LISTENING IN LARGE ROOMS: A NEUROPHYSIOLOGICAL INVESTIGATION OF ACOUSTICAL CONDITIONS THAT INFLUENCE SPEECH INTELLIGIBILITY

by

Benjamin Michael Hammond

B.A., Chemistry (1989)

Harvard College

Submitted to the Division of Health Sciences and Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Health Sciences and Technology

at the

Massachusetts Institute of Technology

Feb, 1997

Signature of Author ....Posthumous Degree Award..
Division of Health Sciences and Technology
May 13, 1996

Certified by ....
Bertrand Delgutte
Associate Professor of Harvard Medical School
Thesis Supervisor

Accepted by ....
Martha L. Gray
Committee on Graduate Students
J. W. Kieckhefer Associate Professor of Electrical Engineering and Computer Science
Co-Director, Divsion of Health Sciences and Technology

HARVARD UNIVERSITY — MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DIVISION OF HEALTH SCIENCES AND TECHNOLOGY

Speech and Hearing Sciences Program

The following document was submitted
as a Doctoral Thesis Proposal by
Benjamin Michael Hammond
who passed away during the Fall of 1996.
It was accepted by the Division of Health
Science and Technology as a Masters
Thesis

LISTENING IN LARGE ROOMS: A NEUROPHYSIOLOGICAL INVESTIGATION OF ACOUSTICAL CONDITIONS THAT INFLUENCE SPEECH INTELLIGIBILITY

by

Benjamin M. Hammond

A thesis proposal for the degree of

Ph.D.

Harvard University – Massachusetts Institute of Technology
Division of Health Sciences and Technology
Speech and Hearing Sciences Program

May 24, 1996

TABLE OF CONTENTS

Thesis Proposal

Benjamin M. Hammond
May 23, 1996

Introduction

In 1895, Harvard's President Eliot asked a newly-appointed assistant professor of physics if he might be able to "do something" about the acoustical difficulties in the lecture room of Harvard's new Fogg Art Museum. Not only did the young professor solve the problem, but as a result of an exhaustive experimental effort far exceeding the scope of his assigned chore, Wallace Clement Sabine also discovered the fundamental relationship between the decay of sound energy in a room and the room's volume and sound absorbing properties. This relationship is now known as Sabine's reverberation equation. Using this discovery, Sabine went on to define the field of modern architectural acoustics in a career that is a role model for the practice of applied science. University presidents should never underestimate the significance of their simple requests.

Another prominent Harvard physicist and acoustician, Frederick V. Hunt, claims that "the Roman architect, Vitruvius, ... was the only one among either the ancients or the pre-Sabinites who came close to matching the clarity of Sabine's statement of the conditions for good hearing." (Hunt, 1964) Writing in the first century BC, Vitruvius described the listening experiences at different seating positions within a theater (Vitruvius, Liber V, Cap. VIII):

> *... we must see with even greater care that a position has been taken where the voice falls softly and is not so reflected as to produce a confused effect on the ear. There are some positions offering natural obstructions to the projection of the voice, as for instance the dissonant, which in Greek are termed κατηχουντεζ; the circumsonant, with them are named περιηχουντεζ; and again the resonant, which are termed αντηχουντεζ. The consonant positions are called by them συνηχουντεζ.*
>
> *The dissonant are those places in which the sound first uttered is carried up, strikes against solid bodies above, and, reflected, checks as it falls the rise of the succeeding sound.*
>
> *The circumsonant are those in which the voice spreading in all directions is reflected into the middle, where it dissolves, confusing the case endings, and dies away in sounds of indistinct meaning.*
>
> *The resonant are those in which the voice comes in contact with some solid substance and is reflected, producing an echo and making the case terminations double.*
>
> *The consonant are those in which the voice is supported and strengthened, and reaches the ear in words which are clear and distinct.*

Sabine identified Vitruvius' terms, *dissonance*, *circumsonance*, and *resonance*, with their modern equivalents, *interference*, *reverberation*, and *echo*. *Consonance*, he noted, has no single corresponding modern term, though it is the "one acoustical virtue that is positive" (Sabine, 1993 p. 187). Today we might call it *integration.*

This historical introduction serves several purposes. In the first place, Vitruvius' comments remind his modern reader of the venerable relationship between the spoken word and the spaces within which communication occurs. In fact, this relationship has existed for music as well, and its importance cannot be overstated. For example, Helmholtz believed that his theory of musical harmony could not account for the invention of the musical scale and its use in the Homophonic music of the early Christian eras. He thought this because "scales existed long before there was any knowledge or experience of harmony" and because "the individual parts of melody reach the ear in succession." Sabine resolved Helmhotz's dilemma by noting that, because of reverberation, "it is not necessarily true that (tones produced in succession) were heard as isolated notes." Harmonic context is provided by reflected sound. As a result, Sabine suggested, differences in the "physical environment" for making sound may account for the "racial and national differences" in musical scales (Sabine, 1993 pp. 113-114). The potentially significant role of room acoustics in the evolution of speech, language and music is a subject that awaits further study.

Vitruvius' quotation is also useful because it introduces the acoustical terms that are central to this thesis proposal. Specifically, this project is concerned with the neurophysiological mechanisms underlying the influence of echoes and reverberation on the intelligibility of speech. By *echo* we mean a reflected sound wave (or group of reflections) that can be detected by the listener and assigned a location. Thus when we speak of *echo threshold*, we will mean the delay time and intensity level of a reflection such that its presence and direction can be perceived. *Reverberation* is the diffuse, decaying sound field that consists of superposed and non-localizable reflections following an initial sound. Because single, isolated reflections occur only in experimental, anechoic conditions, and never in real rooms, we have chosen to concentrate on the effects of reverberation and of echoes occurring in a reverberant field. The term, "large rooms", in the proposal title points out that the acoustical conditions to be studied are those associated with lecture halls, theaters, concert halls, and churches, in which a single talker addresses an audience.

Sabine's turn-of-the-century discovery reveals how recently the study of room acoustics became an empirical science. Even more recently came the techniques and technology that allow the sound-evoked activity of single units in the mammalian auditory nervous system to be recorded (Kiang, Watanabe et al., 1965). This thesis will attempt to bring together the separate disciplines of architectural acoustics and auditory physiology. The mediator in this union will be subject of speech intelligibility.

I. Historical Background

1.1 Speech Intelligibility in Echoic Conditions

1.1.1 Sabine on Echoes

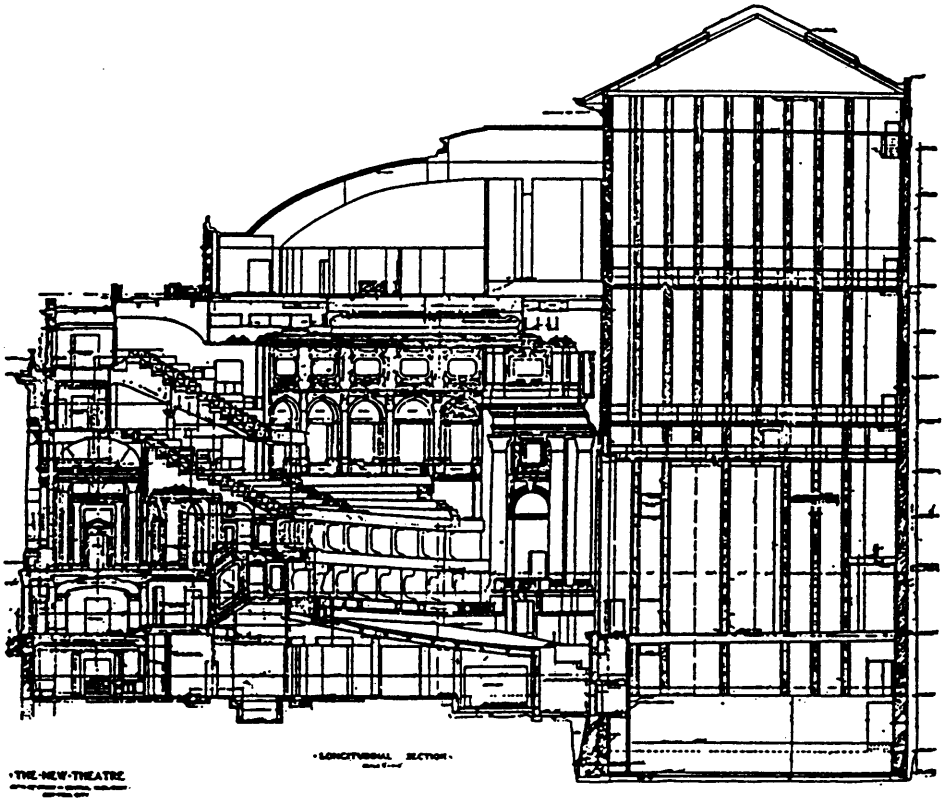

Figure 1. Section of the New Theatre, New York.
Carrere and Hastings, Architects. (Sabine, 1993 p. 179)

Vitruvius and Sabine were aware that certain late-arriving reflections could "double the case terminations" and "produce a confused effect on the ear". The benefits for speech intelligibility of avoiding such reflections, or echoes, guided Sabine's acoustical consulting practice. The New Theatre in New York is a case in point. The architect's sketch of the theater, which was intended for opera and spoken drama, is shown in Figure 1.

When the theater opened, it was met with criticism of its acoustical quality because speech could not be heard clearly. Sabine, using a scaled model of the theater to investigate the radiation of sound from the stage, found that the flat ceiling portrayed in the architect's sketch gave rise to a significant late reflection into the balconies and box seats. This reflection could be eliminated in the presence of a ceiling canopy. Figure 2 presents Sabine's results using the Toeppler-Boys-Foley method of photographing air disturbances. These photographs of the expanding sound wavefront and its reflections were taken by removing the sides of the model, placing a photographic plate across one opening, and illuminating the other opening "instantaneously by the light from a very fine and somewhat distant electric spark." The light of the spark was refracted by the expanding sound wavefronts, making them visible in the photographs. The left column shows the direct sound and its reflections at three times following the initial sound. The ceiling reflection has just reached the upper balcony by the third panel. In the right column, the canopy eliminates this reflection. Needless to say, Sabine's recommendation for a canopy was implemented, and the criticisms subsided (Sabine, 1993 pp. 177-187).

