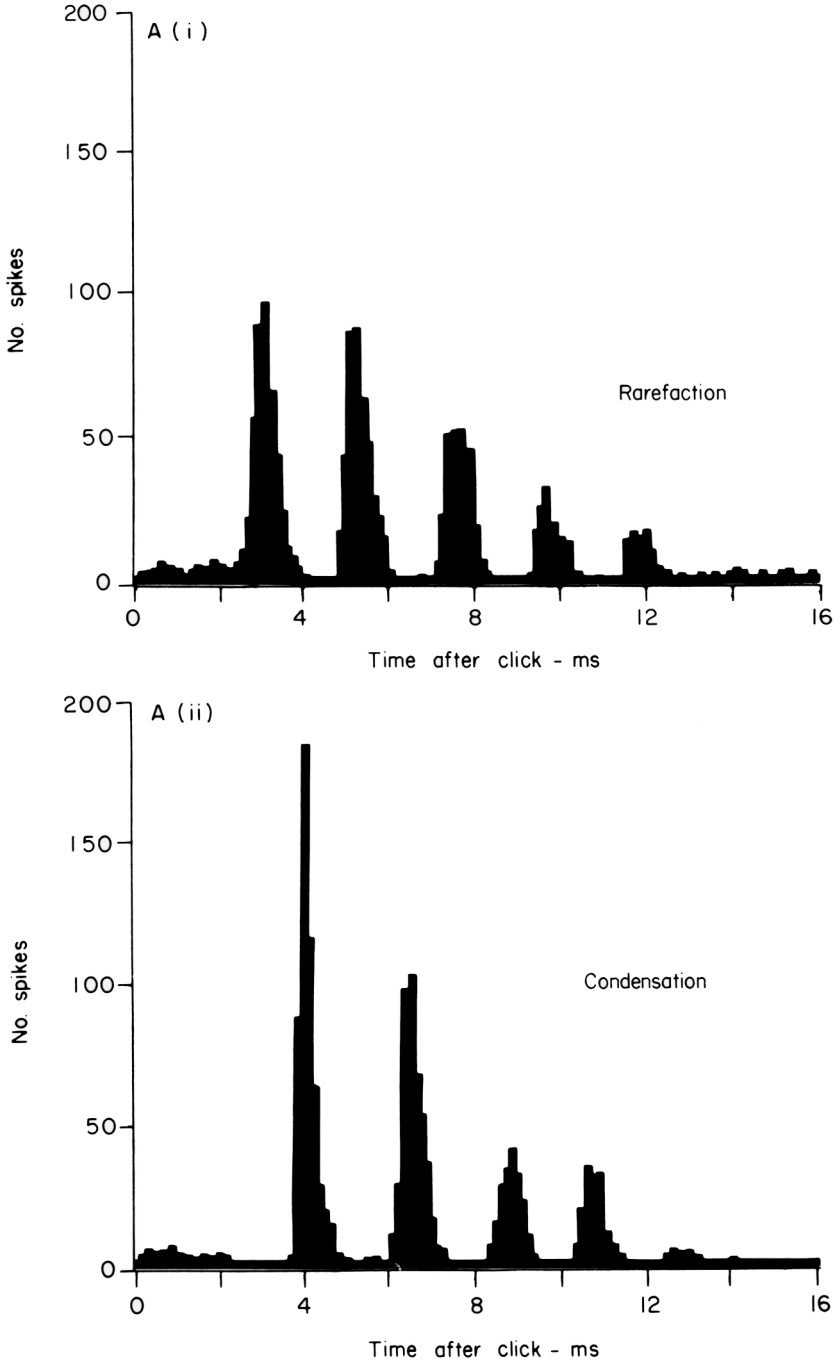

A (i)
200
150
100
50
0
No. spikes
Rarefaction
0
4
8
12
16
Time after click - ms
A (ii)
200
150
100
50
0
No. spikes
Condensation
0
4
8
12
16
Time after click - ms

Fig. 2.7 A. Poststimulus-time histograms to (i) rarefaction and (ii) condensation clicks show that the peaks and troughs occur in complementary places for the two stimuli. Fiber characteristic frequency: 450 Hz.
B. A compound histogram is formed by inverting the histogram to condensation clicks under that to rarefaction clicks. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 86 - 87, Fig. 4.10 A, B.]

Kim *et al.* [6] demonstrated the propagation of a combination tone by recording the responses of auditory nerve fibers to a two-tone stimulus. The responses of a large number of fibers were sampled, producing a 'neurogram,' or a display of activity in the whole nerve fiber array. Phase-locking to the two fixed primaries and to the combination tone were measured. Fig. 2.9 shows neurograms for the primaries and for the $2f_1 - f_2$ combination tone. For each of a large number of nerve fibers, they calculated the number of spikes phase-locked to the stimulus tone or combination tone of interest, divided by the number of spontaneous spikes. The results were then sideways averaged over fibers of a small range of characteristic frequencies, to

produce the running averages illustrated. Note that while the two primaries produced separate peaks of activity at their characteristic place in the cochlea, the combination tone produced a peak at *its* characteristic place in the cochlea as well.

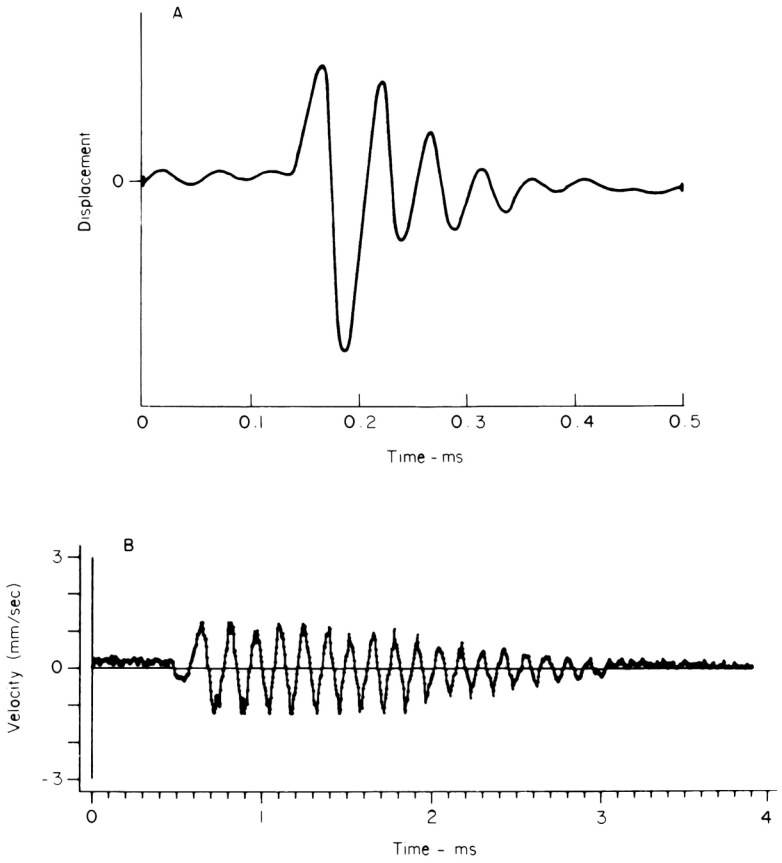


Fig. 2.8. Impulse responses of the basilar membrane show ringing.
A. Guinea-pig, measured by capacitive probe, at the 23 kHz point on the membrane.
B. Squirrel monkey, *velocity* of the impulse response, measured by the Mössbauer technique.
[J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 46, Fig. 3.12 A, B.]

The phase data in the neurogram produced very strong proof that the distortion component was transmitted along the basilar membrane by a travelling wave, just as though it were a tone introduced externally. For externally introduced tones, the phase of the activation in auditory nerve fibers increases steadily along the cochlea, due to

the time taken by the travelling wave (line f_s in Fig. 2.10). Highest to the site of generation, the phase of the response to the combination tone varied in the same exact manner, after an arbitrary phase shift which was necessary (since there is no reference phase - line $2f_1 - f_2$ in Fig. 2.10). We conclude that the distortion tone is propagated along the cochlea just as though it was an externally introduced tone, producing a resonance in the cochlea at the characteristic frequency point.

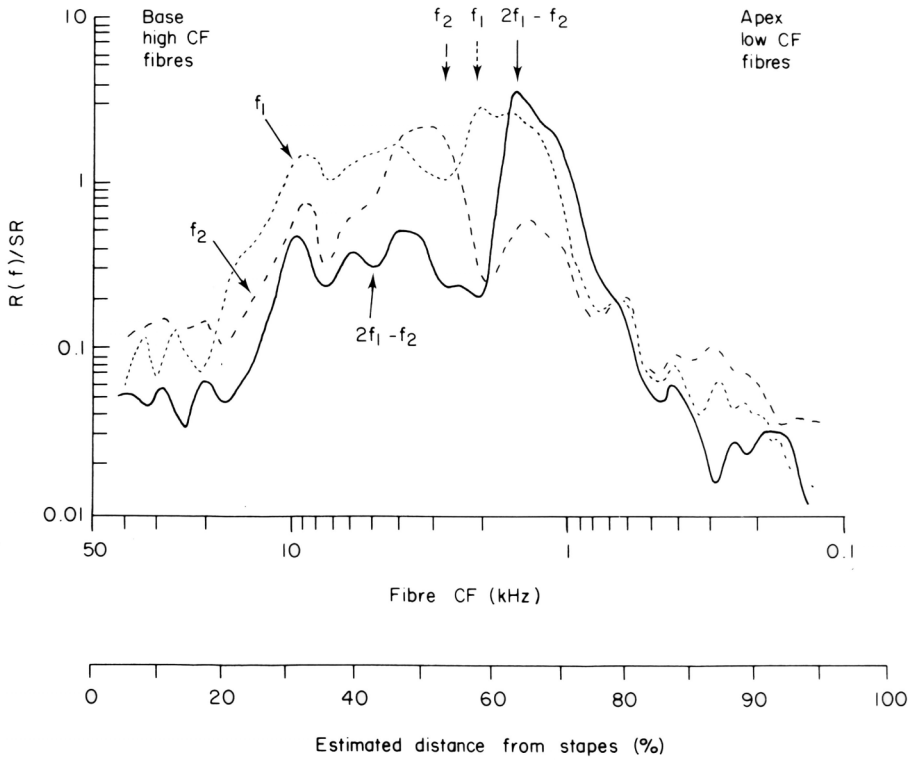


Fig. 2.9. Neurograms for the primaries and for the $2f_1 - f_2$ combination tone. The activity phase-locked to the primaries (f_1 and f_2) was most prominent in fibers of those characteristic frequencies (arrows). Activity phase-locked to $2f_1 - f_2$ was most prominent in fibers tuned to $2f_1 - f_2$. (There was also phase-locking to $f_1 - f_2$ which was deleted for the purposes of the illustration.) The frequencies of f_1, f_2, and $2f_1 - f_2$ are indicated by the arrows. Note that fibers of high characteristic frequency are plotted to the left of the figure, so that points on the left refer to the base of the cochlea. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 141, Fig. 5.14.]

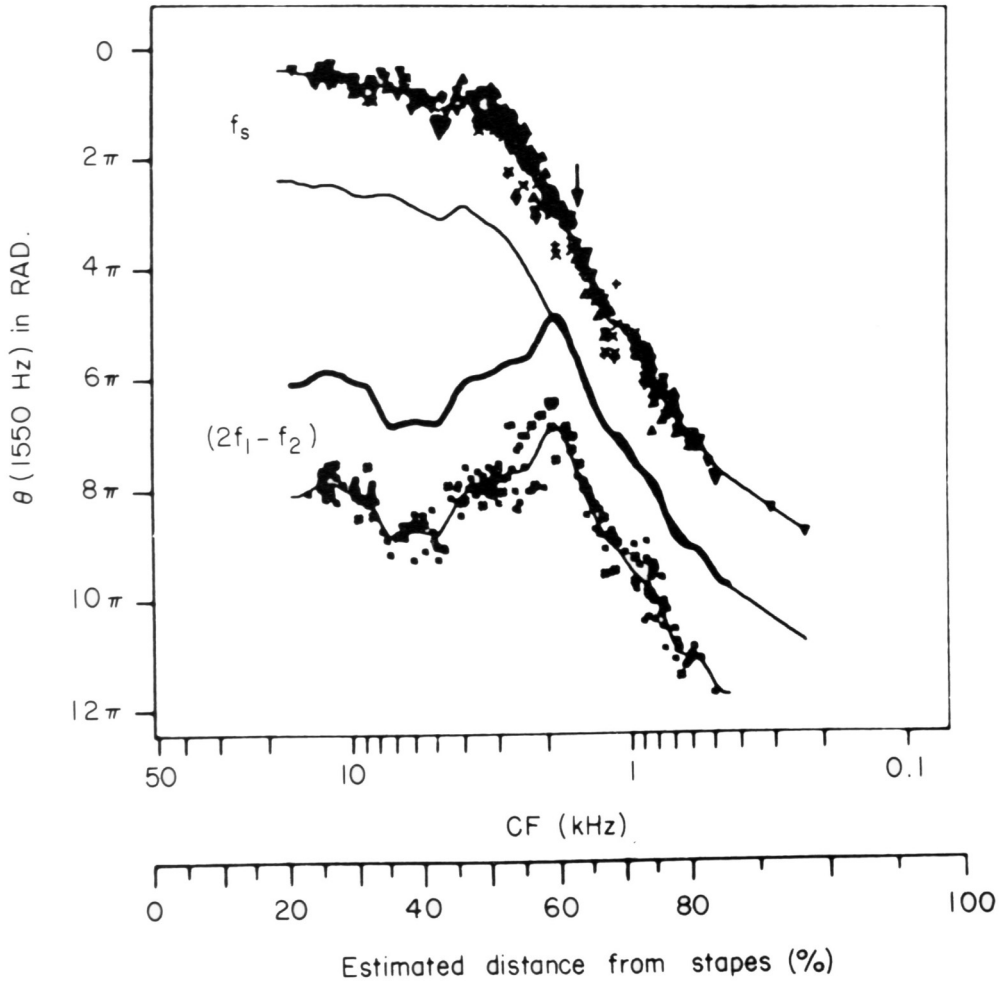


Fig. 2.10. The phase of the neurogram of the $2f_1 - f_2$ combination tone increases with distance apically to the site of generation, in just the same way as does the phase of an introduced tone (f_s). The points for the individual neurons are shown as well as the means; in the center the mean curves have been shifted so as to coincide at the frequency indicated by the arrow. 'CF' stands for 'characteristic frequency.' The stapes is one of three small bones of the ossicles which is connected to the oval window. [J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982), pp. 142, Fig. 5.15.]

Since the amplitude of the combination tone depends highly on the frequency separation of the primaries (f_1 and f_2), the generation of the combination tone seems to occur after one stage of frequency filtering. It can be supposed therefore that the overlap occurs only if both primaries get through the same first filter, possibly to be identified with the mechanical resonance of the cochlear partition. Mechanical energy

will have to be fed back to the basilar membrane in this instance. However, a reverse flow of energy from the transducer in the evoked cochlear mechanical response has also been shown. In this case, the energy was detected in the ear canal. Although the mechanism of the distortion tone re-emission is not certain, it would have to act very quickly since there is no indication that combination tones become weaker at high frequencies.

2.3 Previous Research

Although not in large numbers, previous research in investigation of the audibility of phase distortion has proven that it is an audible phenomenon. Lipshitz *et al*. [7] has shown that on suitably chosen signals, even small midrange phase distortion can be clearly audible. Mathes and Miller [8] and Craig and Jeffress [9] showed that a simple two-component tone, consisting of a fundamental and second harmonic, changed in timbre as the phase of the second harmonic was varied relative to the fundamental. The above experiment was replicated by Lipshitz *et al*., with summed 200 and 400 Hz frequencies, presented double blind via loudspeakers resulting in a 100% accuracy score. An experiment involving polarity inversion of both loudspeaker channels resulted in an audibility confidence rating in excess of 99% with the two-component tone, although the effect was very subtle on music and speech. Cabot *et al*. [10] tested the audibility of phase shifts in two component octave complexes with fundamental and third-harmonic signals via headphones. The experiment demonstrated that phase shifts of harmonic complexes were detectable.

Another very simple experiment conducted by Lipshitz *et al*. was to demonstrate that the inner ear responds asymmetrically. Reversing the polarity of only *one* channel of a pair of headphones markedly produces an audible and oppressive effect on both monaural and stereophonic material. This effect predominantly affects frequency components below 1 kHz. Because reversal of polarity does not introduce dispersive or time-delay effects into the signal, but merely reverses compressions into rarefactions and vice versa, these audible effects are due only to the constant 180° phase shift that polarity reversal brings about. Since interaural cross-correlations do not occur before the olivary complexes to which the acoustic nerve bundles connect, it must be concluded that what is changed is the acoustic nerve output from the cochlea due to polarity reversal. This change owes to two factors: cochlear response to the opposite polarity half of the waveform, and the waveform having a shifted time relationship relative to the signal heard by the other ear. This reaffirms the half-wave rectifying nature of the inner ear.

A frequent argument to justify why phase distortion is insignificant for material recorded and/or reproduced in a reverberant environment is that reflections cause gross, position sensitive phase distortion themselves. Although this is true, it is also true that the first-arrival *direct* sound is not subject to these distortions, and directional and other analyses are determined during the first few milliseconds after its arrival, before the pre-dominant reverberation's arrival. Lipshitz *et al*. do not believe that the reverberation effects render phase linearity irrelevant, and there exists confirmatory evidence [12].

Lipshitz *et al.*'s research involved analog implementations of first- and second-order unity-gain all-pass networks ranging in frequency from roughly 100 Hz to 3

kHz, with frequency switchable in $\sqrt{2}$ steps. The Q of the second-order networks was switchable in $\sqrt{2}$ steps from 1/2 to $2\sqrt{2}$. Transducers used in the experiment were electrostatic Stax headphones and Quad loudspeakers, for their notable phase linearity.

Test material used and notable results include:

- low-frequency square waves of around 150 Hz - Displayed pitch or timbre changes as the all-pass networks were introduced into their chain for *all* Q and f_0 settings, with both first- and second-order networks. The effect was most audible at high levels, although it was detected down to about 60 dB SPL. For the above test with Quad electrostatic loudspeakers, the effects were more audible for near-field auditioning.
- very low repetition rate square waves (2 - 5 Hz) - Demonstrated audible phase effects which sound like the ringing of the all-pass network at f_0. Low selected all-pass frequencies (113Hz - 529Hz) were the most audible, and detection became more difficult above 1kHz. The highest Q positions were the most audible.
- raised cosine modulated pulse - Although signal changes were audible on the second-order all-pass from 160 to 353 Hz at higher Q ($\geq$1), it was not very audible at low Q values or with first-order networks.
- assorted waveforms - The sawtooth waveform proved to be the most audible synthesized waveform. 440 Hz half-wave rectified sine wave with eight harmonics was audible on both first- and second-order networks from 113 to 1037 Hz.
- pre-recorded music:
  - male and female singing (with second-order all-pass with Q of 0.5 set to 160 and 240 Hz) - These were audible with a 95% confidence level.
  - variety of live sounds (recorded anechoically) - male voices, handclaps, metal rods struck against each other, and blocks of plastic and wood clicked together, etc. The unpitched signals had a higher degree of audibility. Male voice changes were not detectable.