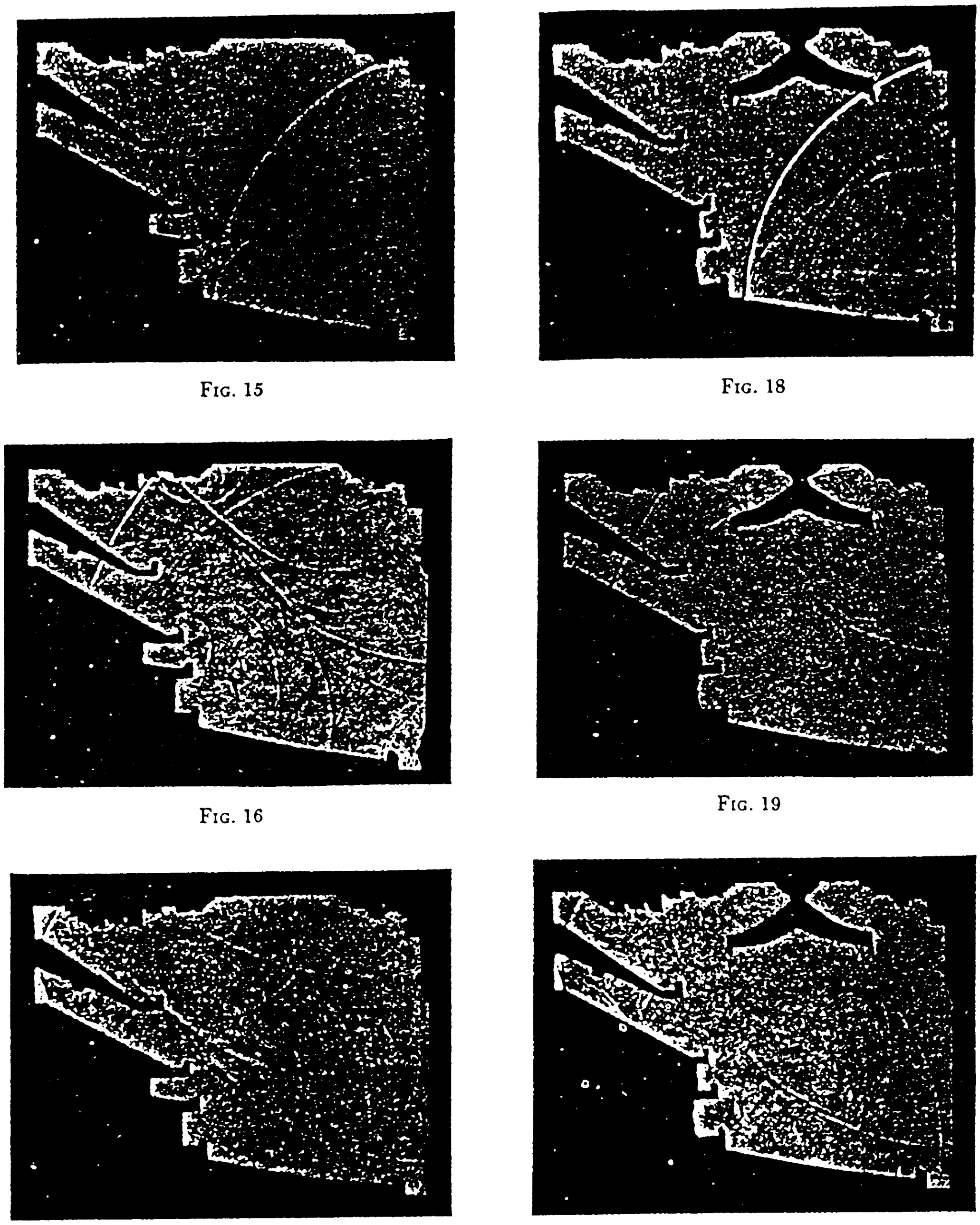

Fig. 15

Fig. 18

Fig. 16

Fig. 19

Fig. 17

Fig. 20

Figure 2. Two series of photographs of the sound and its reflections in the New Theatre, — 15 to 17 before, 18 to 20 after the installation of the canopy in the ceiling. The effect of the canopy in protecting the balcony, foyer chairs, boxes, and the orchestra chairs back of row L is shown by comparing Figs. 19 and 20 with Figs. 16 and 17. (Sabine, 1993 p. 181)

Ever since Sabine, it has been a rule-of-thumb in architectural acoustics to avoid late-arriving reflections that could be perceived as echoes.

1.1.2 The Precedence Effect

A question as relevant today as it was in the first part of this century is why late-arriving reflections are perceived as echoes and are detrimental to speech intelligibility, while early reflections tend to fall below the echo threshold and enhance the clarity of speech. The earliest reports of experimentation with speech and early reflections date back to the 1930's, when vacuum tube amplifiers and direct-radiating loudspeakers were first being used for amplified public address. Several reports from this time period describe a method for maintaining the "illusion that the sound comes from the speaker's mouth." (Fay, 1936; Hall, 1936; Snow, 1936) These reports found that if the arrival of the amplified sound was slightly delayed relative to the direct sound, the listener would perceive the sound as coming entirely from the talker. Because the first-arriving sound appeared to dominate the perception of localization, this "illusion" was called the "law of the first wavefront".

It was not until the middle of the century that the phenomenon known today as the precedence effect was studied systematically (Wallach, Newman et al., 1949; Haas, 1951; Meyer and Schodder, 1952; Lochner and Burger, 1958). Using the configuration shown in Figure 3, these early researchers could measure the effects of a single artificial reflection on speech perception.

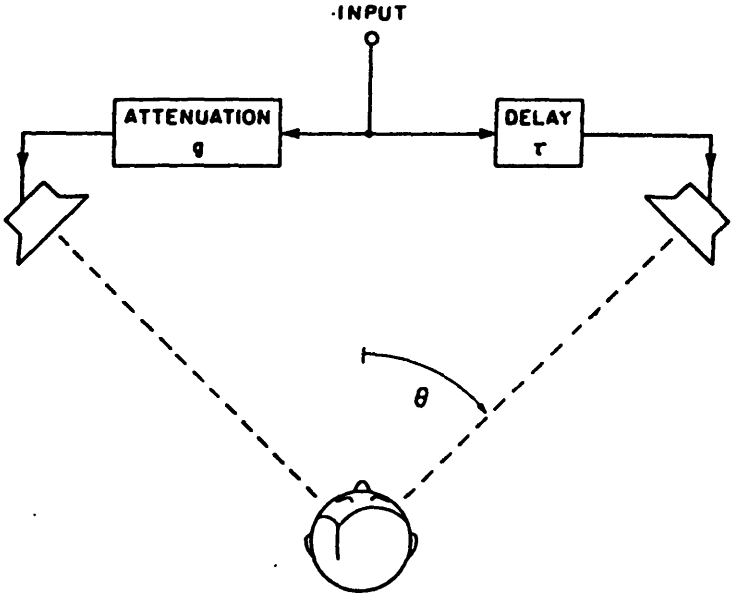

Figure 3. The two-loudspeaker configuration used in demonstrations and measurements of the precedence effect. The left loudspeaker provides the direct (or primary) sound. The reflection is simulated by the right loudspeaker. (Zurek, 1987)

Lochner and Burger determined the echo thresholds (i.e., the delay time and intensity level of a reflection such that its presence and direction can be perceived) for Harvard PB 50 articulation test lists. Their results for overall speech levels of 25 and 50 dB HL are shown in Figure 4. Early reflections (< 30 - 50 ms) must be up to 12 dB more intense than the direct sound in order to be perceived as an echo. With increasing delay time, reflections become easier to detect as echoes.

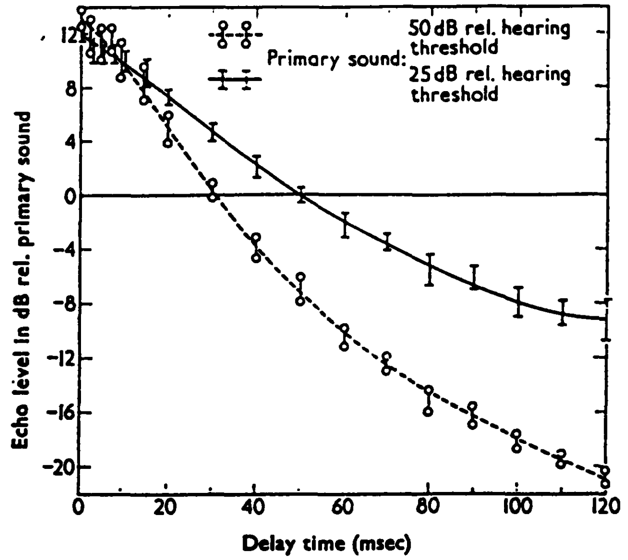

Figure 4. Curves showing the just perceptible level of a speech echo (i.e., echo threshold) in dB rel. the primary sound, at different delay times. (Lochner and Burger, 1964)

Haas showed for running speech that reflections occuring within 30 ms could be up to 10 dB more intense than the direct sound before listeners were disturbed by the reflection. His data on the percentage of listeners disturbed by a delayed speech signal are shown in Figure 5 as a function of delay time, with reflection level as the variable.

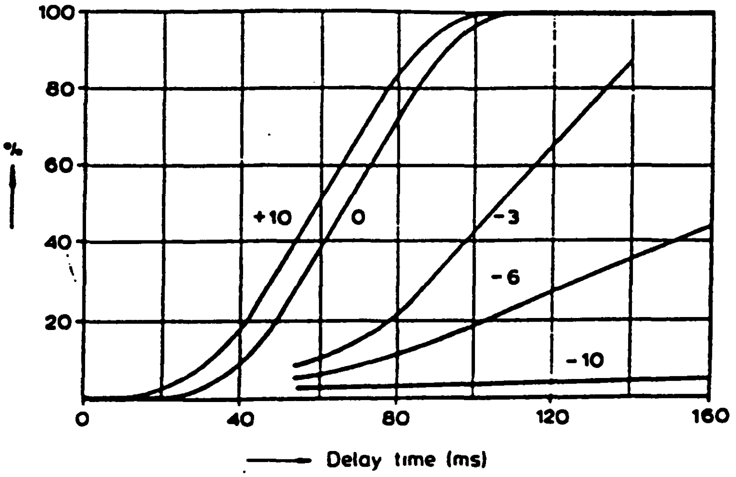

Figure 5. Percentage of listeners disturbed by a delayed speech signal. Speaking rate is 5.3 syllables per second. The relative echo levels (in dB) are indicated by numbers next to the curves. (from Kuttruff, 1991)

Considered with Lochner and Burger's results on echo thresholds, Haas' data suggest that reflections become disturbing when they can be perceived as echoes. Lochner and Burger performed additional experiments that test this hypothesis. They determined the intensity levels of speech that would produce any given articulation score on the Harvard PB 50 test. They then set the direct level for 50% articulation and measured the articulation scores in the presence of reflections with varying delay times and levels. The results of these experiments are shown in Figure 6. The ordinate is the increase in effective level due to the reflection – positive values indicate that the reflection improved speech intelligibility by an amount comparable to an increase in speech level; negative values indicate that speech intelligibility was degraded.

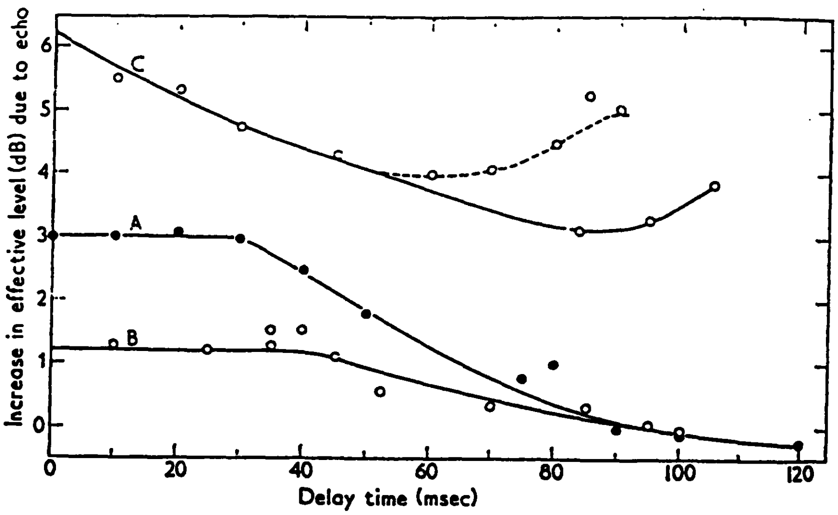

Figure 6. Integration curves for a single echo of (A) the same intensity as the primary sound, (B) 5 dB below the primary sound level, (C) 5 dB above the primary sound level. (Lochner and Burger, 1964)

Curve A is for a single reflection at the same intensity as the direct. For delays less than 30 ms, the effect of the reflection is the same as a doubling (3 dB increase) of speech intensity. In Lochner and Burger's terminology, the reflection is completely integrated with the direct sound, an interpretation that is a wholly consistent with Vitruvius' description of *consonance*:

> *The consonant are those (reflections) in which the voice is supported and strengthened, and reaches the ear in words which are clear and distinct.*

At delay times greater than 30 ms, the integration effect decreases, and by 95 ms, the reflection begins to degrade speech intelligibility. Compared with echo thresholds for an equal-level reflection occurring between 30 and 50 ms (see Figure 4), the 95 ms delay time required for a reduction in speech intelligibility is notably longer. However, it should be noted that the data in Figure 6 were collected at a speech level for 50% articulation, approximately 15 dB HL, which is 10 dB less than the closest speech level shown in Figure 4. In fact, the trend indicated by Figure 4 is that echo threshold delay times increase with decreasing overall level. Consequently, a 95 ms equal-level echo threshold for a 15 dB HL overall level may not be unreasonable.

A direct answer to the question of how echo thresholds relate to the reflection conditions that degrade speech intelligibility has not been provided. We may conclude, however, from the data presented, that the two appear to be closely related.

1.1.3 Early-to-Late Energy Ratios

Lochner and Burger found that speech articulation scores in listening conditions with multiple reflections or reverberation were similar to their single reflection data and consistent with the hypothesis that sound energy arriving within certain time windows after the direct could be either "perfectly" integrated with the direct signal, in which case the energy could be considered as part of the signal; "partially" integrated; or not integrated, in which case the late-arriving energy would be detrimental to speech intelligibility and could be considered as noise. The "critical delay times" defining these time windows are determined by the level of the reflected energy relative to the direct. Using these critical delay times, they developed a measure of acoustical quality based on an effective signal-to-noise ratio derived for echoic conditions. The measure, now known as U$_{95}$, is a weighted ratio of early-to-late sound energy density: (Lochner and Burger, 1964)

$$U_{95} = 10\log\left[\frac{\int_{0}^{95ms}\alpha(t)P^2dt}{\int_{95ms}^{\infty}P^2dt}\right] dB$$

where *P* is the instantaneous sound pressure measured after an impulse excitation of the room and *α(t)* is a weighting factor based on the amount of integration for different levels and delay times derived from the data shown in Figure 6. The measure is versatile in that the effects of background noise may be included by adding a term to the integrand in the denominator for the average power of the noise. For a series of rooms, Lochner and Burger found that the derived values of U$_{95}$ showed strong correlations with speech articulation scores. This being the case, they proposed the measure be used as a predictor of speech intelligibility.

U$_{95}$ is closely related to several other early-to-late acoustical measures. C$_{80}$, thought to correspond with musical clarity in concert hall acoustics, is a ratio of the sound energy in the first 80 ms to that beyond 80 ms: (Reichart, Abdel Alim et al., 1974)

$$C_{80} = 10\log\left[\frac{\int_{0}^{80ms}[g(t)]^2dt}{\int_{80ms}^{\infty}[g(t)]^2dt}\right] dB$$

where *g(t)* is the room response to a broadband impulse measured at a specific seating location. Another objective criterion, called "definition" (D), was claimed to correlate with the distinctness of sound in reverberation: (Thiele, 1953)

$$D = \frac{\int_{0}^{50ms}[g(t)]^2dt}{\int_{0ms}^{\infty}[g(t)]^2dt}\bullet 100\%$$

Bradley has identified at least eight such measures, each with a different integration time in the numerator or denominator, or a different weighting scheme for early energy (Bradley, 1986a). Santon has presented a similar early-to-late energy approach to predicting the intelligibility of speech in rooms using computer-based modeling of room acoustics (Santon, 1976).

1.1.4 Speech Transmission Index

A somewhat different approach to predicting speech intelligibility in echoic conditions was taken by Houtgast and Steeneken in the 1970's and 80's. Their criterion, the Speech Transmission Index (STI), measures the degradation of the temporal envelope of the speech signal caused by factors such as noise, echoes, reverberation, or amplitude compression (Houtgast and Steeneken, 1972; Houtgast and Steeneken, 1973; Houtgast and Steeneken, 1980; Plomp, Steeneken et al., 1980; Rietschote, Houtgast et al., 1981; Wattel, Plomp et al., 1981; Rietschote and Houtgast, 1983; Houtgast and Steeneken, 1985).

Speech is a highly dynamic signal, characterized by large fluctuations in amplitude between and within vowels and consonants. Houtgast and Steeneken measured the amount of modulation in the intensity level of running speech as a function of the modulation frequency (*F*). The resulting envelope spectrum for a 1/3-octave band of speech is shown in Figure 7.

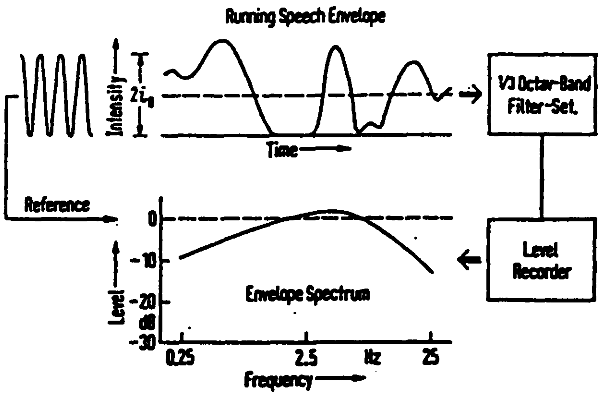

Figure 7. The fluctuations of the envelope of running speech as represented by the envelope spectrum. The spectrum is normalized by defining 100% (intensity) modulation as 0 dB.

They found that regardless of the talker or the center frequency of the 1/3-octave band, the envelope spectrum of running speech had most of its energy in the range from 0.63 - 12.5 Hz, with a maximum near 3 Hz – the average rate of syllable formation. Noting that reverberation and noise would reduce the amount of envelope modulations by "filling-in" the soft and silent segments of the speech signal, they hypothesized that this modulation reduction is the basic mechanism by which speech intelligibility is degraded. Furthermore, for a given system, whether linear or nonlinear, the amount of modulation reduction as a function of modulation frequency is described by the modulation transfer function (MTF).

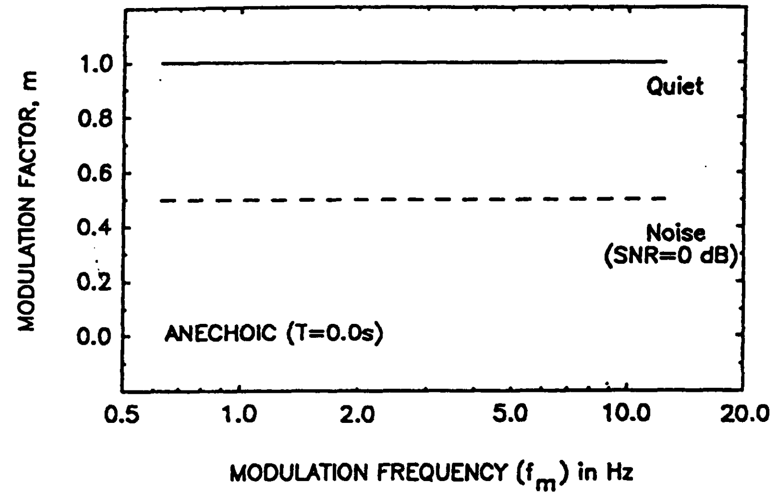

Figure 8. Modulation transfer function (MTF) obtained in an anechoic environment in quiet (solid line) and in noise at a signal-to-noise ratio (SNR) of 0 dB (dashed line). (Humes, 1993 p. 206)

By way of example, the MTF for a quiet anechoic environment is unity for all modulation frequencies; that is, envelope modulations are transmitted between a source and receiver without degradation. On the other hand, in white noise, modulations are degraded equally at all modulation frequencies and the MTF is flat, with a value that depends on the signal-to-noise ratio (SNR). Figure 8 shows the MTFs for these simple conditions. The modulation factor, *m*, is the amount of modulation at the receiver relative to that at the source.