The conclusion drawn by Lipshitz *et al*. was that midrange phase distortion can be heard not only on simple combinations of sinusoids, but also on many common acoustical signals. This audibility was far greater on headphones than on loudspeakers in a reverberant listening environment.

Hansen and Madsen [13], [14] have conducted several experiments that have been completed regarding the audibility of phase distortion. A displaced sine wave (non-

zero DC component) will have the time and frequency functions as shown in Fig. 2.11.

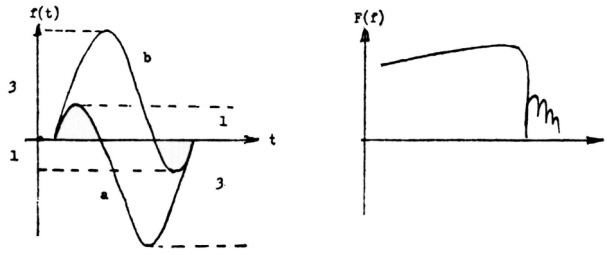


Fig. 2.11. Single-sine pulses with differing displacements (left) and their common spectral plot (right). [V. Hansen and E. R. Madsen, "On Aural Phase Detection," *J. Audio Eng. Soc.,* vol. 22, pp. 10-14 (1974 Jan./Feb.), pp. 12, Fig. 7.]

Frequency analysis demonstrated that there was no difference in the frequency spectrum. However, listening tests conducted with an electrostatic loudspeaker on both signals disclosed a clearly audible difference in timbre.

Hansen and Madsen's second experiment [14] used a very narrow spectrum with three bars, as shown in Fig. 2.12 for a listening test.

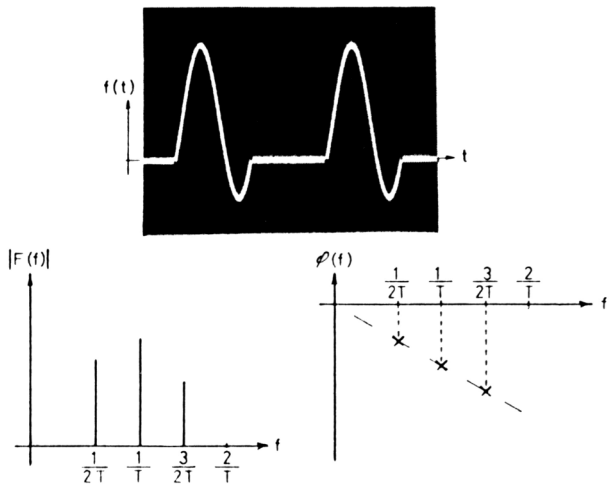


Fig. 2.12. Time function giving a three-bar spectrum. [V. Hansen and E. R. Madsen, "On Aural Phase Detection: Part II," *J. Audio Eng. Soc.*, vol. 22, pp. 783-788 (1974 Dec.), pp. 784, Fig. 3.]

Seven mutually overlapping ranges were selected, each of them containing three harmonics:

1. 50 – 150 Hz
2. 100 – 300 Hz
3. 200 – 600 Hz
4. 400 – 1200 Hz
5. 1 – 3 kHz
6. 2 – 6 kHz
7. 5 – 15 kHz

The quantity $h = A/B$, where the listeners had a chance of switching time functions as shown in Fig. 2.13, could take the following values:

$h_1 = 19/21$
$h_2 = 18/22$
$h_3 = 17/23$
$h_4 = 16/24$
$h_5 = 15/25$

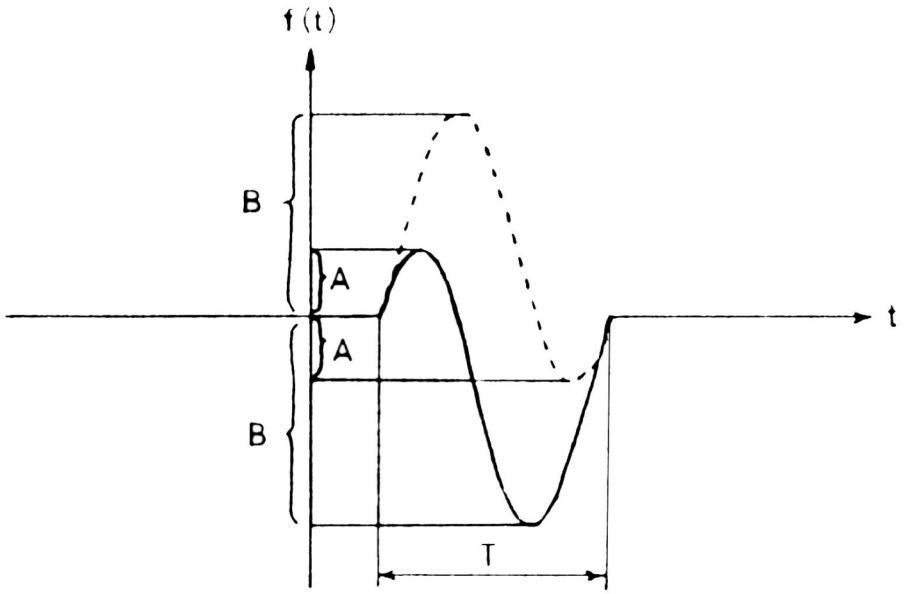


Fig. 2.13. Test signals. [V. Hansen and E. R. Madsen, "On Aural Phase Detection: Part II," *J. Audio Eng. Soc.,* vol. 22, pp. 783-788 (1974 Dec.), pp. 783, Fig. 1.]

The greater the difference between *A* and *B*, the greater was the phase difference, as shown in Fig. 2.14.

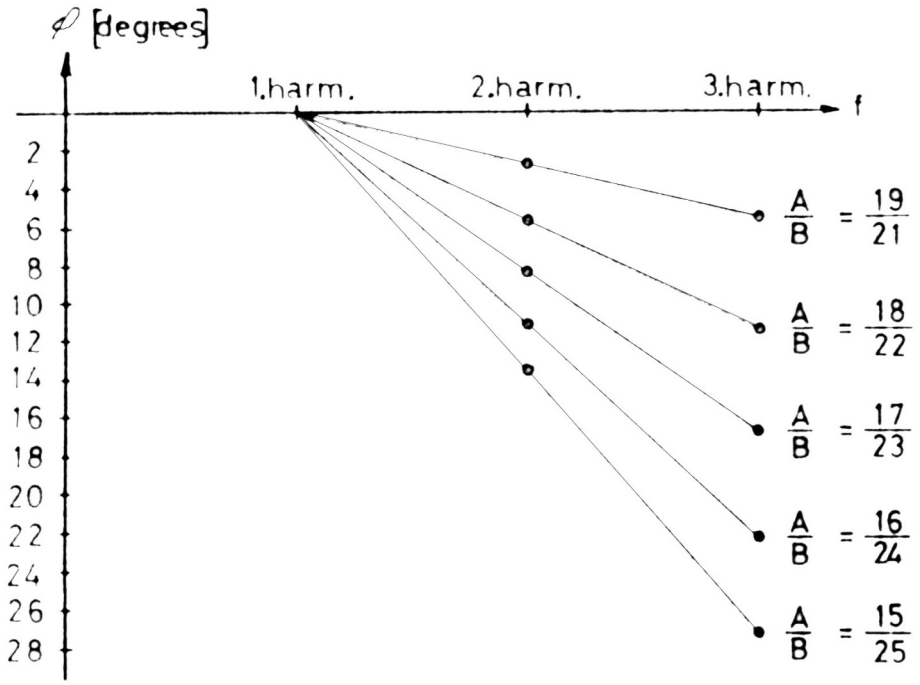


Fig. 2.14. Phase relationship between harmonics for different *h* values. [V. Hansen and E. R. Madsen, "On Aural Phase Detection: Part II," *J. Audio Eng. Soc.,* vol. 22, pp. 783-788 (1974 Dec.), pp. 784, Fig. 4.]

A test was conducted with a Quad electrostatic loudspeaker in a standard living room. Average results for all listeners and the resulting plots of permissible phase distortion levels and phase deviations are shown in Figs. 2.15 and 2.16. The five curves on Fig. 2.15 represent the various quantities of phase difference ratio $h = A/B$ used as the plot parameter. The five curves on Fig. 2.16 represent the relative minimum sound pressure levels for just noticeable detection of phase change between the signals. As an interesting side note, it was found that the tests revealed noticeably increased phase sensitivity with loudspeaker tests in reverberant environments as compared to headphone tests. This increased phase sensitivity may be due to reflections, or standing waves converted into amplitude shift present in the reverberant room.

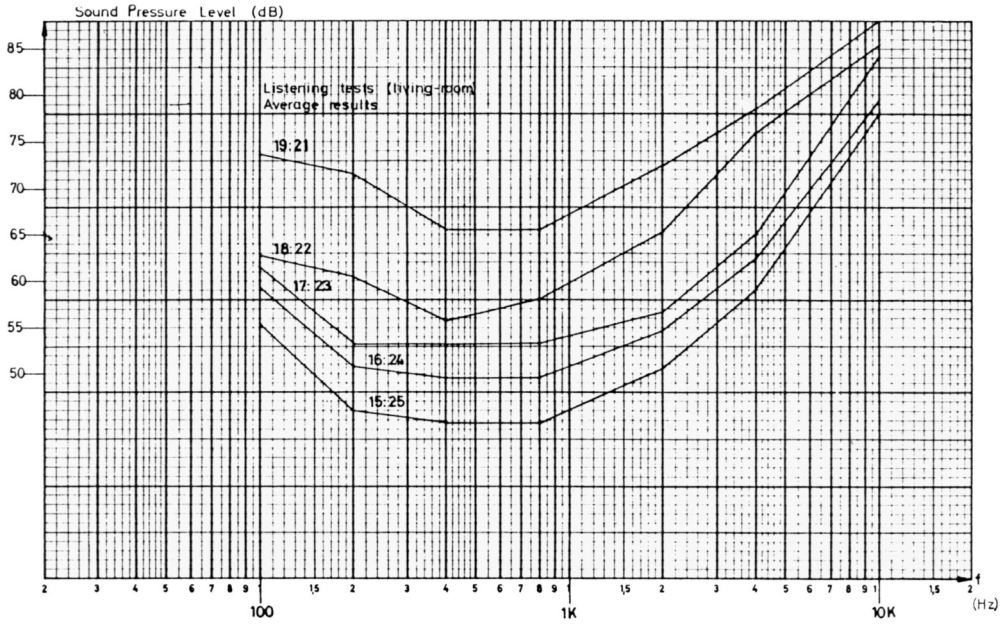


Fig. 2.15. Average results obtained from listening test in a reverberant environment with phase as a parameter. [V. Hansen and E. R. Madsen, "On Aural Phase Detection: Part II," *J. Audio Eng. Soc.,* vol. 22, pp. 783-788 (1974 Dec.), pp. 787, Fig. 7.]

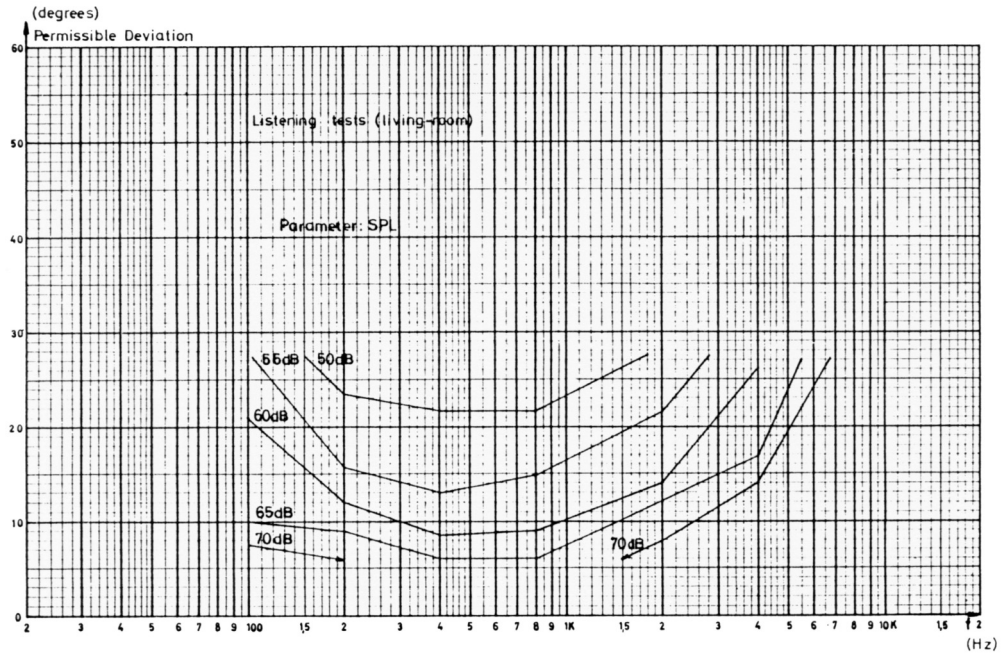


Fig. 2.16. Permissible phase deviation (living room) with sound pressure as a parameter. [V. Hansen and E. R. Madsen, "On Aural Phase Detection: Part II," *J. Audio Eng. Soc.,* vol. 22, pp. 783-788 (1974 Dec.), pp. 787, Fig. 8.]

Suzuki *et al.* [15] conducted a phase distortion perception experiment with transient signals of short duration as shown in Fig. 2.17. The time interval T_0 for each signal was chosen so that $T_0 = 2/f_0$, where f_0 is the 90° phase shift frequency of an analog phase-lag type all-pass filter defined by

$$f_0 = \frac{1}{2\pi R_0 C_0} \tag{2.1}$$

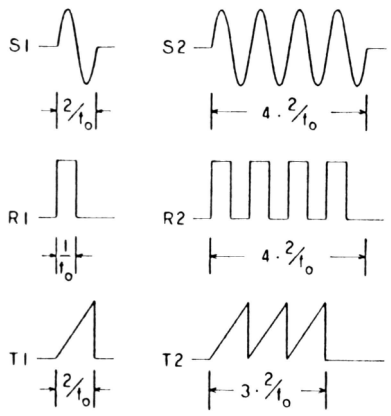


Fig. 2.17. Transient Signals used for hearing test. [H. Suzuki, S. Morita, and T. Shindo, "On the Perception of Phase Distortion," *J. Audio Eng. Soc.,* vol. 28, pp.570-574 (1980 Sep.), pp. 572, Fig. 4.]

These transient signals were then phase shifted by a single-pole phase-lag type all-pass filter. Transient signal intervals were changed according to the frequencies of 300 Hz and 1 kHz, the value of f_0 explained previously.

The procedure involving filtered (A) and unfiltered (B) signals were presented as A·A·A·A·A·B·B·B·B·B ·A·A·A·A·A ··· B·B·B·B·B, where '·' denotes a one-second interval. The correct answers percentages for each subject are shown in Fig. 2.18, where f_0 = 300Hz, and a loudspeaker in a listening room was used. Fig. 2.19 displays

the results of the listening test conducted in an anechoic chamber, where f_0 = 300Hz, and the signals S2 and R2 were used.

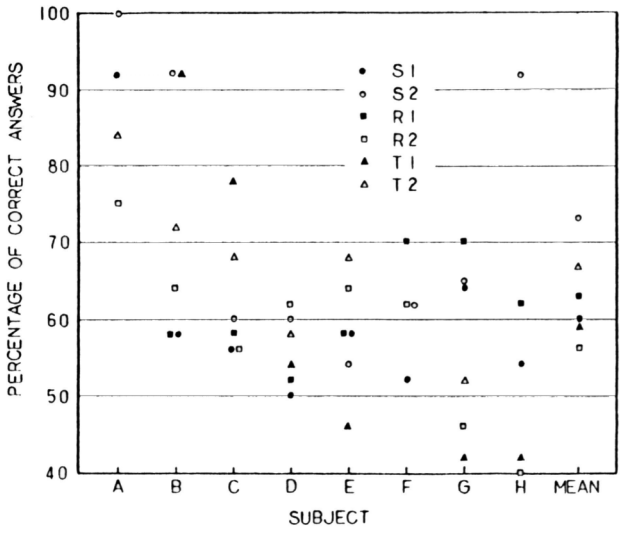


Fig. 2.18. Percentages of correct answers of loudspeaker listening for various signals in a listening room, where f_0 = 300Hz. [H. Suzuki, S. Morita, and T. Shindo, "On the Perception of Phase Distortion," *J. Audio Eng. Soc.,* vol. 28, pp.570-574 (1980 Sep.), pp. 572, Fig. 7.]

Certain people who participated in the test clearly heard the phase distortion present in the low frequencies of a single-pole all-pass filter when highly artificial signals were used. In this sense, for high-fidelity reproduction, phase distortion is not permissible. Another conclusion made by Suzuki *et al*. was that phase effects were highly individual and headphone listening showed much greater sensitivity than loudspeaker listening.

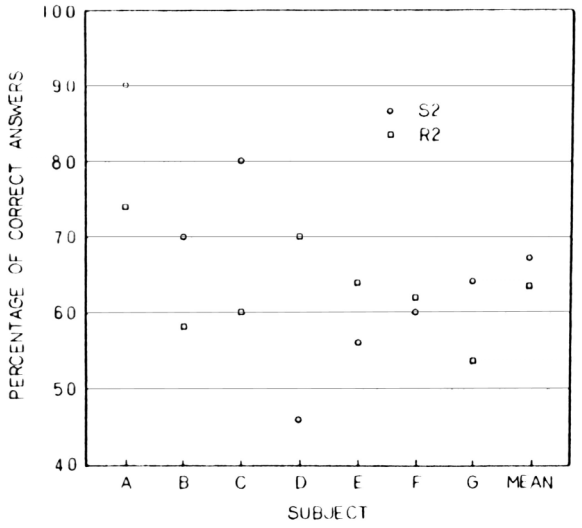


Fig. 2.19. Percentages of correct answers of loudspeaker listening for S2 and R2 in an anechoic chamber, where f_0 = 300Hz. [H. Suzuki, S. Morita, and T. Shindo, "On the Perception of Phase Distortion," *J. Audio Eng. Soc.,* vol. 28, pp. 570-574 (1980 Sep.), pp. 573, Fig. 8.]

Fincham [16] tested the effect of the reduction in group-delay distortion in the audio record/reproduction chain by means of a minimum phase-shift equalizer in carefully controlled conditions. These effects were clearly heard but quite subtle. In another test, a 8 cycles of a 40-Hz tone burst was used which was cascaded with an all-pass filter with significant group delay around 40-50Hz. Loudspeakers were used. Distinct audible differences in sound quality were observed by most of the lecture theater audience.

Preis *et al.* [17] conducted the audibility of phase distortion produced by minimum-phase 4-kHz and 15kHz anti-alias filters. In his experiment, group-delay distortion was doubled progressively until 67% mean correct discrimination was attained. Fig. 2.20 shows the mean correct discrimination between phase-distorted (minimum-phase) and undistorted (linear-phase) test signals for three low-pass systems (4-kHz elliptic and Butterworth, 15-kHz elliptic).

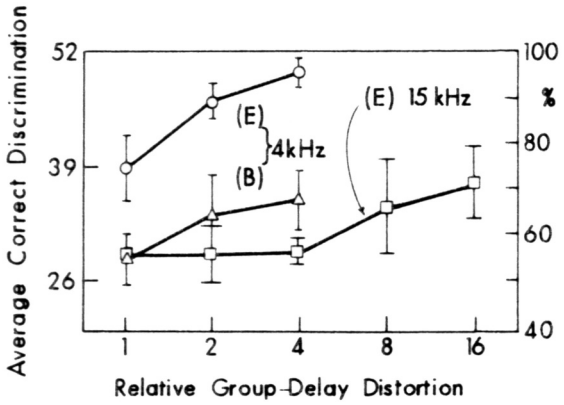


Fig. 2.20. Average correct discrimination between signals with no group-delay distortion and progressively doubled group-delay distortion. 52 presentations per subject of each of 11 test-signal pairs. 5 subjects; E - elliptic; B - Butterworth. [D. Preis and P. J. Bloom, "Perception of Phase Distortion in Anti-Alias Filters," *J. Audio Eng. Soc.*, vol. 32, pp. 842-848 (1984 Nov.), pp. 844, Fig. 2.]

It was concluded that for the impulsive test signals used and diotic (same signal in both ears) presentation via headphones, the ear is significantly more sensitive in the middle of the audio band (4 kHz) than at the upper edge of the band (15kHz) to group-delay distortion.

Some considerations of the experiment conducted by Preis *et al.* include the fact that no attempt was made to determine the detailed dependence of the perceptual threshold on frequency, the peak and width of the group delay characteristic, or signal intensity, or signal polarity. Secondly, other test signals, such as speech and music, were not used. Finally, other methods of irradiation, such as loudspeakers in non-reverberant or reverberant environments, were not tried.

Deer *et al.* [18] investigated and established perceptual thresholds for the audibility of phase distortion in all-pass filters with broadband impulsive test signals presented diotically over headphones. An analog second-order all-pass filter at 2kHz

(assumed to be most sensitive) that had adjustable peak value as well as bandwidth of group-delay distortion was used. Fig. 2.21 displays the mean correct discrimination between phase-distorted and undistorted (linear-phase) test signals for the second-order all-pass systems. Fig. 2.21 reveals that there exists a statistically significant effect perceived when peak group delay at 2 kHz is greater than 2 msec.

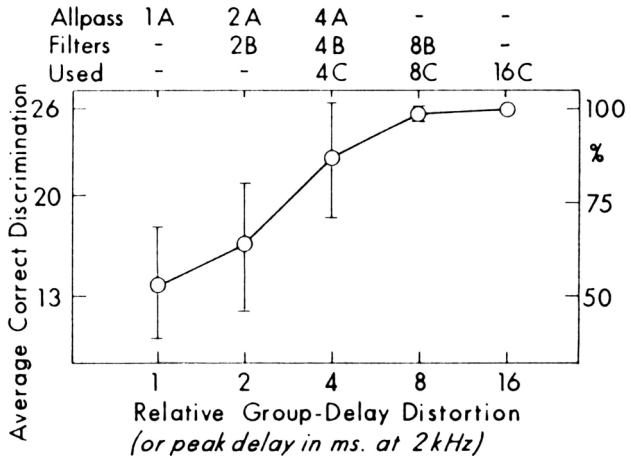


Fig. 2.21. Average correct discrimination between signals with no group-delay distortion and progressively doubled peak group-delay distortion. 26 presentations per subject, 6 subjects. [J. A. Deer, P. J. Bloom, and D. Preis, "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.*, vol. 33, pp. 782-786 (1985 Oct.), pp. 783, Fig. 1.]

Fig. 2.22 displays the result of the listening test of the all-pass filter having a fixed peak group delay at 4 msec and the delay bandwidth varied. The pair [4A, 4C], with Q difference of 0.71, was distinguished as possessing a statistically significant difference.

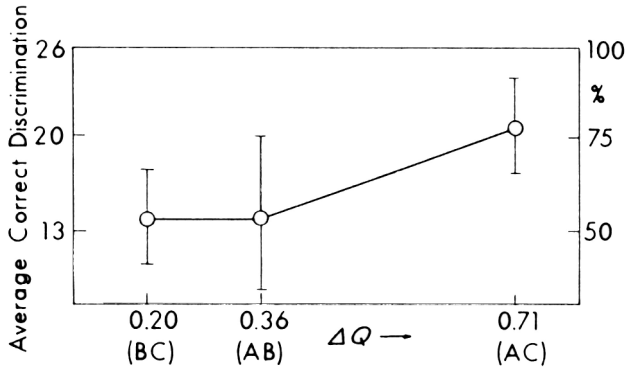


Fig. 2.22. Average correct discrimination between signals with identical peak group delays and differing relative bandwidth. 26 presentations per subject, 6 subjects. [J. A. Deer, P. J. Bloom, and D. Preis, "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.,* vol. 33, pp. 782-786 (1985 Oct.), pp. 785, Fig. 4.]

A new and important contribution made here was the interrelationship between group delay distortion (a frequency-domain measure) and the corresponding impulse response (time-domain measure) for each all-pass filter, particularly the envelope of the impulse response, and perceived phase distortion.

A consideration of this research was that for the all-pass filter, only 2 kHz for f_c was used, because this frequency was assumed to provide maximum sensitivity of the ear. Effects of increased or decreased signal of polarity were not determined. Finally, loudspeakers in non-reverberant or reverberant environments were not tested

Preis *et al.* [19] has also conducted Wigner distribution analysis on various filters with perceptible phase distortion. The responses of several anti-alias and all-pass filters were displayed jointly in time and frequency using the Wigner distribution. The Wigner distribution of a signal contains four useful properties:

1) Frequency response
2) Group delay
3) Instantaneous power
4) Instantaneous frequency.

These properties can each be estimated visually by taking a “slice” of the elevation contours of the Wigner distribution parallel to the horizontal time axis or the vertical frequency axis in the time-frequency plane. Fig. 2.23 and 2.24 display the interplay between the Wigner distribution and impulse response, frequency response, envelope power or energy-time curve, group delay, and instantaneous frequency for a 4 kHz low-pass filter and a 2 kHz all-pass filter with a peak group delay of 8 msec, respectively.

This chapter introduced the audibility of phase distortion in audio signals. The physiology of the human ear was examined. The time-based processes of sound perception, such as timbral sensation in the human auditory process, were discussed. The phase-locking property of the auditory process was presented. Finally, previous research investigating the audibility of phase distortion was presented. The physiological foundation and previous research of phase distortion detection may provide further insight in the human temporal auditory process and assist in the relevant design of further research. The results from this chapter will aid in partial formulation of the listening test used in the thesis research, as will be shown in the next chapter.

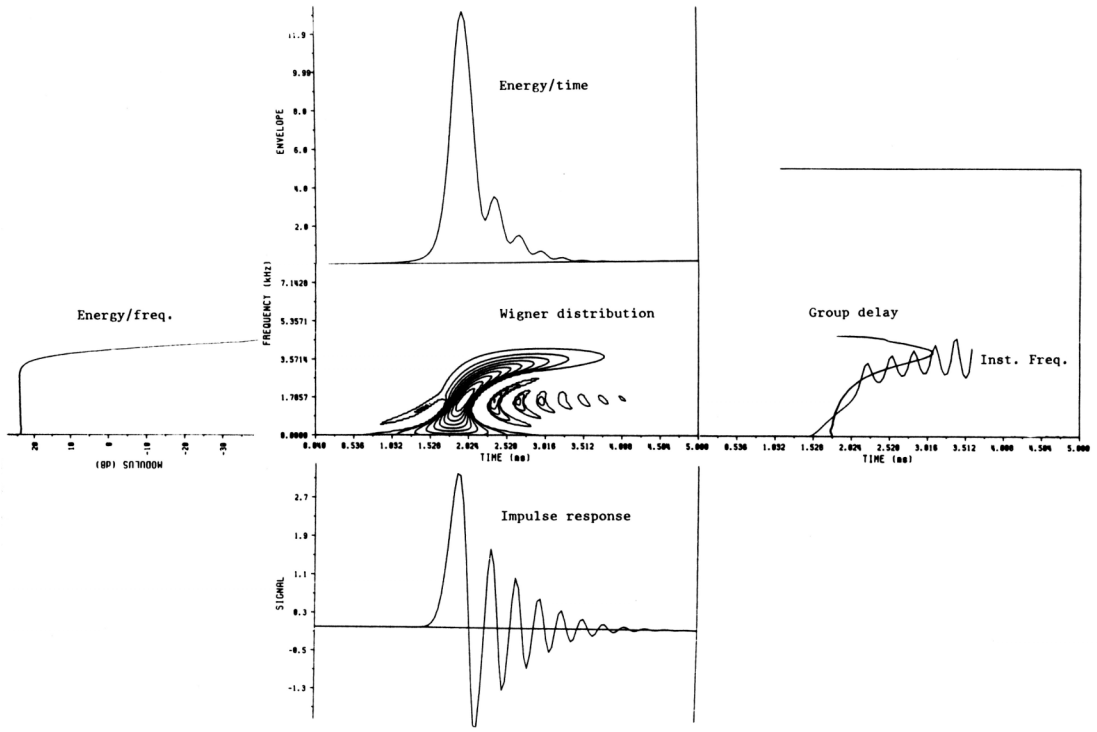


Fig. 2.23. Interplay between Wigner distribution and impulse response, frequency response, envelope power or energy-time curve, group delay, and instantaneous frequency for a 4 kHz anti-alias filter. [D. Preis, F. Hlawatsch, P. J. Bloom, and J. A. Deer, "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *J. Audio Eng. Soc.*, vol. 35, pp. 1004-1012 (1987 Dec.), pp. 1011, Fig. 9]

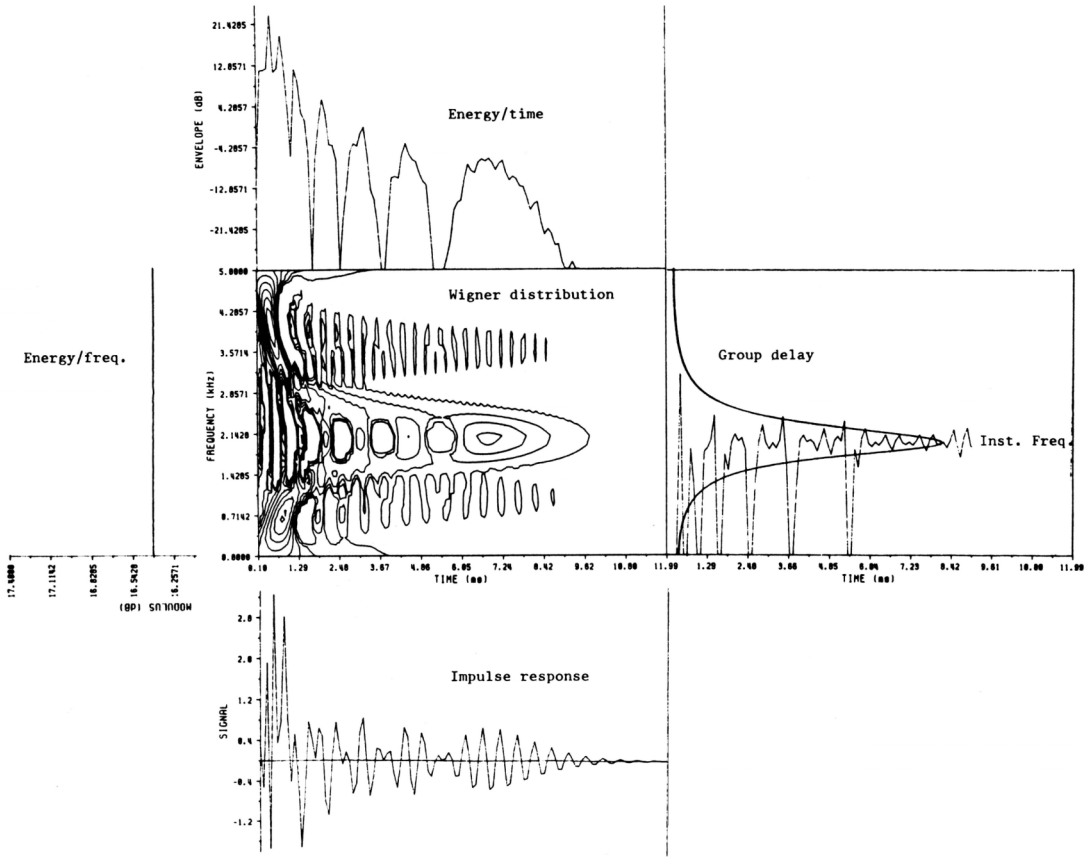


Fig. 2.24. Interplay between Wigner distribution and impulse response, frequency response, envelope power or energy-time curve, group delay, and instantaneous frequency for a 4 kHz anti-alias filter. [D. Preis, F. Hlawatsch, P. J. Bloom, and J. A. Deer, "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *J. Audio Eng. Soc.*, vol. 35, pp. 1004-1012 (1987 Dec.), pp. 1011, Fig. 10]

3. All-Pass Filter and Test Signals

The all-pass filter and audio test signals used in the listening test will now be presented. The theory in formulation of a tunable second-order all-pass filter facilitates implementation of the test signals. Test signal selection and implementation is justified by previous research results.

3.1 All-Pass Filter Formulation

All-pass filters [20] pass all frequencies and rejects none as shown in Fig. 3.1. Although its magnitude characteristic is flat over the frequencies zero to infinity, the phase characteristic is shifted in accordance with the all-pass filter's characteristic.

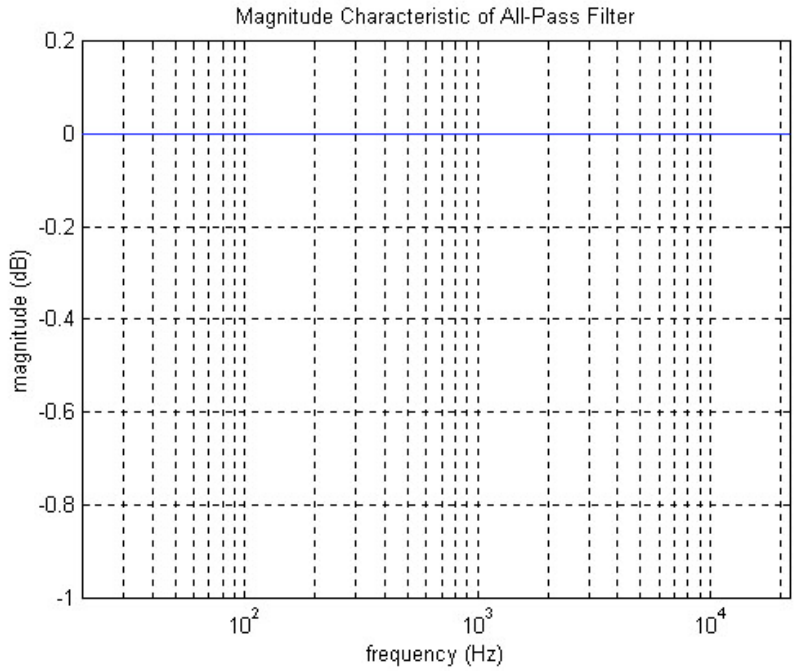


Fig. 3.1. Magnitude characteristic of an all-pass filter.

Suppose that the all-pass filter has gain function H(s) where

$$H(s) = k\frac{N(s)}{D(s)} \tag{3.1}$$

If the filter is to be of the all-pass type, then its ac steady-state magnitude characteristic $|H|$ must be constant over all frequencies and satisfy

$$|H(j\omega)|^2 = H(s)H(-s)\big|_{s=j\omega} = \left.\frac{N(s)N(-s)}{D(s)D(-s)}\right|_{s=j\omega} = k^2 \tag{3.2}$$

Then, the numerator and denominator of an all-pass filter are related as

$$N(s) = kD(-s) \tag{3.3}$$

such that the zeros are the images of the poles. This means, if $H(s)$ has a pole at $s = s_p$, then it *must* also have an image zero at $s = -s_p = s_p e^{j\pi}$. Therefore, image zeros lie diagonally opposite their respective poles in the s-plane.

The ac steady-state phase characteristic of the all-pass filter can be determined easily. From Eq. (3.1), its phase equals

$$\arg H(j\omega) = \arg N(j\omega) - \arg D(j\omega) \tag{3.4}$$

Since

$$\arg N(j\omega) = \arg D(-j\omega) = -\arg D(j\omega) \tag{3.5}$$

from Eq. (3.3),

$$\arg H(j\omega) = -2\arg D(j\omega) = 2\arg H_{LP}(j\omega) \tag{3.6}$$

such that the phase of the all-pass filter is double that owing to the poles acting alone (its associated low-pass filter having gain $H_{LP}(s) = k/D(s)$). The delay of the all-pass filter equals

$$\tau(j\omega) = -\frac{d\arg H(j\omega)}{d\omega} = 2\frac{d\arg D(j\omega)}{d\omega} = 2\tau_{LP}(j\omega) \tag{3.7}$$

which is double that of its associated low-pass filter.

Stable, causal systems cannot have RHP (right-hand plane) poles and have only simple $j\omega$-axis poles. Their zeros may lie anywhere in the s-plane and possess any multiplicity. Complex poles and zeros exist always in conjugate pairs. In general, stable, causal all-pass filters have the additional constraints that $N(s) = kD(-s)$ from Eq. (3.3). Since the zeros are the images of the poles, following facts can be concluded regarding stable, causal all-pass filters:

1. All zeros must lie in the right-half plane. Any LHP (left-hand plane) zero would have to have a RHP image pole which leads to instability.
2. Thus for stability, all poles must lie in the left-half plane and possess image zeros in the right-half plane.
3. Poles or zeros cannot lie on the $j\omega$-axis because the image zeros or poles would produce pole-zero cancellation.
4. All-pass filters are only capable of introducing positive delay (envelope retardation), but never negative delay (envelope advancement). The delay characteristic is always twice that of their associated low-pass filters having gain $H(s)=k/D(s)$.
5. Thus, the nature of the phase characteristic of all-pass filters must be monotonically non-increasing.

The first-order all-pass filter has gain

$$H(s) = k\frac{s-\omega_n}{s+\omega_n} = -k\frac{\omega_n - s}{\omega_n + s} \tag{3.8}$$

The delay characteristic of an all-pass filter is double that of a low-pass filter having gain $H(s) = 1/(s + \omega_n)$ so

$$\tau(j\omega) = 2\frac{d\left[\tan^{-1}(\omega/\omega_n)\right]}{d\omega} = \frac{2}{\omega_n}\frac{1}{1+(\omega/\omega_n)^2} \tag{3.9}$$

The pole-zero pattern for the gain and delay are shown in Fig. 3.2.