The MTF for a linear system is the complex Fourier transform of the squared impulse response divided by its total energy: (Houtgast and Steeneken, 1980; Schroeder, 1981)

$$m(F) = \frac{\int_0^\infty h^2(t) e^{-j2\pi F t} dt}{\int_0^\infty h^2(t) dt}$$

The impulse response *h(t)* of a room may be determined experimentally or modeled stochastically as an exponentially decaying white noise process: (Schroeder, 1981)

$$h(t) = e^{-t/\tau} w(t),$$

where w(t) is a sample function from a stationary white-noise process. The relationship between the time constant τ and the reverberation time RT_{60} can be shown to be

$$\tau = \frac{2RT_{60}}{13.8},$$

where RT_{60} is related to the room volume *V* and absorption *A* by Sabine's reverberation equation:

$$RT_{60} = \frac{0.16V}{A}$$

Thus, for a room with known reverberation time, a simplified approximation of the impulse response is

$$h(t) = e^{-13.8t/2RT_{60}} w(t),$$

in which case, the MTF reduces to

$$m(F) = \left[1 + \left(2\pi F * RT_{60}/13.8\right)^2\right]^{-1/2}$$

Using the preceding equations, the MTF for any existing or theoretical room may be derived. To calculate a single STI index between 0 and 1, a 7 × 14 matrix of *m* values consisting of MTFs (with 14 modulation frequencies) for each of 7 octave-bands within the speech audio frequency range are combined. In the first step, each *m* value is converted into an apparent signal-to-noise criterion SNR_{app}

$$SNR_{app} = 10\log\left(\frac{m}{1-m}\right)$$

These SNR_{app} values are truncated to within the range of ±15 dB and then averaged across the modulation frequency matrix dimension. A weighted sum of the seven octave-band average SNR_{app} values is then computed, where the weighting factors applied to the audio frequency bands depend upon empirical data indicating the band's relative contribution to speech intelligibility and upon masking effects between adjacent frequency bands. Finally, the overall SNR value SNR_{tot} is converted into the STI:

$$STI = \left[SNR_{tot} + 15\right]/30.$$

As for the early-to-late energy measures, the STI has been shown to correlate with speech intelligibility scores. Figure 9 shows the correlation between the STI and Harvard PB-word intelligibility test scores for various listening conditions including reverberation.

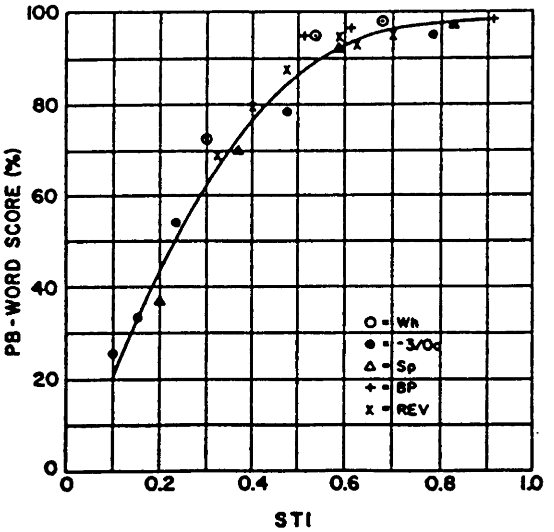

Figure 9. STI/speech intelligibility curve for the Harvard PB-word test plotted with the averages for each condition (Wh = white noise, -3/oc = -3 dB/octave noise, Sp = speech spectrum shaped noise, BP = bandpass filtering, REV = reverberation). (Anderson and Kalb, 1987)

It is interesting to note that the Lochner-Burger definition of echoic signal-to-noise as an early-to-late energy ratio falls out of the STI treatment as a natural consequence. Figure 10 shows the relationship between the Lochner-Burger early-to-late ratio and the STI as a function of reverberation time. The STI curve (solid line) falls within the 60 to 90 ms "critical delay times" found by Lochner and Burger (dashed lines).

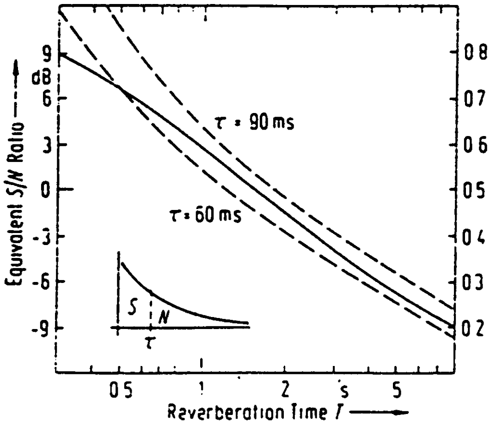

Figure 10. The solid curve represents the relation between reverberation time T and STI (right-side ordinate) or between T and equivalent S/N ratio (left-side ordinate) as resulting from the STI calculation scheme. The dashed curves illustrate the traditional approach in which an equivalent S/N ratio is derived from the ratio between the early and the later part of the echogram, for a temporal boundary at τ = 60 ms or τ = 90 ms. (Houtgast and Steeneken, 1980)

Several reports have demonstrated that the Lochner-Burger signal-to-noise method has similar predictive value to STI (Latham, 1979; Bradley, 1986a; Jacob, 1989). In a study of speech intelligibility in 12 and 13 year-olds in classrooms, Bradley found that U_{35}, U_{50}, U_{95}, and STI were the most accurate predictors and had "essentially equivalent predictive accuracy" (Bradley, 1986b).

Despite the generally positive findings regarding STI as a predictor of speech intelligibility in reverberant conditions, several researchers have raised questions about its limitations. One study has shown that speech intelligibility in the presence of a disturbing echo is dependent upon the horizontal angle of incidence of the echo, with speech intelligibility improving as the angle approaches 90° (with 0° straight ahead) (Nakijama and Ando, 1991). This effect was shown to be independent of the speech intelligibility improvement predicted by STI. The authors argue that STI has limited predictive value because it does not account for adjustments due to binaural listening. For example, STI would not predict the 5% improvement in percent word correct scores, known as the binaural advantage, when reverberant speech is presented binaurally vs. monaurally (Nabelek and Robinson, 1982). Zurek has proposed a conceptual approach which might be used to incorporate binaural interactions into the STI (Zurek, 1993).

Schmidt-Nielsen (1987) has argued that the standard error of STI, which correlates to speech intelligibility scores of approximately ±10%, is too large to make reliable comparisons between, for example, competing professional speech reinforcement systems which may typically vary in performance by 5% or less. He points out that there is no evidence that small improvements in STI are correlated with improvements in speech intelligibility or vice-versa. Furthermore, STI does not reflect talker differences and, therefore, does not predict how specific (reverberant) conditions may affect the speech reception of different talkers.

Payton *et al.* (1994) have reinforced Schmidt-Nielsen's third argument about speaker differences. In this study, the authors compared speech intelligibility scores for clearly articulated (clear) speech versus normally articulated (conversational) speech under various signal-degrading conditions including reverberation. They showed that clear speech is more resistant to degradation by noise and reverberation. However, the calculated STI could not account for the differences due to speaking style. They suggest that since STI is based on the modulation spectrum of the speech signal, which reflects these style differences, it "has the potential to account for intelligibility differences due to speaking style." Furthermore, they point out, based on results from hearing impaired subjects, that STI could not fully account for the differences in speech intelligibility due to hearing loss, even after the specific frequency band weightings of the STI measure were modified to reflect the subject's audiogram and the listening conditions. Perhaps the most significant point of this paper is that the specific features of clear speech most responsible for improved intelligibility have not yet been isolated.

Humes *et al.* (1986) suggested an improved version of the STI, called mSTI, in which the audio frequency resolution of the index is increased from full-octave to 1/3-octave bands and the weighting factors developed by French and Steinberg for the articulation index (AI) are used for the weighted summing process across these 1/3-octave bands. They presented promising results on the prediction of speech intelligibility for hearing-impaired listeners using mSTI.

Regarding listener differences, it has been shown that speech perception by children, the elderly, and the hearing impaired is affected by reverberation more adversely than young normal-hearing adults (Nabelek and Donahue, 1984). Furthermore, listeners' linguistic backgrounds affect the results of speech perception tests (Bergman, 1980). Nabelek and Donahue showed that non-native listeners scored 10% lower than native listeners on consonant perception tests with RT_{60} values of 0.8 or 1.2 seconds. In another study, native Japanese speakers showed similar scores in similar conditions (RT_{60} = 1.2 seconds) (Takata and Nabelek, 1990). It was found that Japanese listeners made similar errors to native listeners, but in addition, showed typical confusions involving phonemes not found in Japanese, especially /θ/, /f/, and /l/. This finding has been explained in the context of categorical phonemic perception, which may also account for the reduced speech perception of children in reverberant environments. Needless to say, STI does not take into account differences in lexical resources between listeners.

In summary, while STI is a reasonably accurate predictor of group performance in reverberant conditions, it does not at this time seem to be a valid predictor for individual cases.

1.1.5 Phonemic Analysis of Speech Degraded by Reverberation

1.1.5.1 Consonants

Vitruvius wrote that reverberation confuses "the case endings". Since the early 1900's it has been reported that word-final consonants are more degraded by reverberation than word-initial consonants (Knudsen, 1929). Figure 11 clearly reveals this trend.

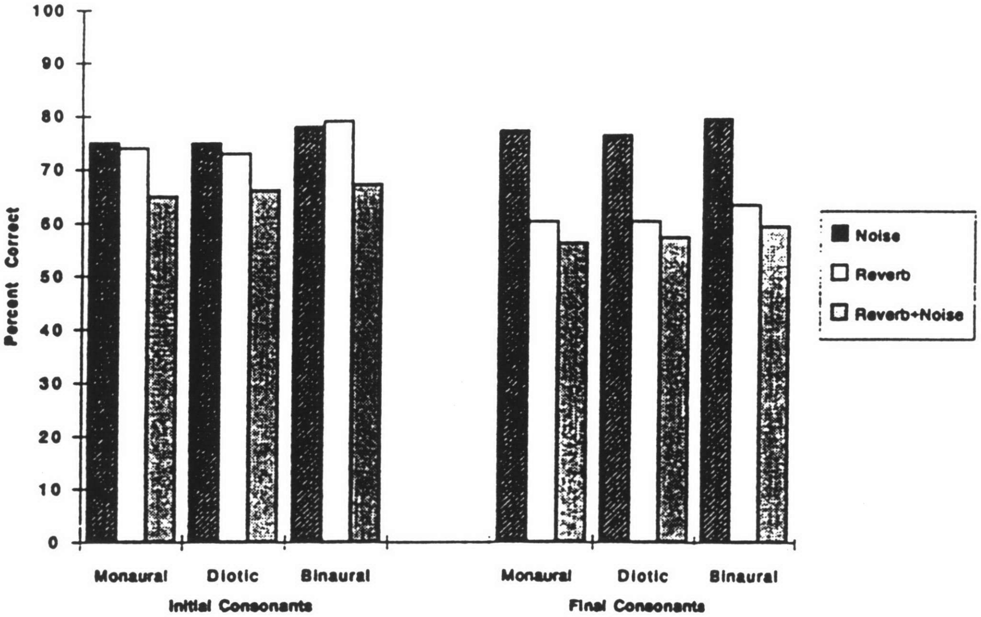

Figure 11. Accuracy of consonant identification by listening condition, presentation mode, and consonant position. (Helfer, 1994)

This effect is generally explained as resulting from reverberant overlap-masking, where decaying energy from a preceding phoneme overlaps the following phoneme. Overlap-masking is seen as distinct from reverberant self-masking, the internal temporal smearing within a phoneme. Self-masking, but not overlap-masking, may occur in initial consonants (Bolt and MacDonald, 1949; Nabelek, Letowski et al., 1989).