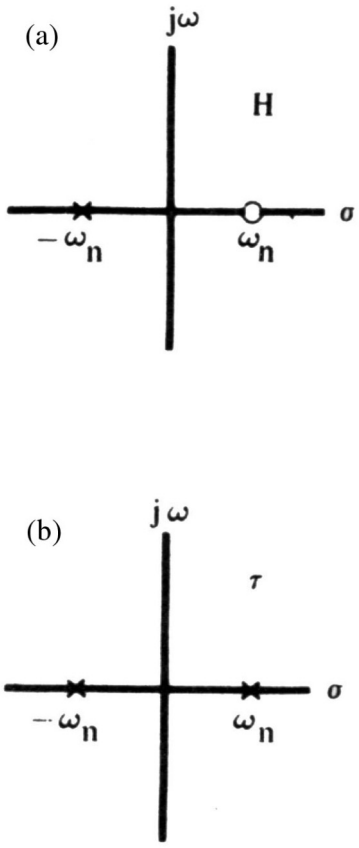


Fig. 3.2. Pole-zero pattern for (a) gain and (b) delay for a first-order all-pass filter. [C. L. Lindquist, *Active Network Design*, (Steward & Sons, California, 1977), pp. 145, Fig. 2.11.1 (a)]

The second-order all-pass filter has complex poles and gain

$$H(s) = k\frac{s^2 - 2\zeta\omega_n + \omega_n{}^2}{s^2 + 2\zeta\omega_n + \omega_n{}^2} \tag{3.10}$$

Its delay equals

$$\tau(j\omega) = \frac{4\zeta}{\omega_n}\frac{1 + (\omega/\omega_n)^2}{1 + (2\zeta^2 - 1)2(\omega/\omega_n)^2 + (\omega/\omega_n)^4} \tag{3.11}$$

The pole-zero pattern for the gain and delay are shown in Fig. 3.3. Since the poles of $H(s)$ lie on a circle of radius ω_n, so do the poles of the delay. Also note that the RHP zeros of $H(s)$ possess the same delay poles as the LHP poles of $H(s)$. The delay zeros lie a s = ± ω_n.

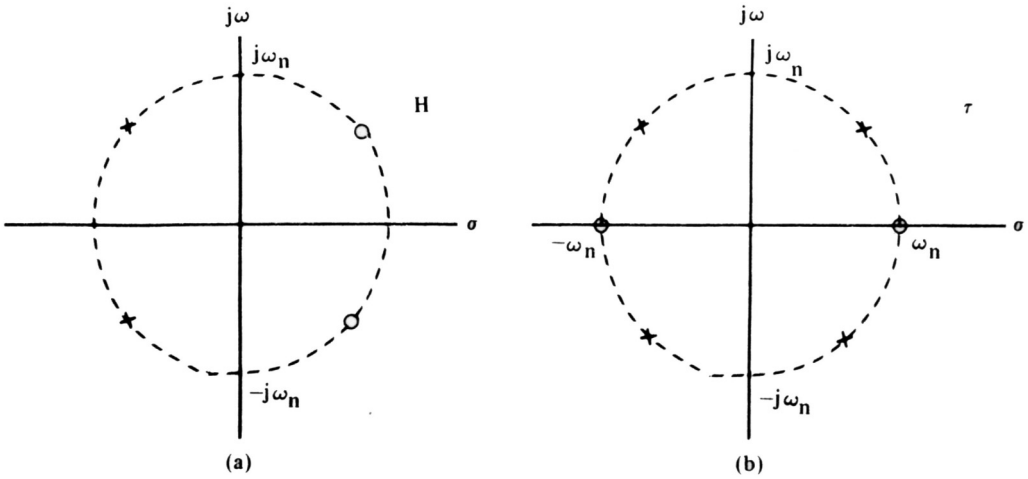


Fig. 3.3. Pole-zero patterns for (a) gain and (b) delay for a second-order all-pass filter. [C. L. Lindquist, *Active Network Design*, (Steward & Sons, California, 1977), pp. 146, Fig. 2.11.2 (a), (b)]

A wide variety of applications that require tunable all-pass filters (variable Q and constant ω_0 or vice versa) exist. One application that requires tunable all-pass filters is delay equalization. Delay equalizers are all-pass filters where its delay is adjusted to equalize the delay of another filter to render the overall delay constant.

A formulation of a tunable second-order all-pass filter will now be shown [20]. From

$$\tau_p = \tau(j\omega_p) = \frac{\zeta}{2}\frac{1}{(1-\zeta^2)^{1/2}\left[1-(1-\zeta^2)^{1/2}\right]} = \frac{1}{2\zeta}\left[1+(1-\zeta^2)^{-1/2}\right], \quad |\zeta| < 0.866 \tag{3.12}$$

that describes the maximum delay of a second-order low-pass gain function $1 / (s^2 + 2\zeta s + 1)$ and

$$\tau(j\omega) = \frac{1}{\omega_n}\tau_n(j\omega/\omega_n) \tag{3.13}$$

that describes the frequency-denormalized delay or a first-order low-pass gain function $1 / (s + 1)$, the maximum delay and resonant frequency are

$$\tau_{\max} \cong \frac{2}{\zeta\omega_n} = \frac{4Q}{\omega_n}, \quad \omega_n \cong \omega_p \tag{3.14}$$

where $Q >> 1$ or $\zeta << 1$ is assumed.

The unnormalized transfer function of an all-pass filter is

$$H_{AP}(s) = \frac{s_n^{\,2} - s_n/Q + 1}{s_n^{\,2} + s_n/Q + 1} \tag{3.15}$$

where $s_n = 1/\omega_0$ and $\omega_0 = 1/2\pi f_0$. These substitutions yield

$$H_{AP}(s) = \frac{\left(\frac{s}{2\pi f_0}\right)^2 - \frac{1}{Q}\left(\frac{s}{2\pi f_0}\right) + 1}{\left(\frac{s}{2\pi f_0}\right)^2 + \frac{1}{Q}\left(\frac{s}{2\pi f_0}\right) + 1}. \tag{3.16}$$

Rearranging Eq. (3.14) yields

$$Q \cong \frac{\tau_{\max} \pi f_0}{2}. \tag{3.17}$$

Now, the all-pass filter can be specified in terms of desired center frequency f_0 and peak group delay $\tau_{\max}$. It is advantageous to apply the above result in the test signals for the listening test since what is tested here is the *amount* of group-delay (or phase) distortion. It is of importance to note here that the above approximation has larger errors at frequency extremes. Thus, some manual compensation by inspection is necessary in the formulation of all-pass filters with either very low center frequencies ($f_0 < 100$ Hz) or with very high frequencies ($f_0 > 10000$ Hz).

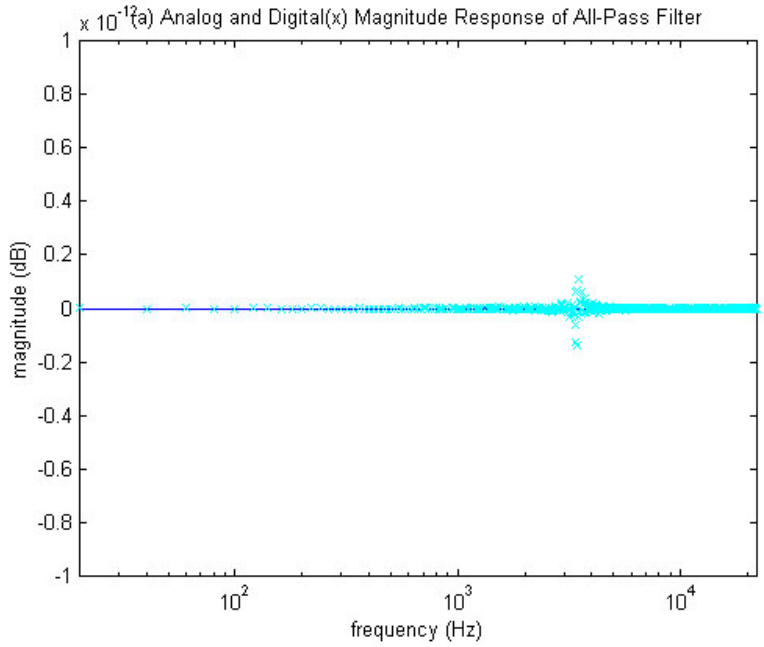
A numerical integration transformation considered here is the bilinear transform. The transformation from the analog to digital domain is:

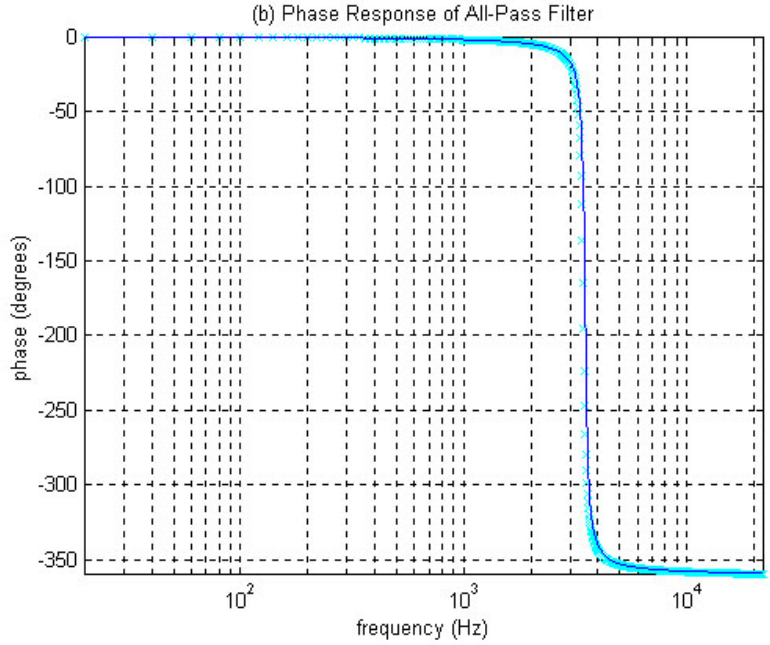
$$p_n = \frac{1}{C}\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right) \tag{3.18}$$
$$C = \tan(BT/2)$$

where p_n is the normalized analog frequency and C is the analog frequency normalization constant. Since the bilinear transform maps the p_n plane imaginary axis

on to the unit circle and the left-half p_n-plane maps inside the unit circle, stable and causal analog filters $H(p)$ will always produce stable and causal digital filters $H(z)$. This transformation exhibits good impulse response characteristics and low implementation complexity. Because of the good correlation with the analog response, the bilinear transform is a good candidate for implementation of the digital all-pass filter.

MATLAB was used in the formulation of the digital all-pass filter. Fig. 3.4 (a), (b), and (c) display the magnitude, phase, and group-delay response respectively for a second-order all-pass filter with peak group delay set to 4 msec and center frequency at 3500 Hz. Both analog and digital plots are shown. It is clear that the bilinear transform offers excellent correlation for both the analog and digital plots. Fig. 3.5 (a), (b), and (c) display the magnitude, phase, and group-delay response respectively for a second-order all-pass filter with peak group delay set to 8 msec and center frequency at 3500 Hz. Again, excellent correlation for both the analog and digital plots are displayed.


x 10^-12
(a) Analog and Digital(x) Magnitude Response of All-Pass Filter
magnitude (dB)
frequency (Hz)


(b) Phase Response of All-Pass Filter
phase (degrees)
frequency (Hz)

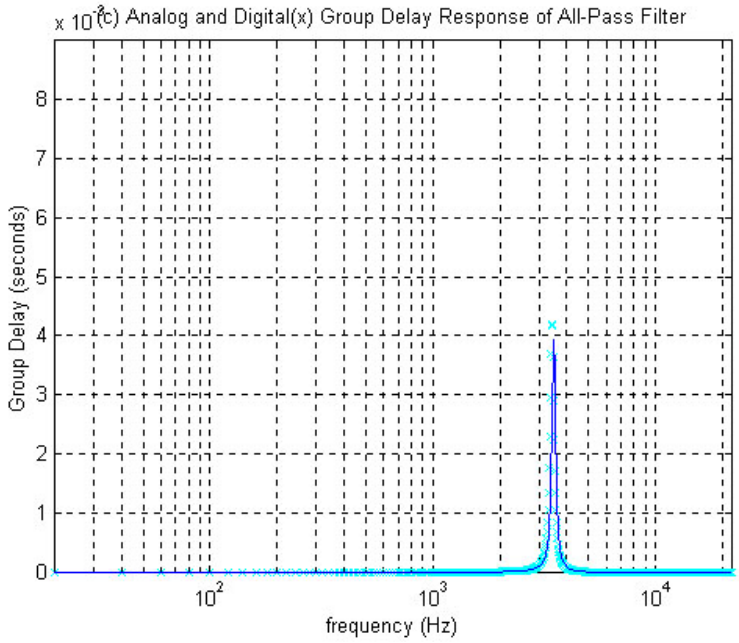
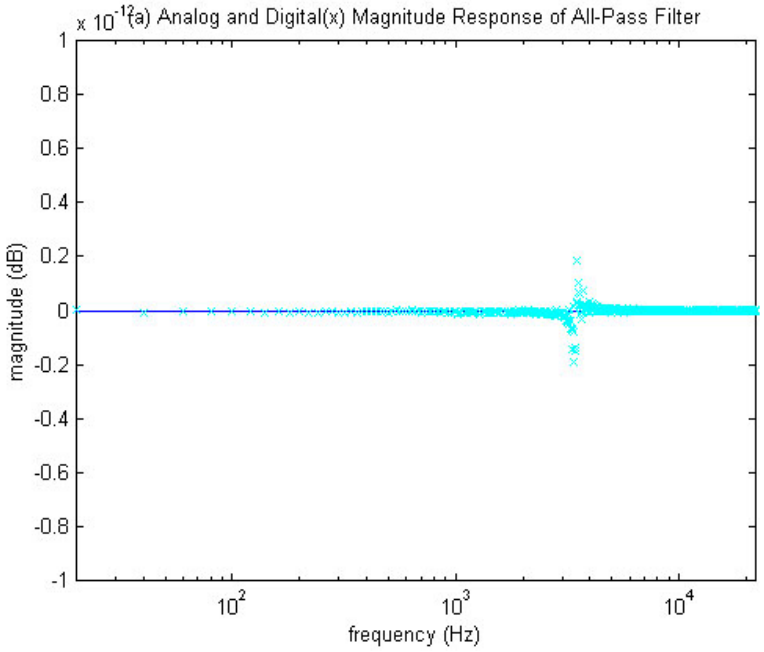


Fig. 3.4 (a), (b), and (c). The magnitude, phase, and group-delay response respectively for a second-order all-pass filter with peak group delay set to 4 msec and center frequency to 3500 Hz.



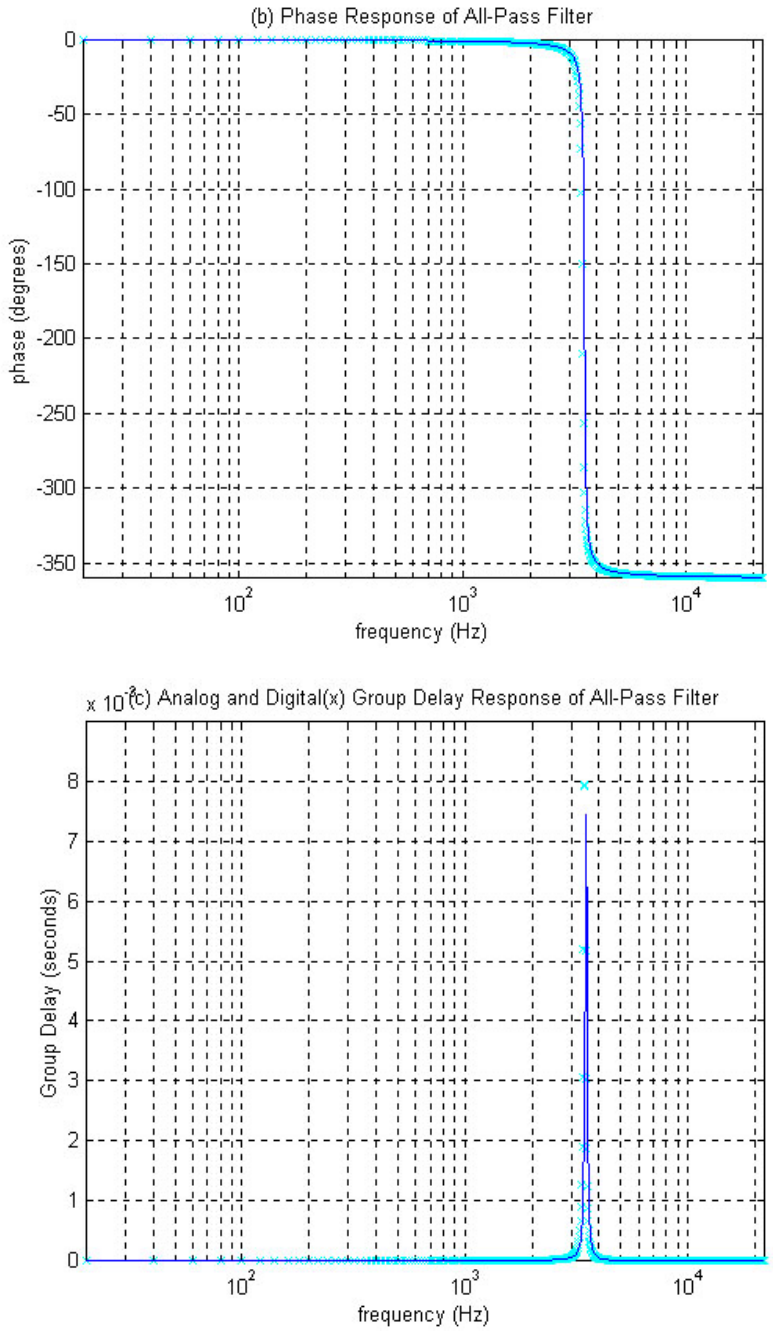


Fig. 3.5 (a), (b), and (c). The magnitude, phase, and group-delay response respectively for a second-order all-pass filter with peak group delay set to 8 msec and center frequency to 3500 Hz.

3.2 Test Signal Formulation and Analysis

The test signals used in the experiment are:

1. 70 Hz sawtooth wave
2. 3.5 kHz sawtooth wave
3. 10 kHz sawtooth wave
4. Impulse
5. Jazz-vocal group
6. Percussion instruments.