Different phonemic errors are seen between initial and final consonants. Errors for final consonants are reported as confusions between semivowels and nasals (Helfer and Huntley, 1991), dentals and their associated stops (esp. /p/ for /f/ and /b/ for /v/), and backed and fronted stops (esp. /k/ for /t/ and /t/ for /p/) (Helfer, 1994). Fricative confusions are explained as reverberant overlap-masking of high frequency energy. Nasal and semivowel confusions are seen as masking of F2 spectral information.

Typical initial errors are /p/ for /t/ or /k/ and /θ/ for /z/ or /v/ (Helfer, 1994), and /w/ for /m/ (Nabelek, Letowski et al., 1989). Self-masking is invoked in these cases.

Significant individual variability was observed in all of these studies.

1.1.5.2 Vowels

Similar to the initial/final distinction in consonants, vowel identification in reverberation is correlated better with the early part of the vowel than the later part (Nabelek and Letowski, 1988; Nabelek, Czyzewski et al., 1993). Significant effects of reverberation are seen in vowels with steep changes in upward direction of one of the formants, such as F1 in /æ/, /Λ/, /a/ and /au/ and F2 in /i/ in a /b-t/ context. It is suggested that reverberant energy maintains the initial steady-state portion of the vowel or diphthong into the transition region, causing self-masking of the transition.

In one report, the perception of a synthetic two-formant diphthong /ai/ in a /b-t/ context was studied (Nabelek, 1994). In general, diphthongs in reverberation may be confused with the monophthong having

F1 and F2 of the same or approximate value as the beginning of the diphthong. In this study, the perception of the diphthong was measured as a function of the rate of formant transition. Figure 12 shows a schematic for the synthetic stimuli used in the experiment.

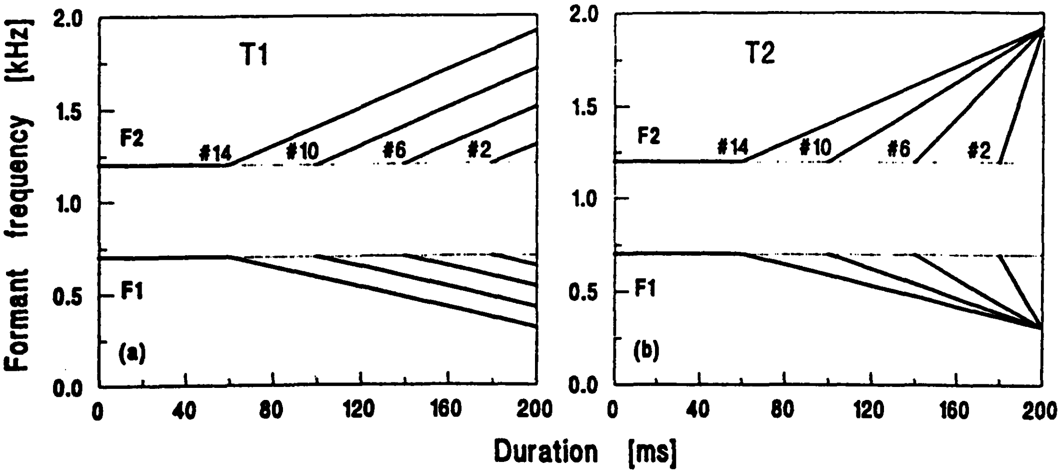

Figure 12. Schematic representation of two types of synthetic diphthong stimuli: (a) T1, with formants changing at the same rate and in the direction of target frequencies; and (b) T2, with formants changing at different rates and reaching target frequencies. Numbers refer to the stimulus numbers. (Nabelek, 1994)

T1 stimuli all had the same rate of transition, but did not reach the target formants. T2 stimuli had different rates, but always reached the targets. The authors found that normal hearing (NH) subjects needed slower transitions to perceive the diphthong clearly in reverberation. It is interesting to note that while noise and reverberation had similar effects for the NH group, errors in reverberation were worse than in noise for a hearing-impaired (HI) group, which required slower T2 transitions (110 ms) in reverberation. Furthermore, if the transition did not reach the target (T1 stimuli), even longer transitions (130 ms) were required for the HI group, whereas NH listeners were not affected by noise or reverberation with the T1 stimuli, as the rate of change was sufficiently slow for clear identification. "Clear speech", then, is related to the rate of formant transition for the diphthong, /ai/, and the requirements for clear speech change with increasing amounts of reverberation and with the listener.

In a follow-up paper, the perception of diphthongs in reverberation was studied as a function of the attenuation of the later part of the vowel (Nabelek, Ovchinnikov et al., in press). Figure 13 shows the two extreme tokens in a continuum of 16 synthetic diphthongs, where the later part of the diphthong was attenuated in steps of 1 dB. A linear attenuation slope was applied from the beginning to the end of the vowel.

It was found, as shown in Figure 14, that diphthongs with significantly attenuated formant transitions were always identified correctly in quiet, but, depending upon the intensity level of the transition, could be mistaken for the initial monophthong in reverberation, supporting the hypothesis that reverberation masks the formant transitions that are responsible for the perception of the diphthong.

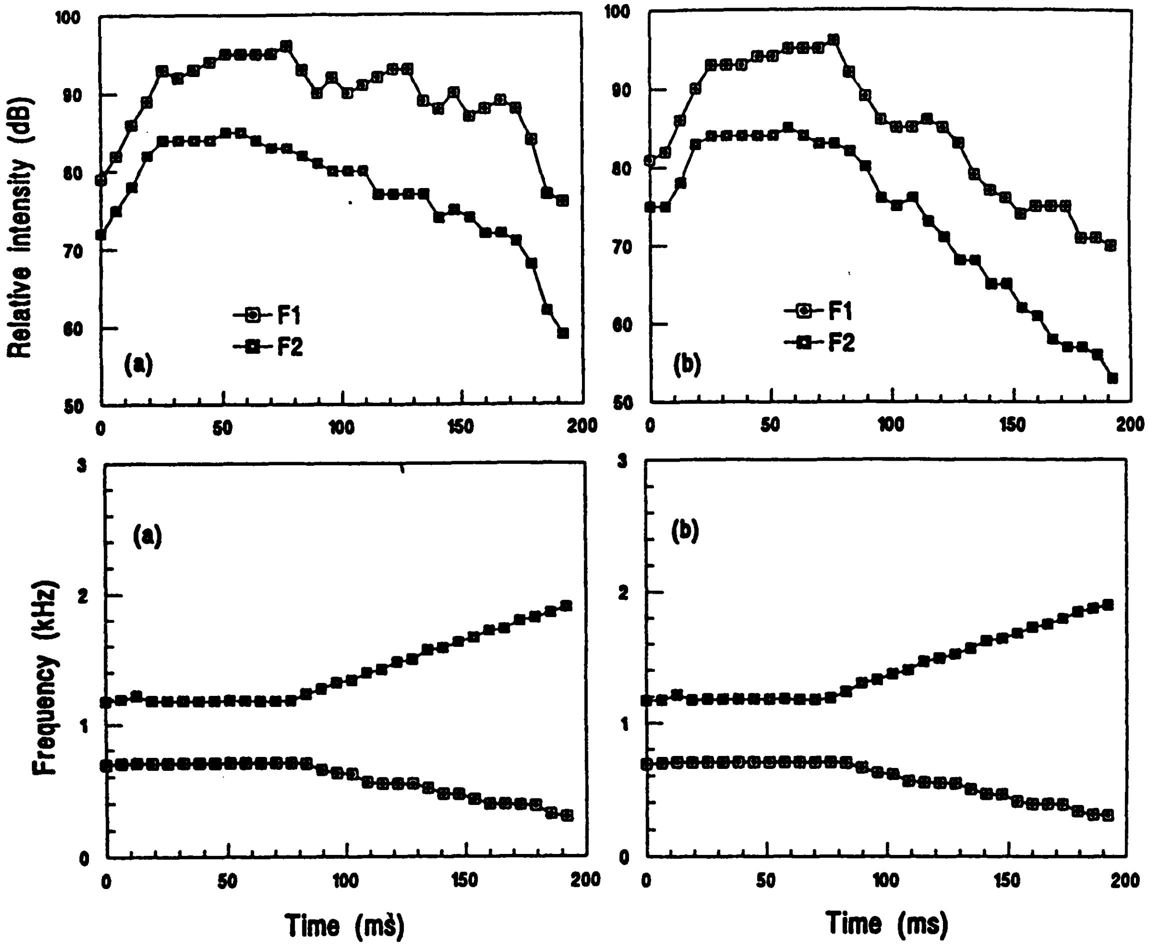

Figure 13. Relative intensity levels and frequencies of F1 and F2 as functions of time of: a) synthesized /ai/ stimulus (No. 16) without attenuation and b) synthesized /ai/ stimulus (No. 1) with 15-dB attenuation of the transition. (Nabelek, Ovchinnikov et al., in press)