The sawtooth wave was chosen as a suitable test signal for its desirability in Lipshitz *et al.*'s [7] research results. Due to its rich harmonic content, the sawtooth wave was found to be the most audible waveform when phase distortion was present. The frequency 70 Hz was chosen as one of the frequencies for the sawtooth wave since signals with low-frequency content were found to have characteristics where phase distortion was readily audible [16]. 3.5 kHz was chosen for the next sawtooth wave since the Robinson-Dadson equal-loudness curves (Fig. 3.6) revealed that the human ear is most sensitive to this area, albeit for pure tones. Each curve is representative of a range of frequencies which are perceived to be equally loud. In selecting 3.5 kHz, it was assumed since this frequency provides maximal sensitivity to loudness (a primary factor), any detection of phase distortion (a secondary factor) would be facilitated.

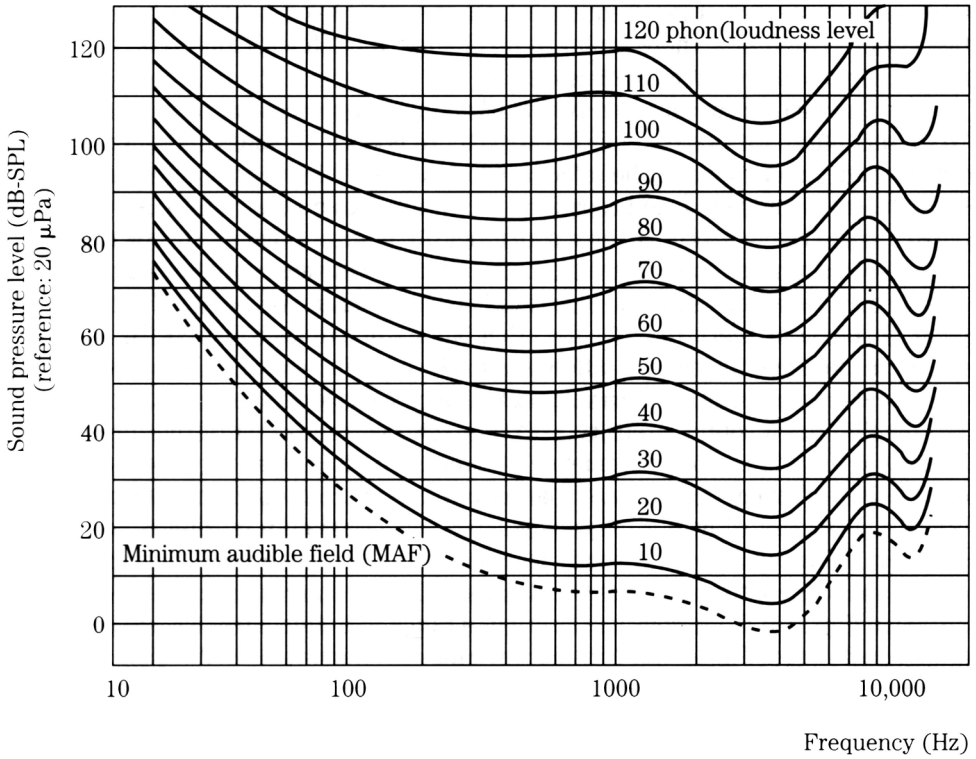


Fig. 3.6. The Robinson-Dadson equal-loudness curves show that the ear has non-linear characteristics with respect to frequency and loudness. These curves are based on psychoacoustic research using pure tones. [K. C. Pohlmann, *Principles of Digital Audio*, (McGraw-Hill, New York, 1995), pp. 360, 11-3.]

10 kHz was chosen for the third and final sawtooth wave because this was exactly twice the frequency of threshold (f > 5kHz) where spike discharges can lock to one phase of the stimulating waveform. Theory [5] states that the ear cannot latch on to the phase of the stimulating waveform at such high frequencies. All sawtooth waves were 1 second in duration.

The impulse (a '1' with trailing zeros) was chosen because of its short time- and broad frequency-domain characteristics. The male and female acapella jazz-vocal group, with its rich spatial content was a good candidate to test the perceived alteration of spatial qualities as a function of phase distortion incurred [22].

The percussion instrument passage was chosen to test the audibility of phase distortion in real instruments of transient content, besides a pure impulse. Both the

jazz-vocal group and percussion instruments were recorded at the University of Miami using the recording facilities available.

Based on the research of Preis *et al.* [17] and Deer *et al.* [18], 4 msec and 8 msec were the delay times chosen for the peak delays of the all-pass filter since they gave significant results for an impulsive test signal. These delay times were assumed to provide a valid reference starting point for other test signals. As referenced to the no peak-delay condition 0 msec, 4 msec is defined as 'medium' phase distortion and 8msec as 'high' phase distortion.

The center frequency, f_0 chosen for the all-pass filters were:

1. 70 Hz Sawtooth Wave: $f_0 = 70$ Hz
2. 3.5 kHz Sawtooth Wave: $f_0 = 3.5$ kHz
3. 10 kHz Sawtooth Wave: $f_0 = 10$ kHz
4. Impulse: $f_0 = 3.5$ kHz
5. Jazz -Vocal Group: $f_0 = 160$ Hz
6. Percussion Instruments: $f_0 = 150$ Hz

3.5 kHz was chosen as a center frequency for the impulse since the introduction of phase distortion in the region assumed most sensitive to the human ear in the Robinson-Dadson curves should maximize audible phase distortion differences. 160 Hz was chosen as an all-pass filter center frequency for the jazz-vocal group since a spectogram analysis (Fig. 3.7) of the signal revealed significant frequency content in this area. Introduction of phase distortion in this region should maximize the audibility of its effect. The same reasoning applies for the 150 Hz center frequency for the percussion ensemble (Fig. 3.8).

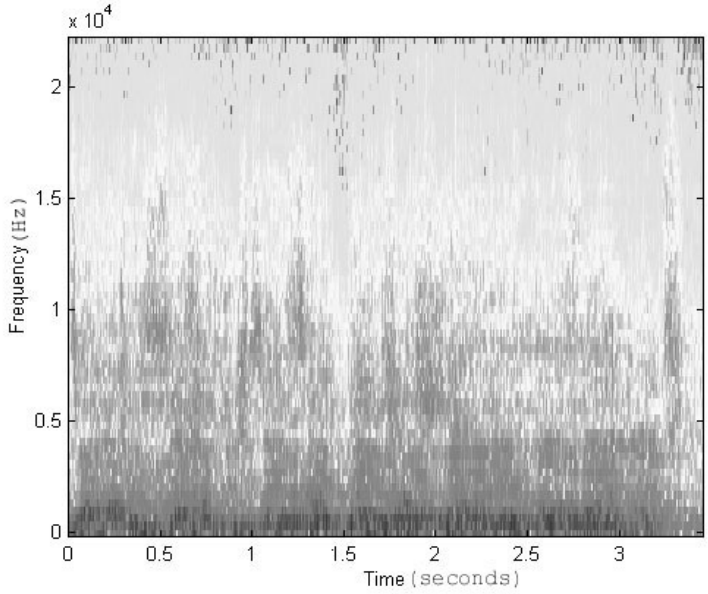


Fig. 3.7. 128 point FFT, 44100 Hz sampling frequency spectogram for the jazz-vocal group, left channel only.

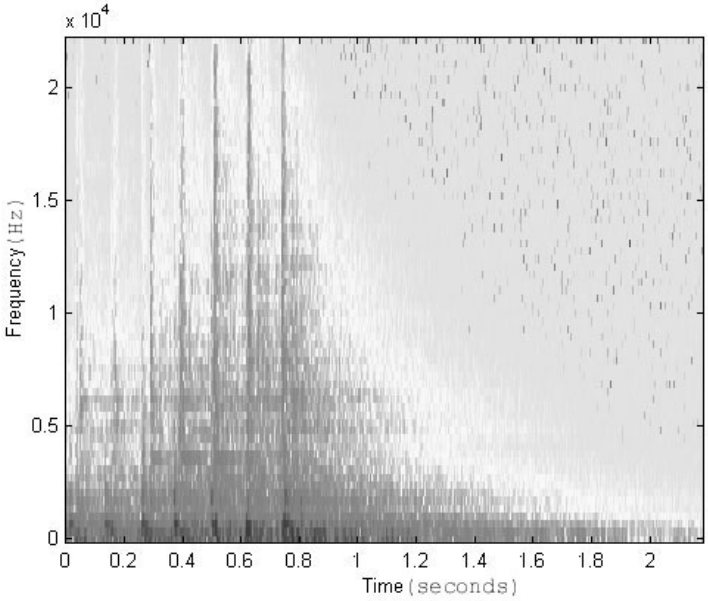


Fig. 3.8. 128 point FFT, 44100 Hz sampling frequency spectogram for the percussion ensemble, left channel only.

Fig. 3.9 and 3.10 are the unfiltered and all-pass filtered impulse test signals for maximum group delay 4 msec and 8 msec respectively. Close inspection of these

plots reveal that the 4 msec all-pass filtered impulse has significantly more waveform distortion that the 8 msec version. This is due to the fact that the 4 msec version has a broader Q (as a result of the all-pass filter formulation that was based on the desired maximum delay τ_{max}) than the 8 msec version. However, obviously the 8 msec all-pass filter has greater maximum delay at its center frequency.

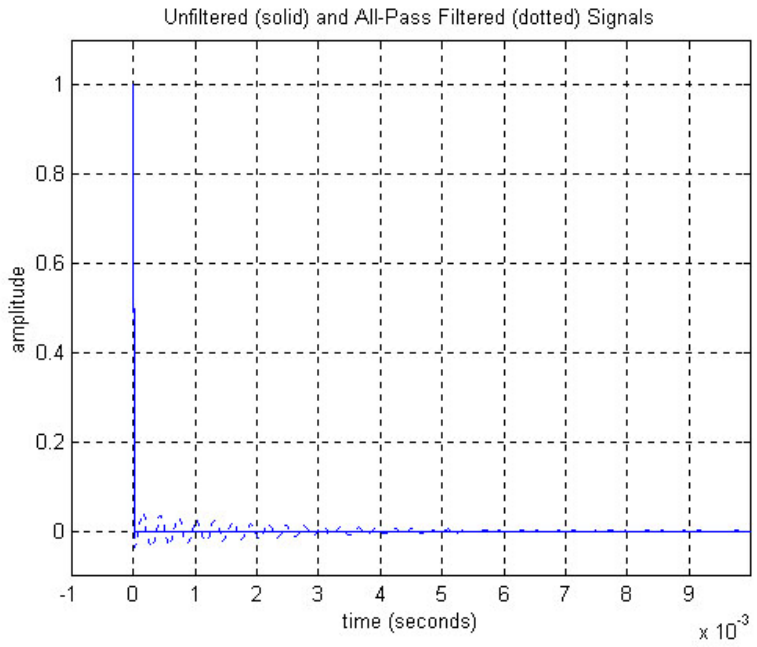


Fig. 3.9. Unfiltered (solid) and all-pass filtered (dotted) impulse. For the all-pass filter, the center frequency, f_0 = 3.5 kHz and the maximum group delay τ_{max} = 4 msec.

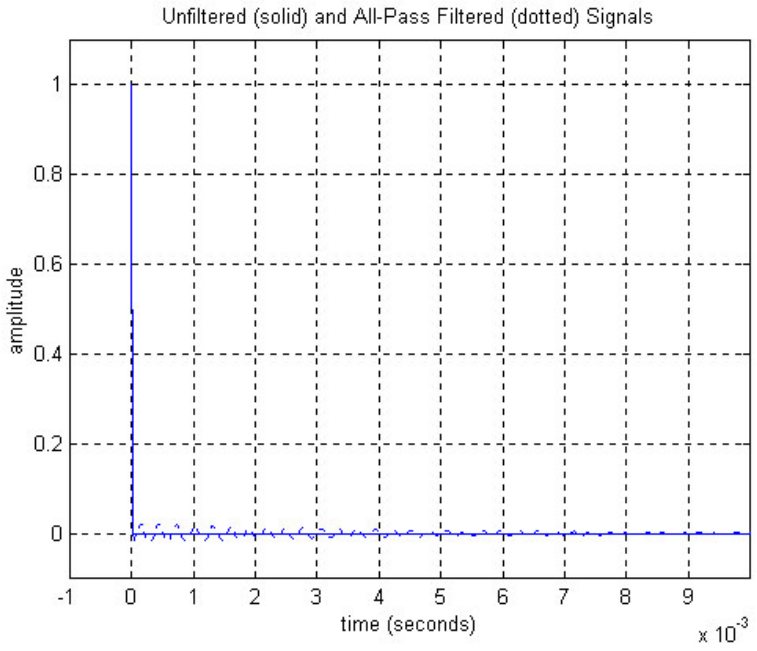


Fig. 3.10. Unfiltered (solid) and all-pass filtered (dotted) impulse. For the all-pass filter, the center frequency, f_0 = 3.5 kHz and the maximum group delay τ_{max} = 8 msec.

Fig. 3.11 and 3.12 are the unfiltered and all-pass filtered 3.5 kHz sawtooth waves for maximum group delay 4 msec and 8 msec respectively. Again, due to the all-pass filter formulation used, it seems that the 4 msec all-pass filtered signal has incurred more severe waveform distortion than the 8 msec version.

This chapter introduced the all-pass filter and audio test signals used in the listening test. The steps in formulating the all-pass filter were outlined. This was followed up by a formulation of a tunable second-order all-pass filter. The bilinear transform was introduced to perform the analog to digital transformation for the all-pass filter. Finally, test signals used in the listening test were introduced and discussed.

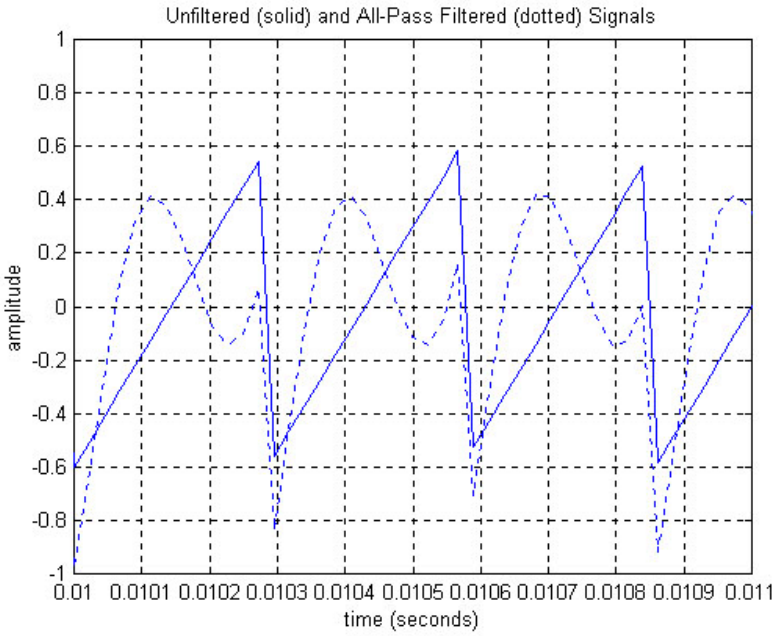


Fig. 3.11. Unfiltered (solid) and all-pass filtered (dotted) 3.5 kHz sawtooth wave. For the all-pass filter, the center frequency, f_0 = 3.5 kHz and the maximum group delay τ_{max} = 4 msec.

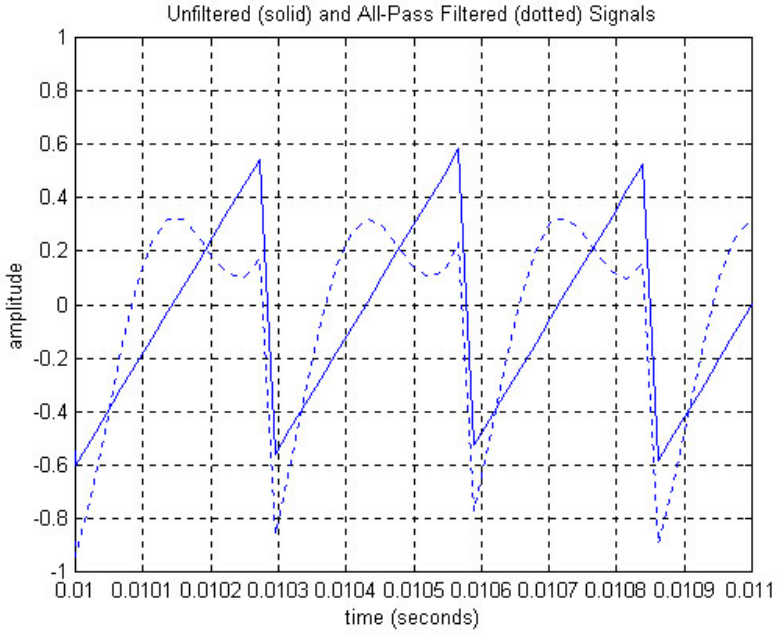


Fig. 3.12. Unfiltered (solid) and all-pass filtered (dotted) 3.5 kHz sawtooth wave. For the all-pass filter, the center frequency, f_0 = 3.5 kHz and the maximum group delay τ_{max} = 8 msec.

It was shown that the formulation of a tunable second-order all-pass filter facilitated implementation of test signals. The next chapter will incorporate the test signals selected in this chapter for the listening test implementation

4. Listening Test Formulation

Previous research regarding a valid listening test design will be investigated. It is vital that a sound psychoacoustic test design is adopted for implementation so that acquisition of the most accurate data is possible. It is paramount in conducting listening tests with the most accurate audio equipment available if the test subject is to hear any audible effects.

4.1 Previous Research

An interesting study was conducted by Sergeant and Boyle [23] regarding comparisons of various psychoacoustic tests of pitch discrimination to see what is "the 'best' measure in relation to efficiency, or validity for general musical behavior" by examination of various test structures, stimulus differences, and response formats. Although pitch discrimination was not an area tested in this research at hand, Sergeant and Boyle's study has relevance in that the phase distortion listening test implementation can be significantly improved by observance of their results. The comparisons of the five pitch tests (Bentley, Colwell, Kwalwasser-Dykema, Seashore, and Sergeant) are summarized in Table 4.1. Of interest here is the variability of task structures among the different test implementations. The K-D test and the Sergeant pitch test require determination of the presence or location of pitch change. However, the others (Bentley, Colwell, and Seashore) require the determination of both the presence or location of pitch change *and* pitch direction. It would appear that the two-step task structure would be more difficult than a one-step task structure. The

response formats used in the various tests also reflect accommodation of the different task structures.