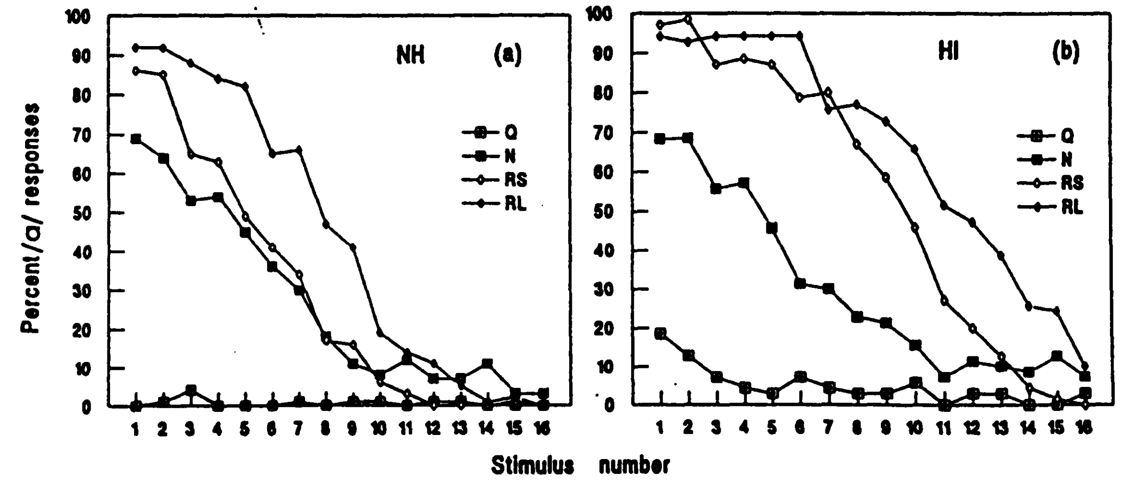

Figure 14. Percent /a/ responses as functions of stimulus number for quiet (Q), noise (N), short reverberation (RS), and long reverberation (RL) conditions. Mean responses for: a) ten normal-hearing (NH) subjects and b) seven hearing-impaired (HI) subjects. (Nabelek, Ovchinnikov et al., in press)

1.1.5.3 Concerns Regarding Phonemic Error Studies

To address the limitations of STI presented earlier, a thorough and systematic study of specific phonemic errors in reverberant conditions must be carried out. Work towards this end has begun, but the field is far from understanding the specific nature of phonemic errors on an individual basis in reverberation. Several problems with current approaches seem to be evident.

At present, there does not seem to be a standard for reverberant conditions used between researchers. In general, reverberation times of between 0.6 and 1.9 seconds are used, but conditions may include: monaural, diotic, or binaural playback (each with a different contribution from the "binaural advantage") and reverberant stimuli recorded at critical distance (where the reverberant level equals the direct) or at given distances beyond the critical distance. The impulse responses of the reverberant spaces used to generate reverberant stimuli are generally not provided; thus, specific information about the early reflection structure of the reverberant conditions and the reverberation times for different frequency bands is not available. In 1987 Humes *et al.* performed a retrospective study that calculated the STI for several speech intelligibility studies using various reverberation conditions (Humes, Boney et al., 1987). The authors found a monotonic relationship between STI and percent correct speech scores for each study but found it difficult to make comparisons between the studies because "a number of assumptions had to be made regarding such things as the spectra of the speech ..., the frequency independence of reverberation time, the directivity of the loudspeakers, and so on."

Regarding the binaural advantage for speech intelligibility in reverberation, Helfer has pointed out that "differences among the three presentation modes (monaural, diotic, and binaural) (are) subtle" (Helfer, 1994). She found that word-final back stop consonants are more likely to be discerned in binaural reverberation and suggested that this difference might be due to enhanced perception of low frequency information binaurally. However, other improvements cannot be explained in this manner; specifically, reduction of the following errors: /p/ for /t/, /g/ for /d/, and /p/ for /k/. Otherwise, "error patterns for phonemes obtained with monaural presentation may be generalized to binaural listening."

1.2 Representations of Speech in the Discharge Patterns of Auditory Nerve Fibers

This section is meant to provide a brief overview of the neural representations of the speech signal at the level of the auditory nerve, with reference to how such representations might be influenced by echoic listening conditions. The reader is referred to several excellent reviews for a more detailed discussion of the auditory processing of speech (Greenberg, 1996; Delgutte, in press).

Information about a stimulus may be represented at the level of a single neuron either by changes in the neuron's discharge rate or in the temporal patterns of its spiking response. The various acoustic features of speech (e.g., formant transition rate, formant level, burst duration and spectrum, voice-onset time, etc.), if they are represented in the discharge patterns of auditory-nerve fibers (ANF), must be coded by either one or both of these mechanisms. "Rate-place" information refers to changes in discharge rates across a population of ANFs tuned to different characteristic frequencies (CF). In this sense, "rate-place" information may resemble the instantaneous spectrum of speech. "Temporal" information refers to changes in the fine time patterns of the spiking responses of ANFs.

1.2.1 Rate-Place Information

The "rate-place" representation of steady-state synthetic vowels has been studied extensively (Sachs and Young, 1979; Delgutte, 1980; Sachs and Young, 1980; Delgutte and Kiang, 1984a; Delgutte and Kiang, 1984b; Delgutte and Kiang, 1984e). These reports have concluded that, at normal and high levels of speech, a "rate-place" representation in high spontaneous rate (SR) fibers does not resolve separate formant peaks because the fibers with their CFs between the formant frequencies have become saturated. Low SR units, on the other hand, have higher thresholds and greater dynamic range than high SR units, and a detectable representation of formant peaks may be observed in these units at levels up to 80 dB-SPL or higher (Sachs, Winslow et al., 1988). The data on dynamic speech-like sounds indicate that average discharge rates can provide more information about the spectra of formant transitions than of steady-state vowels, suggesting that "rate-place" information might be important to the coding of consonants and diphthongs, for which formant transitions are a prominent acoustic cue (Delgutte, 1980; Miller and Sachs, 1983; Sinex and Geisler, 1983; Delgutte and Kiang, 1984c; Delgutte and Kiang, 1984d). Delgutte (1980) has attributed this result to the rapid and short-term adaptation properties of ANFs (i.e., the decay in discharge rate occurring within 5 to 100 ms following an abrupt onset):

> *In general, at the beginning of a speech segment, units tuned to the major frequency components of the preceding (adapting) segment would discharge at lower rates. Thus, short-term adaptation would increase contrast between successive segments in the profile of discharge rate versus CF.*

Adaptation produces prominent transient responses at the onset of speech segments, especially in high SR units (see Section III. Preliminary Results, p. 27). Using a model of peripheral signal processing including adaptation, Delgutte showed that voice-onset time could be reliably detected as the difference in times of burst onsets for high-CF and low-CF fibers (Delgutte, 1986). Delgutte and Kiang (1984d) have suggested that "the peaks in discharge rate that occur in response to rapid changes in amplitude or spectrum might be used by the central processor as pointers to portions of speech signals that are rich in phonetic information."

The potentially important role of adaptation in enhancing spectral contrast in the "rate-place" representation of speech has implications for the mechanism of speech degradation by echoes, reverberation, and noise. In order to respond to a new stimulus with a maximal discharge rate at the onset, ANFs must have adequate time to recover from previous adaptation (Smith and Zwislocki, 1975; Smith, 1977; Relkin and Doucet, 1991). As a highly dynamic signal, speech in quiet provides time-out periods during which such recovery may take place. But when reverberation, for example, "fills-in the soft and silent segments of the speech signal", recovery from adaptation may not be complete, and the "rate-place" representation of important acoustic speech cues may be degraded. This conceptual framework is consistent with the underlying hypothesis of STI that modulation reduction is the basic mechanism by which speech intelligibility is degraded.

1.2.2 Temporal Information

The discharge patterns of ANFs can phase-lock with periodic stimuli at frequencies up to approximately 4000 Hz. This "synchronization" capacity of ANFs allows them to carry temporal information about the fundamental and all important formant frequencies of speech in the timing of their spiking responses. Robust representations of the formants of steady-state vowels have been observed in the synchronous responses of ANFs, even at high levels (Young and Sachs, 1979; Delgutte, 1980; Sachs and Young, 1980; Miller and Sachs, 1983; Delgutte and Kiang, 1984a). Unlike the "rate-place" scheme, in which ANFs carry spectral information that is CF-specific, in a temporal code, ANFs may carry information about off-CF frequencies if their discharge patterns are synchronized to those frequencies. In fact, many units with CF in the range of F1 tend to synchronize to F1, units with CF in the range of F2 synchronize to F2, and so on, although the actual synchronous responses depend upon the context. High CF units receiving little stimulation by frequencies at their CF can show prominent synchronization to the fundamental (Kiang and Moxon, 1974).

At normal and high speech levels, the non-linear effect of synchrony suppression tends to bias the synchronous response towards the dominant low frequency components of the speech signal. Termed "synchrony capture", the effect of synchrony suppression can be to enhance the temporal representation of formant frequencies, especially F1, in populations of ANFs (Greenberg, 1996). "Synchrony capture" may be one of the mechanisms underlying "reverberant self-masking". For example, in reverberation, energy from the initial steady-state portion of a vowel or diphthong is maintained into a subsequent formant transition region. If this transition contains a steep change in the upward direction (recall that such transitions are most significantly effected by reverberation, see p. 13), then the lower frequency reverberant energy may capture the synchronous response long enough to mask the transition.

The somewhat place-specific behavior of temporal coding has prompted the use of the Average Localized Synchronized Rate (ALSR) in data analysis. The ALSR calculates the average synchronous response to a given frequency of fibers whose CFs are within 0.25 or 0.5 of that frequency, which is typically a harmonic of the speech signal (Young and Sachs, 1979; Sachs and Young, 1980). This analysis scheme can overlook the synchronous response to frequencies well off-CF in certain populations of fibers. Alternatively, pooled (or "ensemble") interval histograms may be obtained for fibers across the entire CF spectrum. The Fourier transform of these histograms provides the relative level of the synchronous response to all frequencies for the population of fibers studied (Delgutte, in press).

1.3 The Physiological Basis of Echo Perception

1.3.1 Forward Masking in the Auditory Nerve

The precedence effect has been called "echo suppression" by researchers proposing monaural or binaural neural inhibitory mechanisms for the phenomenon (Zurek, 1979; Wickesberg and Oertel, 1990; Yin, 1994). The question whether echo suppression might be mediated by peripheral adaptation has been addressed indirectly by a forward masking study of the auditory nerve in which the thresholds for detection of a probe tone at CF were measured in the presence of a preceding CF masking tone (Relkin and Turner, 1988). A two-interval forced-choice (2IFC) adaptive up-down procedure was used to determine the just-perceptible difference in spike counts ($d' = 0.82$) during the time window of the probe. The stimulus conditions are shown in Figure 15.

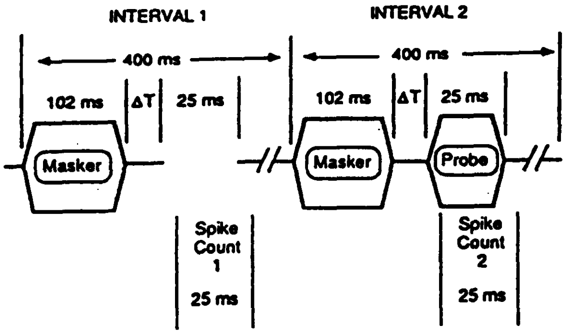

Figure 15. Timing of the electrical signal used to generate the stimuli. All ramps were 2 ms, cosine squared. The probe was randomly presented either in interval 1 or in interval 2. Spike counting intervals were delayed 2 ms relative to the electrical stimulus to allow for the latency of the neural response. (Relkin and Turner, 1988)

For our purposes, the probe may be considered as a reflection and potential echo. Thus, the detection threshold of the probe corresponds to a peripheral "echo threshold". In this experiment, the delay time of the probe from the onset of the masker was always 102 ms. The results are shown in Figure 16.

The thresholds for high SR units are significantly lower than those for low SR units because adaptation in high SR fibers reduces the spontaneous rate, making a post-masker stimulus more detectable because the noise floor for its detection is lowered. The neural thresholds in Figure 16 are low compared with the behavioral data for similar stimulus parameters showing masked thresholds of 50-70 dB SPL in human and similar levels in chinchilla, the experimental animal used in this study. In other words, detectable representations of the probe appear in ANFs before psychophysical detection occurs. The authors point out that "these results imply that behavioral forward masking must result from suboptimal processing of spike counts from auditory neurons at a location central to the auditory nerve." The results also indicate that echo suppression is not mediated by peripheral adaptation because echo thresholds are even higher than forward masking thresholds (Blauert, 1983 p. 225; Zurek, 1987).

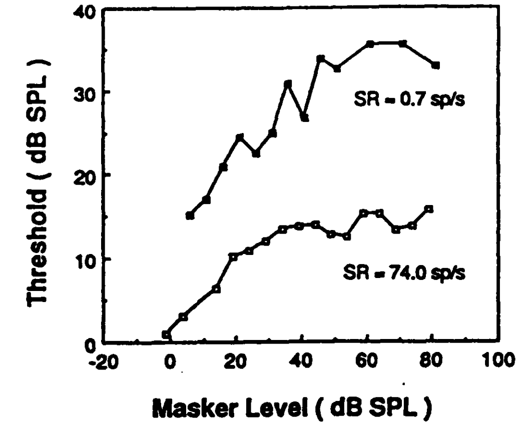

Figure 16. Growth-of-masking functions for two fibers, one with low spontaneous rate and the other with high spontaneous rate. Note that the threshold saturates at high-masker intensities. Also see that the range of thresholds is greater for low spontaneous rate fibers compared to high spontaneous rate fibers. (Relkin and Turner, 1988)

1.3.2 Echo Suppression in the Cochlear Nucleus

Wickesberg and Oertel (1990) have identified an inhibitory neural pathway in tissue slice preparations of mouse cochlear nuclei which they suggest could be a source of monaural echo suppression. The circuit they identified projects in a frequency-specific fashion from the deep dorsal cochlear nucleus (DCN) to bushy and stellate cells in the anteroventral cochlear nucleus (AVCN) and produces 5 ms duration inhibitory post-synaptic potentials (IPSPs) which are "delayed by a additional synaptic delay with respect to the excitation" (approximately 600 μs). The time course of this inhibition compares favorably with the psychophysical time course of monaural echo suppression for clicks (Harris, Flanagan et al., 1963; Zurek, 1980).

Kaltenbach, Maleca *et al.* (1993) recorded from single units in the DCN using a forward masking stimulus protocol. They found that responses to the probe could be suppressed for durations up to 300 ms following the masker, significantly longer than could be accounted for by peripheral adaptation in the ANF inputs to these cells. Their results support the hypothesis that delayed inhibitory mechanisms within the DCN may play a role in monaural echo suppression.

1.3.3 Physiological Correlates of the Precedence Effect in the Inferior Colliculus

Recording from single direction-sensitive binaural units in the inferior colliculus (ICC) of the cat, Yin (1994) observed responses to free-field and dichotic stimuli that correlated with the precedence effect. The stimuli were composed of a leading click, presented at one location, followed by a spatially-separated reflection, whose delay time was varied from 0 up to 150 ms following the leading click. The directional response properties of a typical ICC cell studied by Yin and its response to the precedence stimuli are shown in Figure 17.

An arbitrary echo threshold for the lagging click may be defined as the inter-click delay (ICD) at which the response to the lagging click is 50% that of the lead. For the unit in Figure 17, the echo threshold is approximately 50 ms. Yin found that this threshold ranged from 1.5 ms to 100 ms, with a median of 20 ms. The smallest echo threshold corresponds with that observed behaviorally for clicks – about 2 ms.

Yin's results are consistent with the following statements: The neural representation of early reflections is suppressed or inhibited in binaural cells adapted for directional sensitivity. This suppression of early reflections is a physiological correlate of the precedence effect.

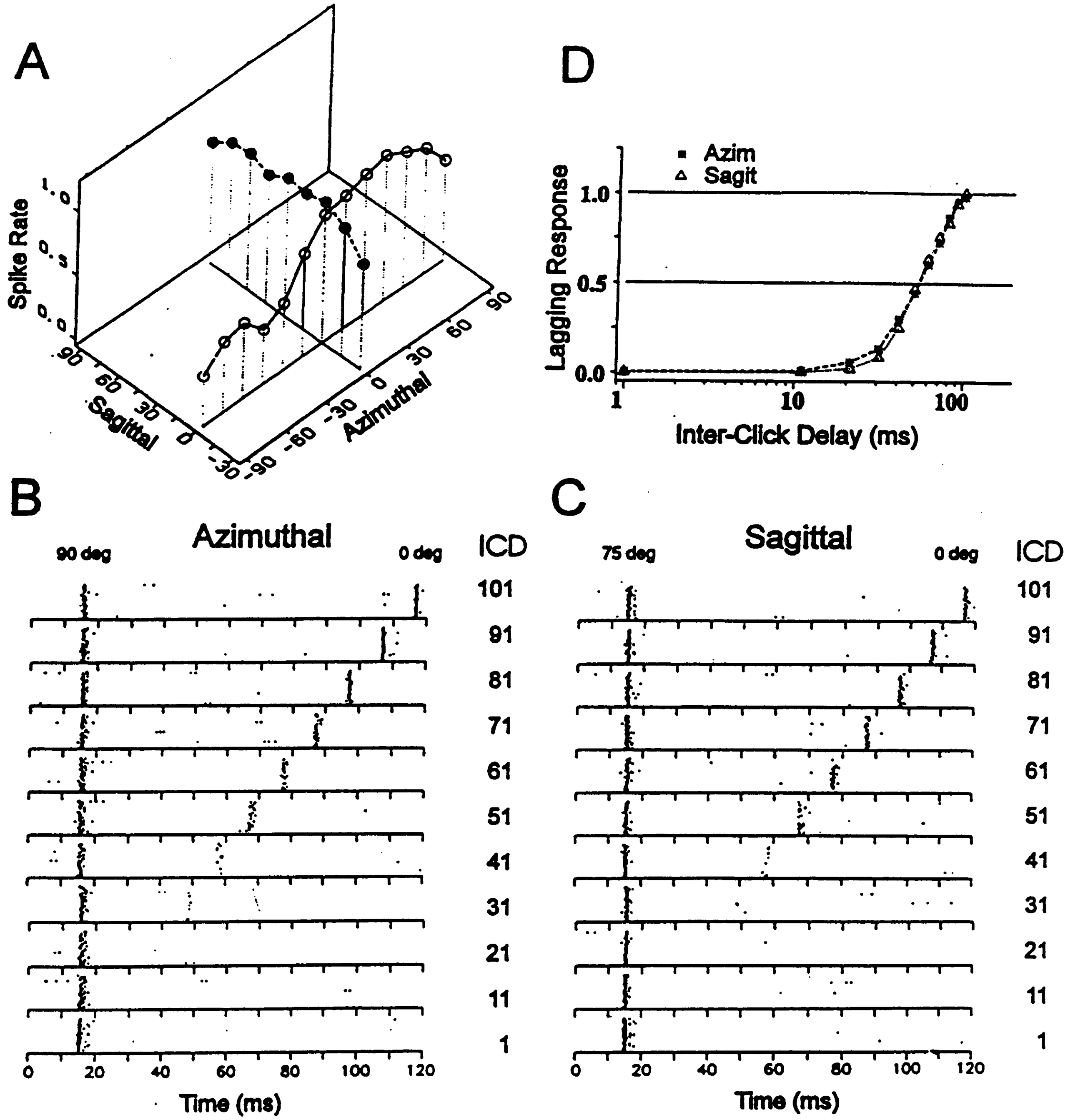

Figure 17. A. Directional response properties of a direction-sensitive binaural unit in the inferior colliculus of the cat: positive azimuth = contralateral field; positive elevation = up; 0° is straight ahead. B. Free-field response to equal-level leading and lagging clicks in the horizontal plane as a function of inter-click delay (ICD); azimuth as indicated. C. Free-field response to equal-level leading and lagging clicks in the sagittal plane as a function of inter-click delay (ICD); elevation as indicated. D. Lagging response normalized to leading response as a function of ICD. Azimuthal and sagittal conditions as indicated. (Litovsky, R., 1996 unpublished figure)

II. Experiments

In this section we will present two sets of experiments that explore the neural mechanisms by which speech intelligibility is affected in the echoic conditions of large rooms. In Experiment I, we will quantify the effects of reverberation on "rate-place" and "temporal" representations of speech in the auditory nerve. Our goal is to relate changes in neural representations, especially their degradation, to conditions affecting speech intelligibility. Experiment I is important because it attempts to determine the "worst-case" condition in which peripheral neural representations are no longer available to the central auditory nervous system. In Experiment II, we will measure and compare the detection thresholds for late-arriving reflections in a reverberant field between ANFs and ICC direction-sensitive cells. This study is important because behavioral echo thresholds may correspond with the reflection conditions that degrade speech intelligibility (see Section 1.1.2 The Precedence Effect, p. 5) and because an explicit comparison of echo thresholds in reverberant conditions between the periphery and the ICC has not yet been performed. Both experiments are straightforward in that the technical requirements for their execution have already been developed, and precedents have been established for the successful completion of such investigations. To be consistent with these precedents, the electrophysiology proposed here will be performed in the anesthetized cat (see Appendix C: Vertebrate Animals, p. 33).