		Stimulus Characteristics							Task	
Test	*Number of items*	(1) *Signal*	(2) *Number of tones*	(3) *Frequency level*	(4) *Pitch deviation range*	(5) *Duration ind. tones*	(6) *Duration between tones*	(7) *Dura. of response interval*	(8) *Tot. dura. between items*	*Determine pitch direction*	*Response format*
Bentley	20	Pure tone	2	440 Hz	Approx. 12–100 cents	1 sec.	None	6 sec.	8 sec.	Yes	Write U, D, or S in box
Colwell	15	Violin or 'cello	2	Varied E Flat³–D' (77·8 Hz–2349·3 Hz)	100 cents (semitone)	1·5 sec.	None	4 sec.	8 sec.	Yes	Mark approp. box
K–D	40	Pure tone	One sustained or with one-third of it changed	500 Hz and 1000 Hz	Approx. 9–80 cents	3 sec. total, 1 sec. each one-third	None	None	3 sec.	No	Write S or D in box
Seashore	50	Pure tone	2	500 Hz	Approx. 9–59 cents	0·6 sec.	Approx. 0·2 sec.	3 sec.	3 sec.	Yes	Mark blank below approp. letter
Sergeant	30	Square wave, A.S.D.R. modulated	5	290 Hz	Approx. 5 cents—semitone	1 sec.	1 sec.	5 sec.	10 sec.	No	Circle dot of tone which differs

Table 4.1. Comparison of five pitch discrimination tests. [D. Sergeant and J. D. Boyle, "Contextual Influences on Pitch Judgement," *J. Soc. for Res. in Psych. of Music and Music Ed.*, vol. 8, pp. 3 - 15 (1980), pp. 7, Table 1.]

Table 4.2 displays the means and standard deviations for the scores of 65 subjects on each of the five tests.

Test	*Mean*	*Standard Deviation*
Bentley Model	22.52	4.73
Colwell Model	22.49	4.18
K-D Model	29.09	2.14
Seashore Model	20.62	5.14
Sergeant Model	28.74	2.15

Table 4.2. Means and standard deviations of subjects' responses to five tests (n = 65). [D. Sergeant and J. D. Boyle, "Contextual Influences on Pitch Judgement," *J. Soc. for Res. in Psych. of Music and Music Ed.*, vol. 8, pp. 3 - 15 (1980), pp. 12, Table 4.]

It can be seen that the Kwalwasser-Dykema model has the highest test mean (out of 30 total) and the lowest standard deviation. Therefore, a phase distortion audibility test implementation based on the K-D test should provide very accurate and hence reliable data acquisition. The Sergeant model was also considered, but was not implemented since it was determined that its implementation may be inappropriately long for some of the music-based test signals.

4.2 Test Equipment

The audio equipment used in the implementation consisted of

1. Personal computer
2. Yamaha 01V digital mixing console
3. Genelec 1030A loudspeakers
4. AKG K270 headphones.

The personal computer contained a studio-quality sound card (Micro Technology Unlimited, Krystal) that was used to play back the test signals. The Yamaha 01V mixing console provided pre-amp gain for the self-powered loudspeakers and headphone amplification for the headphones. The studio-quality Genelec loudspeakers have a free-field frequency response of 55 Hz - 18 kHz (± 2.5 dB). This on-axis (0°) plot, along with off-axis plots of 15°, 30°, and 45° along with the 1/3 octave power response is shown in Fig. 4.1. The off-axis characteristics of the loudspeaker display a smooth and well-behaved transition. Their harmonic distortion at 90 dB SPL at 1 meter was specified at < 3% for 60 - 150 Hz and < 0.5% for > 150 Hz. The loudspeakers incorporate a waveguide on the tweeter, which focus dispersion

and aids in minimizing unwanted first-order reflection contributions in the sound heard in the listening position.

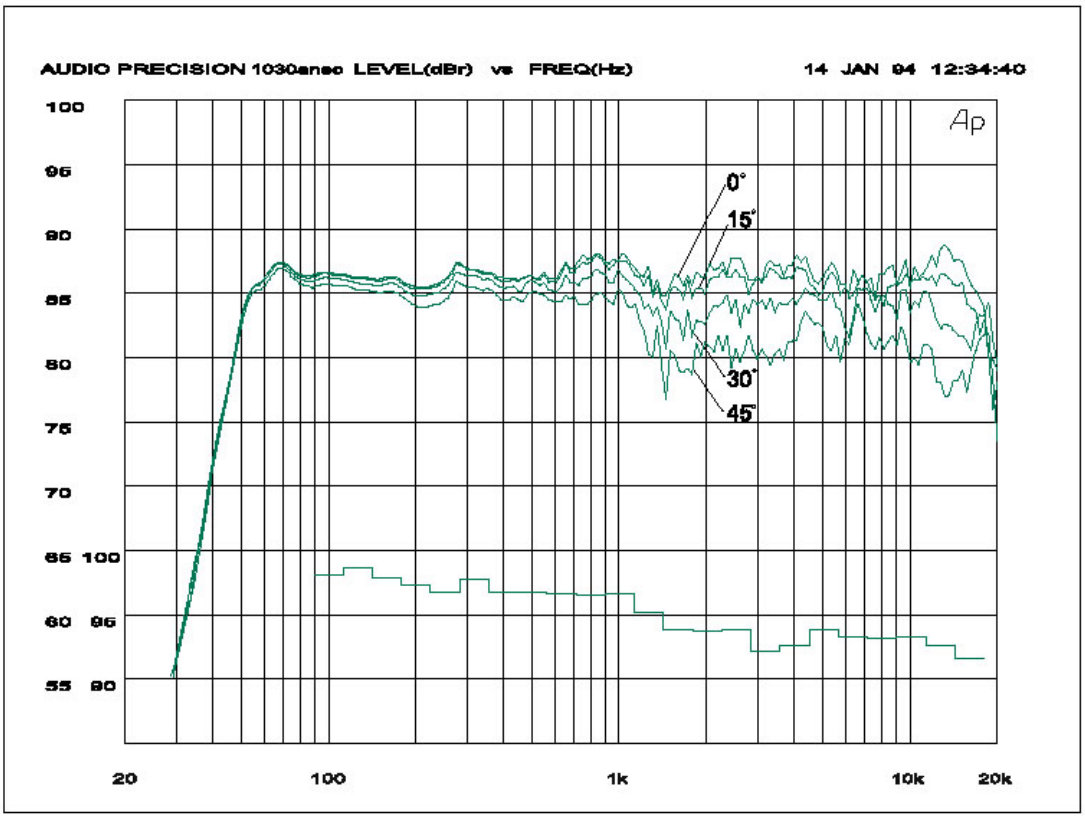


Fig. 4.1. Anechoic frequency response (0°, 15°, 30°, and 45°) and 1/3 octave band power response plot for Genelec 1030A loudspeaker. [Genelec Data Sheet for 1030A Bi-amplified Monitoring System, pp. 3, Genelec Oy, Finland (1999)]

Phase characteristics for this loudspeaker were not investigated. The AKG headphones have a published frequency-response specification of 20 Hz to 28,000 Hz. Again, the phase response for the headphones was not investigated. As far as the frequency response of both the headphones and loudspeakers were concerned, they were both adequate to carry out the test signal content.

4.3 Test Implementation

All test subjects were given clear oral instructions regarding how to proceed with the test. Fig. 4.2 is a copy of the instructions handed to the subjects for viewing purposes.

INSTRUCTIONS

1. Please locate and open (double click) the folder with the appropriate *sound* folder corresponding to the scoresheet.
2. Please locate and open (double click) the folder with the appropriate *version* number folder corresponding to the scoresheet.
3. Lower the headphone or loudspeaker output initially so that excessive levels are not produced.
4. Two sounds, played one after another, are contained in each sound file (a) ~ (d).
5. Play back the sound file (a) ~ (d) by double clicking on them and adjust headphone or loudspeaker output to comfortable levels. (sounds may be played as many times as desired)
6. Locate the corresponding sound (a) ~ (d) on the scoresheet.
7. Please indicate if a difference (i.e. in timbre, loudness, etc.) was heard between the two sounds by checking on the appropriate 'Yes' or 'No' box.
8. When finished with the headphone test, please inform me to set up the loudspeaker test.

IMPORTANT: Take care to make sure that you are in the correct *sound* and *version* folder.

Thank you very much for your participation.

Fig. 4.2. Instruction sheet for the listening test implementation.

The listening test was implemented by a randomized arrangement of A, the unfiltered test signal and B, the all-pass filtered test signal as AA, AB, BA, or BB. The synthesized test signals were presented in mono and acoustic test signals in

stereo. The test subject was free to adjust both headphone and loudspeaker levels for each test signal to comfortable levels. The randomized presentation order of the test signals was different for the headphone and loudspeaker test implementations. Care was taken regarding the onset portion of the test signals, which perception of timbre is highly dependent on, were audible by the subjects. A 'blank' time of 50 msec duration was placed in the beginning and between all test signals except for the jazz vocal, which received 500 msec. The longer 'blank' time was determined necessary for the relatively long and complex jazz vocal test signal to facilitate phase distortion detection.

An implementation where test signals were presented entirely in increasing order of difficulty has the possibility of subjects developing response expectancies regarding the relative phase distortion differences among subsequent items presented. Also, it was found in Sergeant and Boyle's study [23] that wholly random sequencing of test items had "the result that a subject has no chance to acclimatize himself to the test before critical judgements are demanded of him" therefore putting some difficulty in the beginning of a test. Therefore, test items were first sequenced in order of difficulty and then randomized for the rest of the test to optimize data acquisition.

Upon listening to the test signal sequence, the test subjects had to indicate by marking a box if a difference was heard or not by a 'Yes' or 'No' box, as shown in Fig. 4.3. This implementation is similar to the K-D Model except it simplifies the response format slightly further by simply questioning if a difference existed. This was repeated for all six test signals implemented on both headphones and loudspeakers.

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

Fig. 4.3. Task response format used for listening test. Version numbers refer to randomized maximum group delay times of either 4 or 8 msec.

The loudspeakers used in the listening test were set up for near-field listening (~1 meter from left and right loudspeaker to respective ear) and the acoustic centers of the loudspeakers were located on the same plane as the test subject's ears. The test subject sat equidistant from the left and right loudspeakers. In this way, the sound received by the test subject was predominantly the on-axis direct sound from the loudspeakers.

The listening room, which was a control booth for a concert hall, had carpeted floors and concrete walls. The corners closest to the audio equipment were treated with acoustic diffusers/absorbers. Therefore, the listening environment had a semi-reverberant characteristic.

All test subjects apparently had normal hearing, but were not tested. The majority of test subjects was musicians and had recording studio experience (meaning they were critical listeners). Listener fatigue was attempted to be minimized by focusing on relevant aspects of the test implementation such as a clean and simple test design

and relatively short overall length. The average listening test had the duration of 40 minutes with the longest duration being 1 hour.

This chapter introduced the listening test formulation process. A study was shown which compared the validity of various psychoacoustic listening test implementations. Test equipment used in this thesis research was presented. Finally, the listening test implementation used in the thesis research was outlined. It is important that a sound psychoacoustic test design is implemented so that accurate data acquisition is possible. It is also stressed that in listening tests, the most accurate audio equipment available should be used if the test subject is to hear any audible effects. The results of the listening test implementation will be presented in the next chapter and discussed

5. Results and Discussion

Performing a fair statistical analysis is important for useful interpretation of the acquired listening test data and is presented. The results of the listening test are presented, which is followed by a discussion. A discussion of the resultant statistical analysis to bring about implications aided by previous cumulative knowledge provides further explanation of the results.

5.1 Results

When statistically analyzing listening test data using the conventional significance level ($\alpha = 0.05$), employing a small number of trials or listeners may produce a high risk of concluding that audible differences are inaudible (type 2 error) [24]. This risk can be absolutely and relatively large as compared to the risk of concluding that inaudible differences are audible (type 1 error). Care must be taken so that the type 2 error does not obliterate type 1 error.

The *significance test* is statistically testing scientific propositions by deduction of two incompatible statistical hypothesis, H_0 and H_1, and determining the plausibility from research data of rejecting H_0 in favor of H_1. H_0, the *null hypothesis*, says that the proportion of correct identifications p in the conceptual population of trials is 0.5. This has the statistical implication that says the differences between the two components under test are not audible and the listener will perform at chance. Therefore

H_0: $p = 0.5$.

H_1, the *alternative hypothesis,* says that p is greater than 0.5. This has the statistical implication that differences are audible and that the listener will perform above chance. Therefore

H_1: $p > 0.5$.

H_1 is said to be *directional* in this case. The task then, is to decide whether to reject H_0 in favor of H_1.

The experimenter's decision regarding the rejection of H_0 in favor of H_1 can be correct in two ways and incorrect in two ways, as shown in Fig. 5.1. Cells *a* and *c* show two possibilities when H_0 is true. Cell *a* represents the correct decision of not rejecting H_0, which has the probability of a correct decision 1 - α. Cell *c* represents the incorrect decision of rejecting H_0 and is designated as type 1 error, symbolized by α. Cells *b* and *d* illustrate the two possible decisions when H_0 is false. Cell *b* represents the incorrect decision of not rejecting H_0 and is designated as type 2 error, symbolized by β. Cell *d* represents the correct decision of rejecting H_0, which has the probability of correct decision 1 - β.

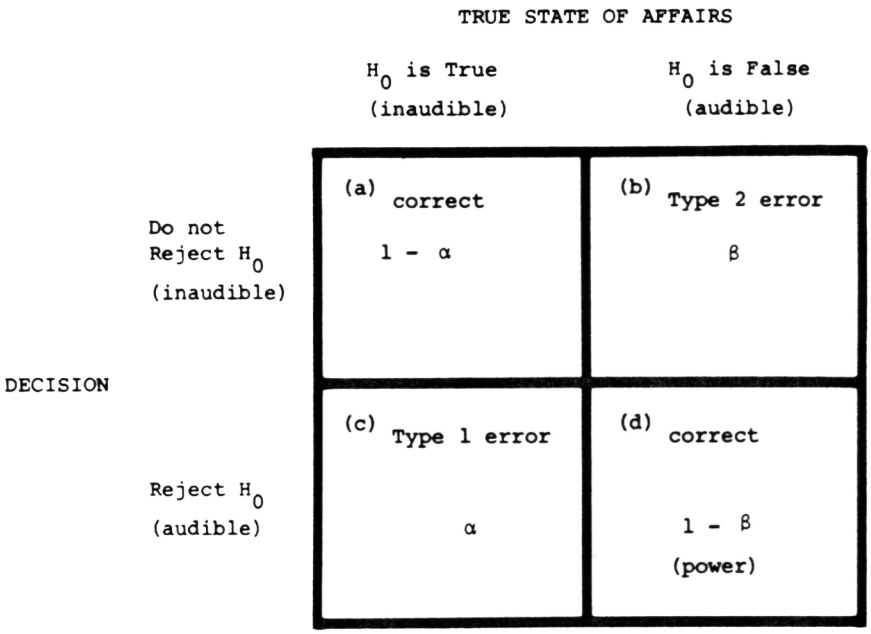


Fig. 5.1. 2 x 2 table illustrating the two correct and two incorrect decisions possible when deciding whether to reject H_0 in favor of H_1. Indicated in parentheses are scientific interpretations frequently made of the column and row headings regarding the audibility of differences. [L. Leventhal, "Type 1 and Type 2 Errors in Statistical Analysis of Listening Tests," *J. Audio Eng. Soc.,* vol. 34, pp. 437-453 (1986 Jun.), pp. 440, Fig. 1.]

The *power* of a statistical procedure is the degree that it produces a high probability of rejecting H_0 when H_0 is false, as indicated in cell *d*. Anything that can reduce the type 2 error probability will increase statistical power. Thus, the power of a statistical analysis of a listening test is the probability that its analysis will uncover the ability to hear differences in the presence of ability. Fig. 5.2 shows that the probability of committing a type 2 error (β) decreases as sample size (N) increases, and therefore stating that the power increases with the size of N. These samples are taken from normal populations with variance σ^2 and the mean under the null hypothesis is symbolized on the figure as μ_0.

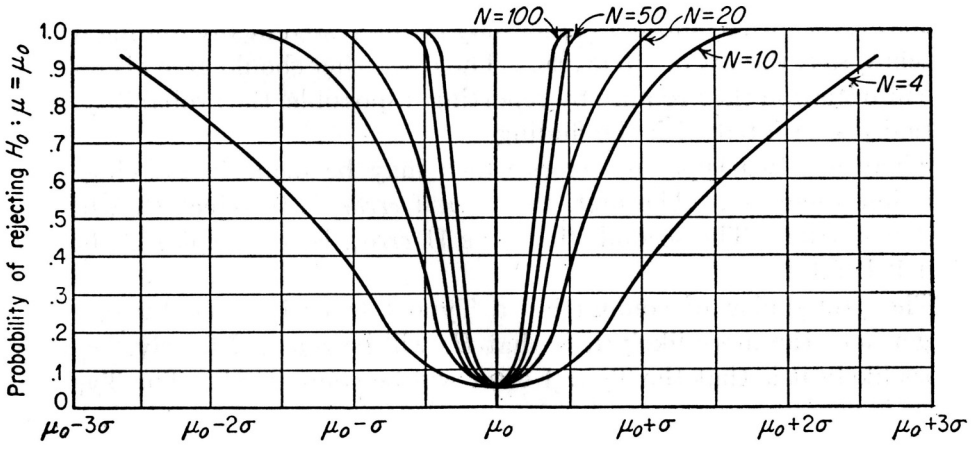


FIG. 1. Power curves of the two-tailed test at $\alpha = .05$ with varying sample sizes.

Fig. 5.2. Power curves of the two-tailed test at $\alpha = 0.05$ with varying sample sizes. [S. Siegel, *Nonparametric Statistics*. (McGraw-Hill, New York, 1956), pp. 10, Fig. 1.]

In studies of small N (number of trials), the $\alpha = 0.05$ significance level usually produces type 2 error larger than type 1 error. Equalizing both errors usually requires reduction of type 2 error, since it is desirable to keep both errors as small as possible. There are three ways of reducing type 2 error in a listening test.

The first way is to increase N. This requires more trials or listeners and is the preferred method to decrease type 2 error. However, one must increase N with care since more trials can be conducive to fatigue or boredom and jeopardize accuracy of results.

The second way is to increase p. While p is not known most of the time, careful design of the listening test can help increase it. For example, utilization of carefully selected audio signals, methods of irradiation (i.e. headphones, loudspeakers), giving the listener warm-up trials, preventing fatigue, boredom, and distractions, and using a familiar room and equipment are some of methods of enhancing this ability. However,

these procedures by themselves cannot be relied upon to reduce type 2 error since their effect is speculative and difficult to measure.

The third way is to increase type 1 error. This method of increasing type 1 error to decrease type 2 error should only be a last resort, upon the attempt to increase N and p. However, this may be necessary, particularly if it is important to avoid type 2 errors that have p values which are just slightly above chance.

A listening test which is fair to both sides of the study should incorporate equal probabilities of type 1 and type 2 errors for the p of interest. For such a study, the *fairness coefficient* FC_p, a measure of the degree to which the two error risks have been equalized for a given p, is a convenient figure of merit. For a study with type 1 error probability α and a type 2 error probability for a given p, β_p, the fairness coefficient is

$$\mathrm{FC}_p = \frac{\text{smaller probability}}{\text{larger probability}}. \tag{5.1}$$

An FC_p of 1.0 represents for a given p, a perfectly fair study, although this is usually impossible to achieve since α and β change abruptly in unhelpful increments for adjacent values of r.

Average correct responses for the headphone listening test data are as follows in Table 5.1. The AA and BB test signal pairs were discarded since they do not provide information for phase distortion detection and thus only the AB and BA pairs were considered for this calculation.

Test Signal		Average Correct Responses ($N = 15$)
70 Hz Sawtooth Wave	$\tau_{max} = 4$ msec	9
70 Hz Sawtooth Wave	$\tau_{max} = 8$ msec	9.5
3.5 kHz Sawtooth Wave	$\tau_{max} = 4$ msec	5.5
3.5 kHz Sawtooth Wave	$\tau_{max} = 8$ msec	4
10 kHz Sawtooth Wave	$\tau_{max} = 4$ msec	1.5
10 kHz Sawtooth Wave	$\tau_{max} = 8$ msec	4.5
Impulse	$\tau_{max} = 4$ msec	13
Impulse	$\tau_{max} = 8$ msec	14
Jazz-Vocal	$\tau_{max} = 4$ msec	6.5
Jazz-Vocal	$\tau_{max} = 8$ msec	4
Percussion Instruments	$\tau_{max} = 4$ msec	4
Percussion Instruments	$\tau_{max} = 8$ msec	4

Table 5.1. Average correct responses for 15 test subjects performing the headphone test.

Average correct responses for the loudspeaker listening test data are as follows in Table 5.2. Again, the AA and BB test signal pairs were discarded and only the AB and BA pairs were considered for this calculation.

Test Signal		Average Correct Responses ($N = 15$)
70 Hz Sawtooth Wave	$\tau_{max} = 4$ msec	7
70 Hz Sawtooth Wave	$\tau_{max} = 8$ msec	7
3.5 kHz Sawtooth Wave	$\tau_{max} = 4$ msec	5
3.5 kHz Sawtooth Wave	$\tau_{max} = 8$ msec	3.5
10 kHz Sawtooth Wave	$\tau_{max} = 4$ msec	2
10 kHz Sawtooth Wave	$\tau_{max} = 8$ msec	2
Impulse	$\tau_{max} = 4$ msec	8.5
Impulse	$\tau_{max} = 8$ msec	6.5
Jazz-Vocal	$\tau_{max} = 4$ msec	4.5
Jazz-Vocal	$\tau_{max} = 8$ msec	3.5
Percussion Instruments	$\tau_{max} = 4$ msec	4
Percussion Instruments	$\tau_{max} = 8$ msec	2.5

Table 5.2. Average correct responses for 15 test subjects performing the loudspeaker test.

5.2 Discussion

Table 5.3 is the minimum number of correct responses *r* for concluding that performance is better than chance, and the resulting type 1 and type 2 error probabilities for *p* values from 0.6 to 0.8 in listening tests for a given *N* of 15.

N	*r*	Type 1 Error (α) actual value	Type 2 Error (β)			
			p = 0.6	p = 0.7	p = 0.75	p = 0.8
15	14	0.0005	0.9948	0.9647	0.9198	0.8329
	13	0.0037	0.9729	0.8732	0.7639	0.6020
	12	0.0176	0.9095	0.7031	0.5387	0.3518
	11	0.0592	0.7827	0.4845	0.3135	0.1642
	10	0.1509	0.5968	0.2784	0.1484	0.0611
	9	0.3036	0.3902	0.1311	0.0566	0.0181
	8	0.5000	0.2131	0.0500	0.0173	0.0042
	7	0.6964	0.0950	0.0152	0.0042	0.0008
	6	0.8491	0.0338	0.0037	0.0008	0.0001

Table 5.3. Minimum number of correct responses *r* for concluding that performance is better than chance, and resulting type 1 and type 2 error probabilities for *p* values from 0.6 to 0.8 in listening tests with *N* at 15. [L. Leventhal, "Type 1 and Type 2 Errors in Statistical Analysis of Listening Tests," *J. Audio Eng. Soc.,* vol. 34, pp. 437-453 (1986 Jun.), pp. 445, Table 3.]

A few points can be illustrated in Table 5.3, regarding the analysis of listening tests with statistical tests of significance. In listening tests, which are brief enough to be practical to conduct and avoid fatigue or boredom (i.e., $N \leq 20$), significance tests conducted at the 0.05 level of significance result, for most values of *p*, in a type 2 error risk which is much larger than actual type 1 error risk. Another point is that type 2 error and power change as a function of *p*, although one never really knows *p*.

If from Table 5.3, if $r = 9$ is selected, for $p = 0.6$ the fairness coefficient FC_p can be calculated as

$$FC_{0.6} = \frac{0.3036}{0.3902} = 0.7781. \quad (5.2)$$

This is a very desirable result as the ideal value is 1 (equal error). The actual type 1 error for the above situation is 0.3036 and type two error is 0.3902 which are both fairly similar. Although a p of 0.6 may seem as a low criterion, it was chosen so that the subtle effects of the audibility of phase distortion were uncovered in the analysis. Therefore, for this study, anything above 9 correct responses (r) out of 15 will be considered statistically significant for $p = 0.6$.

For data analysis of the randomized test signal sequences of AA, AB, BA, and BB presented in the listening test, the AA and BB data were discarded for analysis. Test subject responses to the test signal sequences AA and BB give no information regarding the detection of phase distortion and were only included to have an even random discrimination of the test signal sequences. The results of presentation sequence AB and BA were grouped together for calculation of the average correct listener responses. Thus, order effects were not considered in the analysis.

In a broad sense, the average correct responses for the loudspeaker-based listening test were significantly lower than for the headphone-based test. This was shown by a two-sample t-test assuming equal variances for $\alpha = 0.05$ stating differences between the two types of listening test responses exist. The results of the loudspeaker-based test also seemed to be independent of distortion level ($\tau_{max} = 4$ or 8 msec), even for test signals that showed statistically significant phase distortion audibility with headphones such as the 70 Hz sawtooth wave and impulse. The audibility effect of any Q effects in the all-pass filter, which were not investigated in the research, may

explain why for some signals, the correct responses for the audibility of the phase distortion were higher for the 4 msec delay than for the 8 msec delay.

The audibility of phase distortion for steady-state signals, such as the sawtooth wave was dependent upon frequency (of the sawtooth wave). The impulsive test signals (impulse) displayed phase distortion audibility for a mid-range all-pass filter center frequency. Simple test signals, such as the sawtooth wave, seemed to be more conducive in revealing the presence of phase distortion. In contrast, complex test signals such as the jazz vocal proved to be more difficult.

The audibility of phase distortion in audio signals was also highly dependent upon individual ability, although for statistical analysis individual data was not considered. For example, while most test subjects were very good at recognizing what was in general perceivable as phase distortion such as the impulse and the 70 Hz sawtooth wave, a few others had greater difficulty. Specifically, a few subjects seemed to hear clearly the presence of phase distortion in the jazz-vocal test signal for the headphone listening test, while a few test subjects seemed to perceive phase distortion better than others during the loudspeaker listening test.

Table 5.1 indicates that even for the headphone listening test, phase distortion audibility was of very subtle nature. This is surprising, since there exists gross phase distortion present in the all-pass filtered test signals. In this test, human ears seem to be tolerant of even large phase distortions in audio signals. For the $r = 9$ and $p = 0.6$ for the $N = 15$ criterion described for the statistical analysis earlier, it can be seen that 70 Hz sawtooth wave and the impulse test signals for both 4 and 8 msec maximum group delay times (τ_{max}) were significant. Given the results, the criterion in selecting the all-pass filter center frequency $f_0 = 3.5$ kHz seemed to be valid for the impulse test

signal. Phase distortion detection became progressively difficult for the 3.5 and 10 kHz sawtooth waves. Although it was proved that relative phase has subtle effects on timbre and there exists phase-locking of the auditory fibers of the ear below 5 kHz, the introduction of phase distortion for the 3.5 kHz sawtooth waves did not have a statistically significant result. The assumption that equal-loudness contours provide valid areas of maximal phase distortion sensitivity did not seem to hold true for the sawtooth wave. Other mechanisms in the human auditory system or test equipment may also be responsible for this. Based on the fact that phase-locking of the auditory fibers is lost above 5 kHz, it was hypothesized that the 10 kHz sawtooth wave had negligible results in the audibility of phase distortion. Results of the headphone listening test confirm this as they are all far below chance occurrence ($r < 7.5$). All other test subject's responses were below chance occurrence. Shifting the peak group delay of the fundamental of a spectrally rich signal such as a sawtooth wave in relation to its upper harmonics shows greater audibility for lower frequencies.

Table 5.2 indicates for the loudspeaker listening test that the overall phase distortion audibility was increasingly difficult as compared to the headphone listening test. None of the test subjects' responses qualified for the $r = 9$ and $p = 0.6$ for $N = 15$ statistical analysis criterion used, although the impulse test signal at 4 msec maximum group delay came very close at 8.5 average responses. When reverberant listening conditions are present, the audibility of phase distortion seems to be highly masked by its presence. Irradiation methods obviously do play a part in the audibility of phase distortion. Although the headphone listening process can be considered as a sole entity, room effects also contribute to the loudspeaker listening process and thus should be considered in tandem. Strategic acoustic room treatments and excess-phase

equalization could minimize audible room effects such as reverberation and excess-phase response [26], respectively. These corrections should in turn increase the sensitivity of the audibility of phase distortion of loudspeakers to experimental data comparable to that of the headphone data.

It is of interest to note that on music, even with transient content such as the percussion instruments, did not reveal phase distortion present in the signal. This could be due to in part to the existence of reverberation present in the original recording. Jazz vocals, with its rich spatial content, may have obscured the presence of any phase distortion. Although phase-locking of the auditory nerve fibers is present even at low intensities, the lack of phase distortion audibility in music test signals implies another masking mechanism is perhaps present. The audibility of phase distortion in music test signals is further exacerbated when played back on loudspeakers than on headphones, due to the non-anechoic listening environment. Clearly, for non-anechoic listening environments, there are certain issues to be resolved before consideration of the audibility of phase distortion comes in to play.

Fig. 5.3 shows the permissible phase distortion level for the various test signals used in the headphone listening test. This graph has its limitations as it is a construct based on the constraints of the experimental design (i.e., type of test signals used and limited levels of distortion introduced in the test signals) and is by no means a complete representation of the nature of the audibility of phase distortion. Fig. 5.4 shows the permissible phase distortion level for the various test signals used in the loudspeaker listening test. What is implied by Fig. 5.3 and 5.4 is the minimum allowable level of phase distortion before audibility and therefore actual levels are very likely to be higher.

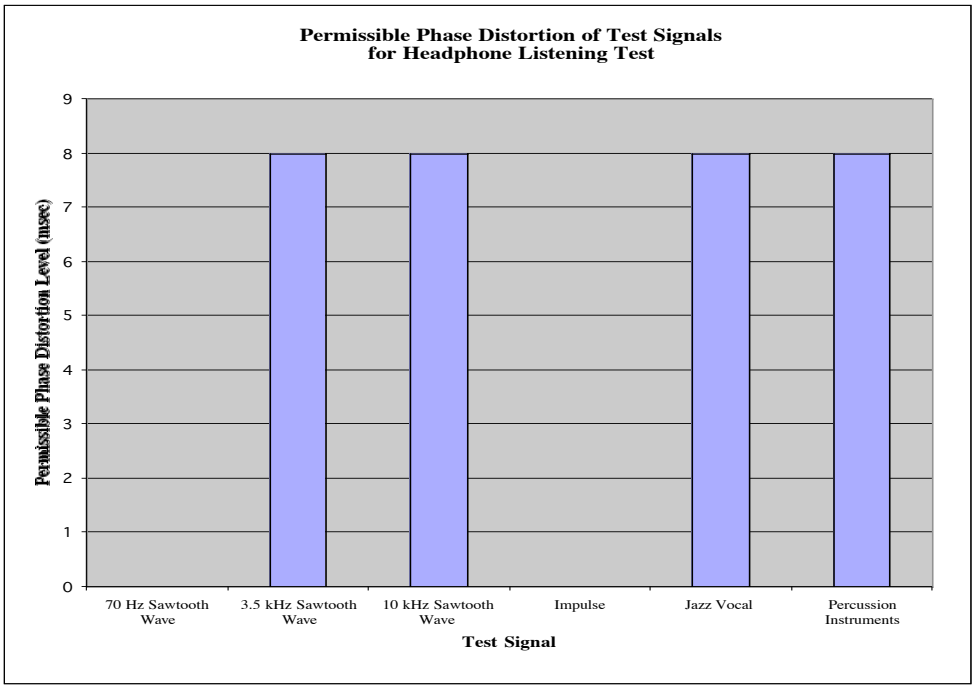


Fig. 5.3. Permissible phase distortion in milliseconds of test signals used for the headphone listening test. This graph has its limitation, as it is a construct based on constraints of the experimental design. Shown is the minimum allowable level of phase distortion before audibility and actual levels are very likely to be higher.

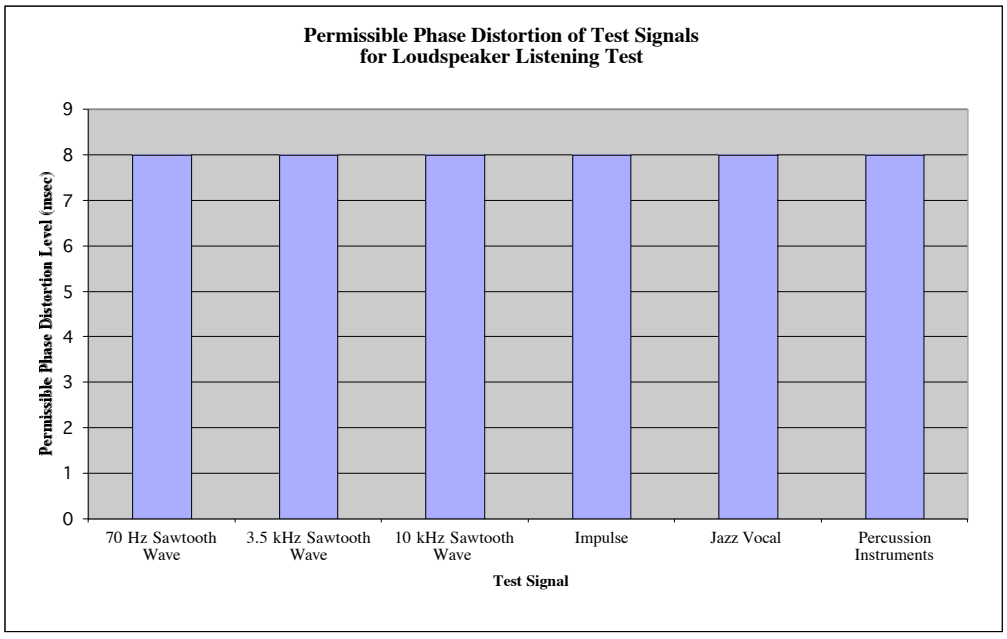


Fig. 5.4. Permissible phase distortion in milliseconds of test signals used for the loudspeaker listening test. This graph has its limitation, as it is a construct based on constraints of the experimental design. Shown is the minimum allowable level of phase distortion before audibility and actual levels are very likely to be higher.

Conducting research of broad test-signal scope and experimental design practicality resulted in limited results. Obviously, more refined research utilizing a greater signal variation base (i.e., more frequencies for sawtooth waves across frequency spectra) and phase distortion levels are necessary to ascertain more accurate permissible levels of phase distortion for these signals.

Another way to show the content displayed in Fig. 5.3. and 5.4 is shown in Table 5.4., which contains the types of test signals used for both headphone and loudspeaker tests, center frequencies of the all-pass filter, and the minimum permissible phase distortion levels.

(a)

Test Signal	Center Frequency (f_0) of All-Pass Filter	Permissible phase distortion level (msec)
70 Hz sawtooth wave	70 Hz	≥ 0
3.5 kHz sawtooth wave	3.5 kHz	≥ 8
10 kHz sawtooth wave	10 kHz	≥ 8
Impulse	3.5 kHz	≥ 0
Jazz-vocal group	160 Hz	≥ 8
Percussion instruments	150 Hz	≥ 8

(b)

Test Signal	Center Frequency (f_0) of All-Pass Filter	Permissible phase distortion level (msec)
70 Hz sawtooth wave	70 Hz	≥ 8
3.5 kHz sawtooth wave	3.5 kHz	≥ 8
10 kHz sawtooth wave	10 kHz	≥ 8
Impulse	3.5 kHz	≥ 8
Jazz-vocal group	160 Hz	≥ 8
Percussion instruments	150 Hz	≥ 8

Table 5.4 (a) Permissible phase distortion levels (msec) for headphone listening for test signals and center frequencies of all-pass filters used in research. (b) Permissible phase distortion levels (msec) for loudspeaker listening for test signals and center frequencies of all-pass filters used in research.

Table 5.4 may be used to identify the permissible level of phase distortion as a function of test signal type and frequency. It can be seen that phase distortion audibility depends on both type of test signal and phase distortion level incurred in the all-pass filter, the former being dominant.

Considerations regarding this research include the fact that phase distortion inherent in the transducers used for the listening tests such as the microphones, headphones, and loudspeakers, were not investigated. The amplifiers used by the headphones and loudspeakers were assumed to be phase linear. Test subjects were assumed to have normal hearing, however were not tested.

Fig. 5.3 and 5.4. indicate regarding design of acoustic transducers and loudspeaker systems, that only for very critical listening conditions is the correction of phase

distortion a requirement and only after extraneous sources of masking (i.e., reverberant listening conditions) are addressed properly. For example, the peak group delay incurred in the transition of a fourth-order Butterworth low to high-pass crossover is 0.2 msec. For accurate perception of audio signals, other primary design requirements exist for acoustic transducer and loudspeaker system design that should take priority before secondary measures such as phase correction are addressed.

Another implication is the phase distortion incurred from studio equalizers. These equalizers are usually of analog type, and essentially behave as band-pass/band-reject filters. Minimum-phase phase distortion is introduced by usage of such a device. Fig. 5.5 shows the group-delay characteristic for a first-order band-pass filter (f_0 = 3.5 kHz) with Q = 1, 10, 50, and 100. These graphs would be analogous to altering the Q of a parametric equalizer. Inspection of Fig. 5.5 (b) reveals that for Q = 50, there was approximately 4 msec of group-delay peaking and for Q = 100, approximately 8 msec. The permissible levels established in this research imply that phase distortion incurred by analog equalization may not be of concern up to a certain permissible level (i.e., Q = 100 in a first-order band-pass filter), especially for most mid-range musical content.

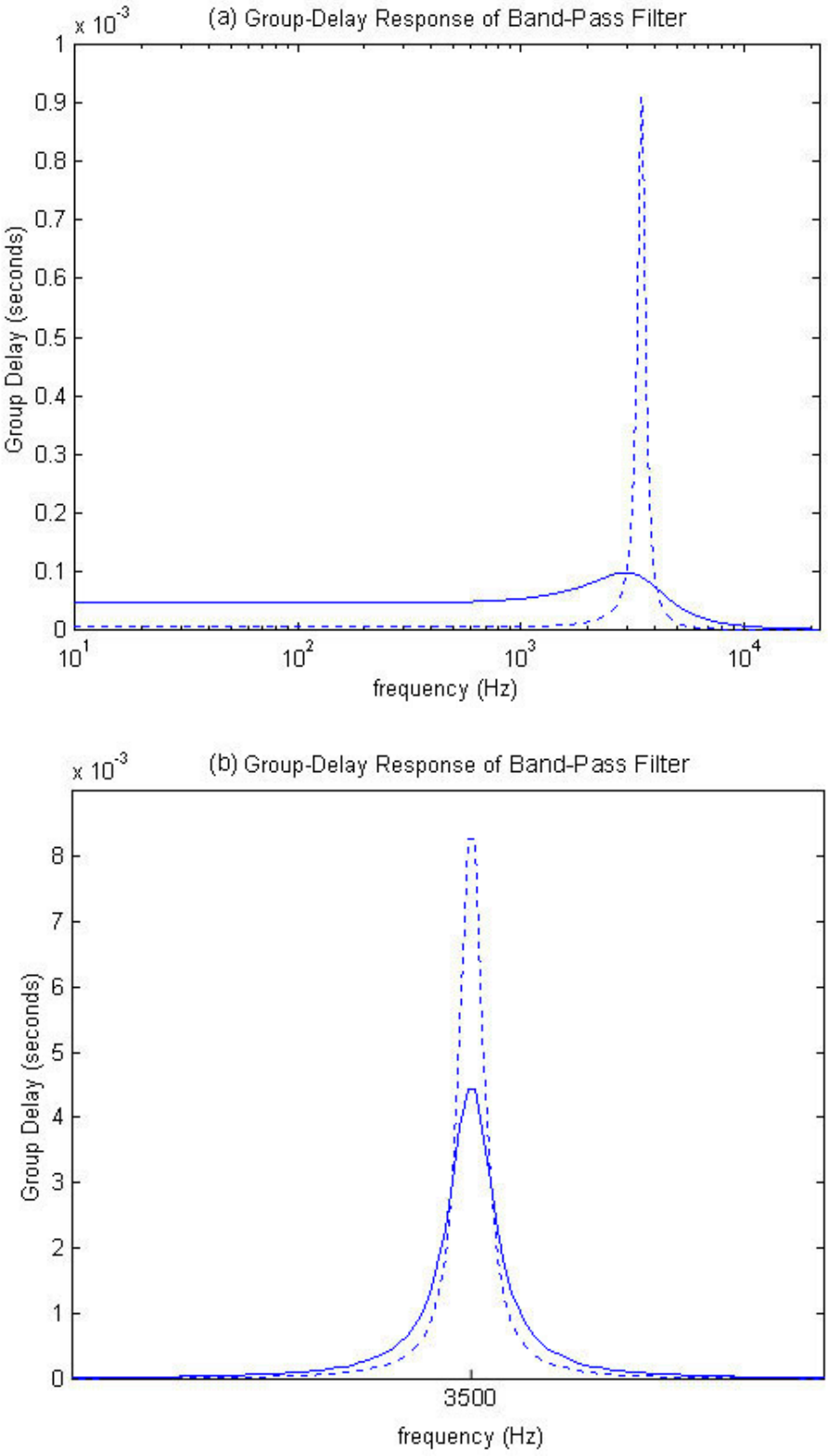


Fig. 5.5 (a). Group delay response for first-order band-pass filter (f_0 = 3.5 kHz) with Q = 1 (solid line) and Q = 10 (dotted line). (b). Group delay response for first-order band-pass filter (f_0 = 3.5 kHz) with Q = 50 (solid line) and Q = 100 (dotted line).

This chapter presented the results and provided discussion. Average correct responses of all test subjects for both the headphone and loudspeaker listening tests were presented. The statistical analysis method of equalizing type 1 and 2 error, which is important for useful interpretation of the acquired data, was presented and implemented. Finally, the results were discussed and implications were brought forward. The results from this thesis research imply that phase distortion is of secondary importance as compared to frequency response irregularities, which is in agreement with previous research results. The next chapter will conclude this thesis research by providing a summary

6. Conclusion

A psychoacoustic experiment and analysis, proposed to ascertain permissible levels of phase distortion in various audio signals, addressed the significance of the audibility of phase distortion. A valid experimental design was justified by the application of the *Kwalwasser-Dykema Music Tests* format. The use of a computer enhanced the accuracy in implementing the digital all-pass filter. The use of studio-quality audio equipment also aided in enhancing the accuracy of the experimental results. The statistical analysis method used in this thesis research of equalizing type 1 and 2 error provided for a fair study. Finally, permissible levels of phase distortion audibility within the confines of the experimental design were established.

These permissible levels may be beneficial in the design and application of audio equipment. It is concluded that phase distortion is an extremely subtle effect, complicated by reverberant listening or original recording conditions. The audibility of phase distortion in audio signals seems to be highly dependent upon individual ability. It is also suggested that since audible only in extreme situations, the audibility of phase distortion may be of concern only when primary factors (i.e., reverberant listening environments) are accounted for.

Although in agreement for the most part with previous research, the phase distortion audibility results in this thesis research did not seem to be as significant as demonstrated by Lipshitz *et al.* [7]. Furthermore, this test does not support the results of Hansen and Madsen [14] since their research revealed increased phase sensitivity with a loudspeaker in reverberant environments as compared to headphone listening tests.

The selection of the all-pass center frequency of 3.5 kHz where the Robinson-Dadson curve displays the lowest threshold of audibility may suggest that for broad-band signals (such as the impulse), mid-range phase distortion is most audible. Since phase-locking of the auditory fibers is lost above 5 kHz, this may be a contributing factor for the 10 kHz sawtooth wave having negligible results in phase distortion audibility. Although phase-locking of the auditory nerve fibers is present even at low intensities, the lack of phase distortion audibility in music test signals implies this theory is suspect to explain phase distortion audibility for complex audio signals and/or another masking mechanism may be present.

Considerations regarding this research include the fact that phase distortion inherent in the transducers used for the listening tests such as the microphones, headphones, and loudspeakers, were not investigated. The amplifiers used by the headphones and loudspeakers were assumed to be phase linear. Test subjects were assumed to have normal hearing, but were not tested.

The human auditory system was found to be extremely tolerant of even gross phase distortion effects. Although the impulse test signals were very revealing of the presence of phase distortion, more refined research utilizing an improved selection, broader frequency range, and various all-pass filtering implementations (wider range of phase distortion levels and Q) of test signals are necessary to ascertain more accurate permissible levels of audible phase distortion. Improved irradiation methods, such as the use of phase-equalized loudspeakers in an anechoic environment, may also aid in ascertaining more accurate permissible levels.

References

[1] D. Preis, “Phase Distortion and Phase Equalization in Audio Signal Processing – A Tutorial Review,” *J. Audio Eng. Soc.,* vol. 30, pp. 774-794 (1982 Nov.).

[2] R. E. Radocy and J. D. Boyle, *Psychological foundations of musical behavior*, (Charles C Thomas Publisher Ltd., Illinois, 1997).

[3] J. O. Pickles, *An Introduction to the Physiology of Hearing* (Academic Press, New York, 1982).

[4] C. A. Elliot, “Attacks and releases as factors in instrument identification,” *J. Res. in Music Education*, vol. 23, pp. 35-40 (1975).

[5] J. G. Roederer, *The Physics and Psychophysics of Music : An Introduction* (Springer-Verlag, New York, 1995).

[6] D. O. Kim, C. E. Molnar, and J. W. Matthews, “Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure,” *J. Acoust. Soc. Am.*, vol. 67, pp. 1704-1721 (1980).

[7] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, “On the Audibility of Midrange Phase Distortion in Audio Systems,” *J. Audio Eng. Soc.,* vol. 30, pp. 580-595 (1982 Sep.).

[8] R. C. Mathes and R. L. Miller, “Phase Effects in Monaural Perception,” *J. Acoust. Soc. Am.*, vol. 19, pp. 780-797 (1947).

[9] J. H. Craig and L. A. Jeffress, “The Effect of Phase on the Quality of a Two-Component Tone,” *J. Audio Eng. Soc.,* vol. 34, pp. 1752-1760 (1962).

[10] 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 Sep.).

[11] J. R. Ashley, “Group and Phase Delay Requirements for Loudspeaker Systems,” *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Denver, CO, 1980 Apr. 9-11), vol. 3, pp. 1030-1033 (1980).

[12] S. G. Bridges, “Effect of Direct Sound on the Perceived Frequency Response of a Sound System,” presented at the 66th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 28, p. 552 (1980 July/Aug.), preprint 1644.

[13] V. Hansen and E. R. Madsen, “On Aural Phase Detection,” *J. Audio Eng. Soc.,* vol. 22, pp. 10-14 (1974 Jan./Feb.).

[14] V. Hansen and E. R. Madsen, “On Aural Phase Detection: Part II,” *J. Audio Eng. Soc.,* vol. 22, pp. 783-788 (1974 Dec.).

[15] H. Suzuki, S. Morita, and T. Shindo, "On the Perception of Phase Distortion," *J. Audio Eng. Soc.,* vol. 28, pp. 570-574 (1980 Sep.).

[16] L. R. Fincham, "The Subjective Importance of Uniform Group Delay at Low Frequencies," *J. Audio Eng. Soc.*, vol. 33, pp. 436-439 (1985 Jun.).

[17] D. Preis and P. J. Bloom, "Perception of Phase Distortion in Anti-Alias Filters," *J. Audio Eng. Soc.*, vol. 32, pp. 842-848 (1984 Nov.).

[18] J. A. Deer, P. J. Bloom, and D. Preis, "Perception of Phase Distortion in All-Pass Filters," *J. Audio Eng. Soc.,* vol. 33, pp. 782-786 (1985 Oct.).

[19] D. Preis, F. Hlawatsch, P. J. Bloom, and J. A. Deer, "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *J. Audio Eng. Soc.*, vol. 35, pp. 1004-1012 (1987 Dec.).

[20] C. L. Lindquist, *Active Network Design*, (Steward & Sons, California, 1977).

[21] K. C. Pohlmann, *Principles of Digital Audio*, (McGraw-Hill, New York, 1995).

[22] R. G. Greenfield and M. O. J. Hawksford, "The Audibility of Loudspeaker Phase Distortion," presented at the 88th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* (1990 Mar.), preprint 2927.

[23] D. Sergeant and J. D. Boyle, "Contextual Influences on Pitch Judgement," *J. Soc. for Res. in Psych. of Music and Music Ed.*, vol. 8, pp. 3 - 15 (1980).

[24] L. Leventhal, "Type 1 and Type 2 Errors in Statistical Analysis of Listening Tests," *J. Audio Eng. Soc.,* vol. 34, pp. 437-453 (1986 Jun.).

[25] S. Siegel, *Nonparametric Statistics.* (McGraw-Hill, New York, 1956).

[26] L. G. Johansen and P. Rubak, "The Excess Phase in Loudspeaker/Room Transfer Functions: Can it be Ignored in Equalization Tasks?," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* (1996 May.), preprint 4181.

Appendix A - MATLAB code for test tone synthesis, selection, and all-pass filtering

```
%-------------------------------------------------
% MATLAB code for test tone synthesis and filtering
% by Daisuke Koya
%-------------------------------------------------

clear all;
h = get(0,'Children');  % get handle of current plots
set(h,'color',[1 1 1]); % set background of plot to white

% test tone generation
fs = 44100;       % sampling frequency of test tone in Hz
time = (0:1/fs:1);% 1 sec. time duration of sine wave, T = 1/44100Hz

%--------
% impulse
%--------

imp = zeros(44100,1);
imp(1) = 1;
% f0 = 3500 is used for the AP filter f0
% (assumed to be most sensitive by the Fletcher-Munson curves)

%-------------------
% 70Hz sawtooth wave
%-------------------

%st70 = sawtooth(2*pi*70*time); % 70Hz sawtooth wave
%st70 = 0.35.*st70;
% reduce amplitude so that filtered signal
% is not clipped beyond -1, +1

%---------------------
% 3.5kHz sawtooth wave
%---------------------

%st3500 = sawtooth(2*pi*3500*time); % 3.5kHz sawtooth wave
%st3500 = 0.6.*st3500;
% reduce amplitude so that filtered signal
% is not clipped beyond -1, +1

%--------------------
% 10kHz sawtooth wave
%--------------------

%st10000 = sawtooth(2*pi*10000*time); % 10kHz sawtooth wave
%st10000 = 0.6.*st10000;
% reduce amplitude so that filtered signal
% is not clipped beyond -1, +1

% note: need to use f0=12100Hz for AP filter formulation
% since approximation has larger frequency errors at higher
frequencies

% note: tmax approximation for 10kHz sawtooth waveform
% fails, therefore,
% for tmax = 0.001 mS, use tmax = 0.000285
% for tmax = 0.002 mS, use tmax = 0.000575
```

```
% for tmax = 0.004 mS, use tmax = 0.00115
% for tmax = 0.008 mS, use tmax = 0.00235

%-----------------------------------------------------
% University of Miami Percussion Ensemble
% use f0 ~ 150Hz for APF formulation since spectogram
% has significant frequency content in this area
%-----------------------------------------------------

%perc = wavread('perc.wav');
%perc = 0.8.*perc;
% reduce amplitude so that filtered signal
% is not clipped beyond -1, +1

%-----------------------------------------------------
% M-Pact (Jazz vocal group)
% use f0 ~ 160Hz for APF formulation since spectogram
% has significant frequency content in this area
%-----------------------------------------------------

%mpact = wavread('mpact.wav');

%--------------------
% test tone selection
%--------------------

x = imp;    % select signal from above to be used

% !! check f0 of APF to make sure it corresponds to above signal !!

%-------------------------------------------------------------
% Sampling frequency for 2nd-order allpass filter formulation
%-------------------------------------------------------------

fs = 44100;       %     sampling frequency in Hz
T = 1/44100;      %     sampling period in seconds

%--------------------------
% Allpass filter parameters
%--------------------------

tmax = 0.004                % peak delay in mS of AP
f0 = 3500;                  % resonant frequency in Hz of AP

%--------------------------
% Allpass filter formulation
%--------------------------

Q = 0.5*tmax*pi*f0          % Q of AP transfer function
w0 = 2*pi*f0;               % resonant frequency in radians
sn = 1/w0;                  % normalized frequency sn
numAP = [sn^2 -sn/Q 1];     % numerator of allpass filter
denAP = [sn^2 +sn/Q 1];     % denominator of allpass filter

f = (0:20:22050);           % frequency vector, to 1/2 Nyquist
w = 2*pi*f;                 % frequency vector in radians
H = freqs(numAP,denAP,w);   % complex frequency response

magH = 20*log10(abs(H));% magnitude of H
phaserad = angle(H);    % phase of complex frequency response (rad.)
```

```
phasedeg = phaserad*180/pi;                % convert to degrees
gd = -diff(unwrap(phaserad))./diff(w);     % calculate group delay

%---------------------------------
% Analog to Digital Transformation
%---------------------------------

[numAPd,denAPd] = BILINEAR(numAP,denAP,fs);%Bilinear transformation
[Gd,F] = GRPDELAY(numAPd,denAPd,f,fs);
Gdsec = Gd*T;
[H,F] = freqz(numAPd,denAPd,f,fs);

%---------------------------------
% Filter signal with allpass filter
%---------------------------------

y = filter(numAPd, denAPd, x);

%------------
% Plot result
%------------

figure(1)
semilogx(f,Gdsec)
axis([0 22050 0 0.01])
grid
title('Digital Group Delay Response of AllPass Filter');
xlabel('frequency (Hz)'),ylabel('Group Delay (seconds)')

tsignal = ( 0:(length(x)-1) ).*T ;  % convert # samples to time in
seconds

figure(2)
plot(tsignal,x,'-')     % snapshot of steady-state response
axis([0.000 0.05 -1 1]);
hold on
plot(tsignal,y,':')
hold off
title('Unfiltered (solid) and filtered (dotted) signals');
xlabel('time (seconds)'),ylabel('amplitude')
```

Appendix B - Scoresheet for Headphone Listening Test

Scoresheet for *headphone* test

Name _____

1. SOUND ONE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

2. SOUND TWO

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

3. SOUND THREE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

4. SOUND FOUR

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

5. SOUND FIVE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

6. SOUND SIX

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

Appendix C - Scoresheet for Loudspeaker Listening Test

Scoresheet for *loudspeaker* test

Name _____

1. SOUND ONE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

2. SOUND TWO

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

3. SOUND THREE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

4. SOUND FOUR

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

5. SOUND FIVE

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

6. SOUND SIX

VERSION ONE	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

VERSION TWO	Different? Yes	No
(a)		
(b)		
(c)		
(d)		

Appendix D - D. Listening Test Results for Headphone Listening Test

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Average
st70_04	1	0	0	1	1	0	0	1	1	1	1	1	0	1	0	9
st70_40	1	0	1	1	0	0	1	0	0	1	1	1	1	0	1	
st70_08	1	0	0	1	1	0	0	1	1	1	1	1	0	0	1	9.5
st70_80	1	1	1	1	0	0	1	1	0	0	1	1	1	1	0	
st3500_04	1	0	1	0	0	1	0	0	0	0	1	0	1	0	1	5.5
st3500_40	0	0	1	0	0	1	0	1	1	0	1	0	0	0	0	
st3500_08	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	4
st3500_80	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	
st10000_04	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1.5
st10000_40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
st10000_08	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	4.5
st10000_80	0	0	1	0	0	0	1	0	1	0	0	0	1	1	0	
imp_04	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	13
imp_40	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
imp_08	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	14
imp_80	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
mpact_04	1	1	1	0	0	1	0	1	1	0	0	0	1	0	0	6.5
mpact_40	1	1	1	0	0	0	1	1	0	0	0	0	0	0	1	
mpact_08	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	4
mpact_80	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	
perc_04	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	4
perc_40	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	
perc_08	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	4
perc_80	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	

Appendix E - Listening Test Results for Loudspeaker Listening Test

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Average
st70_04	0	0	1	0	1	0	1	0	1	0	1	1	1	0	0	7
st70_40	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	
st70_08	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	7
st70_80	1	1	1	0	1	0	0	1	1	0	1	0	0	0	0	
st3500_04	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	5
st3500_40	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	
st3500_08	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	3.5
st3500_80	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	
st10000_04	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	2
st10000_40	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	
st10000_08	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2
st10000_80	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	
imp_04	1	0	1	1	0	1	0	0	1	0	0	1	0	1	1	8.5
imp_40	1	1	0	1	0	1	1	0	1	1	1	1	0	0	0	
imp_08	1	1	0	0	0	0	1	1	0	0	0	0	0	1	0	6.5
imp_80	1	0	1	1	0	1	0	1	1	1	0	0	1	0	0	
mpact_04	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	4.5
mpact_40	0	0	1	0	0	0	0	0	1	0	0	0	1	0	1	
mpact_08	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	3.5
mpact_80	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	
perc_04	1	1	1	0	0	0	1	0	1	0	0	0	1	0	0	4
perc_40	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
perc_08	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.5
perc_80	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	