2.1 Experiment I: Auditory Nerve

2.1.1 Experiment I-a A Population Study of the Effect of Reverberation on Consonant Coding in Auditory-Nerve Fibers of the Cat

2.1.1.1 Description

Preliminary studies have indicated a relationship between the modulation transfer function (MTF) of ANFs, their rapid and short-term adaptation properties, and the envelope of their discharge responses to speech (see Section III. Preliminary Results on p. 27). We found that adaptation produces prominent transient responses at the onset of speech segments, especially in high SR units, and that a model based on the neural MTF could simulate this response characteristic. In addition, these transients were reduced or eliminated when speech was presented in reverberation, and model predictions were consistent with this effect. Based on the hypothesis that the response transients occurring at the onset of speech segments are important both in a "rate-place" representation of dynamic spectral information and as "pointers to portions of speech signals that are rich in phonetic information" (see Section 1.2.1 Rate-Place Information, p. 16), we suggest that the reduction of these transients in reverberation should correlate with the degradation of speech intelligibility. Specifically, the reduction of the transient representation occurring for different word-final consonants in different vowel-consonant (VC) contexts should correlate with the phonemic errors reported in the literature (see Section 1.1.5 Phonemic Analysis of Speech Degraded by Reverberation, p. 13).

2.1.1.2 Experimental Protocol

To test this hypothesis, a quantitative measure of the transient response is required. One such measure, the onset-to-adapted rate ratio (O-A Ratio), has been used in the study of adaptation (Müller and Robertson, 1991). The method used to estimate this ratio is illustrated in Figure 18.

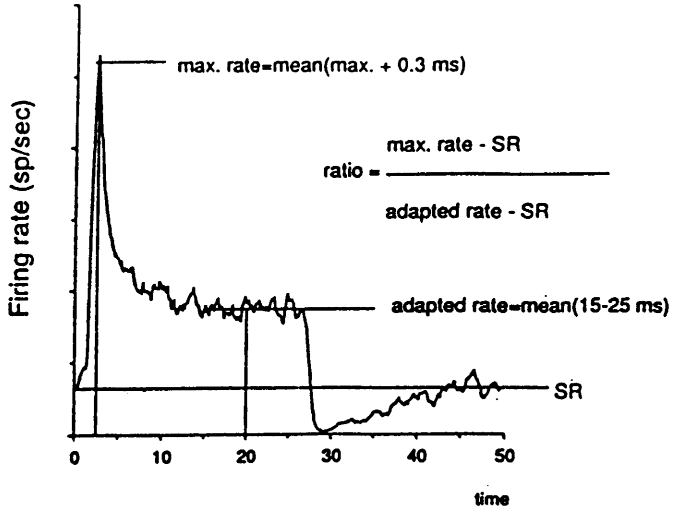

Figure 18. Schematic illustration of the method used to estimate an onset-to-adapted rate ratio from peri-stimulus time histograms (PSTHs). (Müller and Robertson, 1991)

The experimental protocol involves two phases. In the first phase, the MTF model will be used to simulate a population of ANF PSTH responses to natural CVC tokens as a function of the reverberation times of large rooms (e.g., 0.8, 1.5, 2.1 sec.). Using a criterion such as the O-A Ratio, the transient responses in these simulated fiber populations will be analyzed during the word-final consonant. Ensemble and place-specific O-A Ratios may be studied, and the results compared between the reverberant conditions and the non-reverberant control. Once a specific protocol for comparison has been determined, a computer-based analysis of a large number of CVCs, including those reported in the literature, will be generated. From this data, the subset of "most degraded" (MD) and "most resilient" (MR) tokens will be selected.

Phase two involves recording ANF population responses to the MD and MR tokens. The techniques for surgical exposure of the auditory nerve, delivery and calibration of acoustic stimuli, and single-fiber recordings are standard and have been described elsewhere (Kiang, Watanabe et al., 1965). Data from approximately 15 cats will be obtained. The population responses will be analyzed using the same computer-based procedure as in phase one, and the results will be compared with both the model predictions and the word-final errors reported in the literature. Comparisons between the ANF responses and the literature will be used to evaluate the initial hypothesis regarding the role of onset transients in speech coding. Differences between ANF responses and the model, which currently does not include two-tone nonlinear effects, could be used to improve the model.

2.1.2 Experiment I-b A Study of the Effect of Reverberation on the Coding of Diphthongs in Auditory-Nerve Fibers of the Cat

2.1.2.1 Description

Reverberation has a consistent and predictable influence on the perception of diphthongs: the final segments of the phoneme are masked by "internal temporal smearing ... called reverberant self-masking", and the diphthong may be confused with its initial monophthong (Nabelek, 1994). The physiological basis of reverberant self-masking at the level of the auditory nerve may be two-tone rate or synchrony suppression. The relative contribution of these factors can be measured in the discharge responses of ANFs. In this experiment, we will record responses of ANFs using the stimulus paradigm of Nabelek (in press) described on page 15, in which a continuum of the two-formant synthetic diphthong /ai/ is presented in varying conditions of reverberation. The continuum is generated by attenuating the later part of the token relative to the initial part, with a linear attenuation slope applied across the phoneme (see Figure 13). We will quantify the reduction due to reverberation in both rate and synchronous responses to the formant transitions and compare these results with the behavioral data of Nabelek shown in Figure 14.

2.1.2.2 Experimental Protocol

In order to avoid the requirement for a neural population study, a set of stimulus continua will be generated offline (using a Klatt synthesis algorithm) during the ANF recording session in response to the CF of the fiber being studied. The continua within this set will be shifted in frequency around the CF such that each set will contain stimuli having the initial formants and at least three specific points along the formant transition at CF. In this way, a large amount of data on the neural response during the transition may be collected from a single fiber. Although, depending on the CF, the stimuli may no longer sound like /ai/, the responses should still be relevant to the question of reverberant masking. Because unit holding times are limited in ANF experiments, the continua will be limited to between 4 to 8 attenuation steps, and the effects on F1 and F2 will be studied in different units. As in the behavioral study, two reverberant conditions (0.8 and 1.1 sec.) will be included in addition to the non-reverberant control. Data from approximately 5 cats will be obtained.

Data analysis will involve both synchrony and rate computations. For each stimulus condition, the synchronous responses to the initial formant and instantaneous formant transition frequencies will be computed as a function of time, and the ratio of these responses between the reverberant and non-reverberant conditions will be obtained. For the formant transition frequency, this synchrony ratio is a measure of the reduction of temporal information by reverberation. For the initial formant frequency, it measures the amount of "synchrony capture". To quantify the effect of reverberation on the rate-place representation of the formant transitions, the ratio of discharge rates between the reverberant and non-reverberant conditions will be computed for the "transition at CF" stimuli. This ratio is a measure of two-tone rate suppression in that the energy at frequencies maintained by reverberation may suppress the response to energy at other frequencies, namely, the transition frequency. Additionally, we may study the "transition at CF" responses to see if onset transients occurring as a formant passes through the CF are reduced in the reverberant conditions. The O-A Ratio described in Experiment I-a could be used in this instance. These synchrony and rate measures will be compared with the behavioral data to determine whether synchrony and/or rate suppression mechanisms are consistent with "reverberant self-masking".

2.2 Experiment II: Inferior Colliculus

Detection of Echoes in Reverberation: A Comparison Between Auditory-Nerve Fibers and Direction-Sensitive Cells in the Inferior Colliculus of the Cat

2.2.1.1 Description

Blauert (1983, p. 276) has noted that reverberant sound decay is uncorrelated between the two ears. This characteristic of reverberation, as illustrated in Figure 19, likely underlies our inability to form a localized perception of reverberant decay, just as uncorrelated noise presented dichotically sounds diffuse (Durlach and Colburn, 1978).

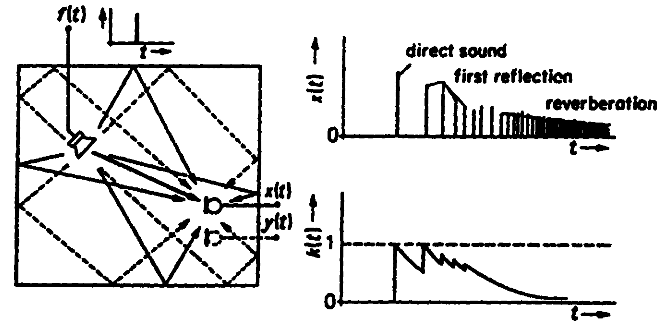

Figure 19. The impulse response of an acoustical transmission path in an enclosed space. Top right: Echogram. Bottom right: Interaural cross-correlation echogram. (Blauert, 1983 p. 276)

Many of the direction-sensitive cells in Yin's ICC precedence effect study were binaural units sensitive to interaural time differences (ITDs). As shown in Figure 20, these low-frequency units respond poorly to binaurally uncorrelated noise. On the other hand, high-frequency binaural units sensitive to interaural level differences (ILDs) can respond strongly to binaurally uncorrelated noise.

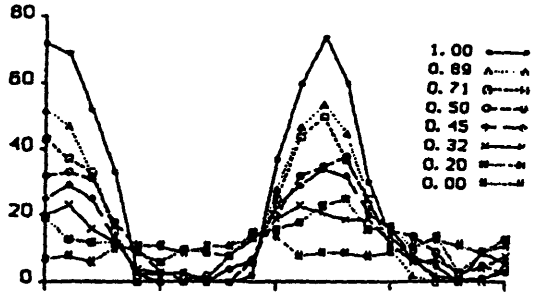

Figure 20. Responses of one cell to ITDs as a function of the cross-correlation of the stimuli to the two ears. The correlation coefficient varied between one (identical stimuli to the ears) and zero (uncorrelated noise). (Yin and Chan, 1988 p. 416)

Given this information and the earlier discussion of the precedence effect, several hypotheses may be formed regarding neural echo thresholds for a late-arriving reflection in reverberation:

1. Echo thresholds in ANFs should be greater in reverberation than in standard forward masking experiments – due to less recovery from adaptation and an elevated "noise floor" in reverberation.
2. Echo thresholds in reverberation should be greater in ICC direction-sensitive cells than in ANFs – consistent with the ICC precedence effect correlate in non-reverberant conditions.
3. ITD-sensitive cells in the ICC should be less responsive to the uncorrelated part of the reverberant decay than ILD-sensitive cells or ANFs – ITD units respond poorly to binaurally uncorrelated stimuli.
4. ICC direction-sensitive cells showing the lowest echo thresholds should "point" to the correct direction of the echo – consistent with their directional-sensitivity.
5. Reverberant echo thresholds in ICC direction-sensitive cells should correlate with the behavioral echo thresholds for similar stimulus conditions – consistent with the precedence effect correlation.
6. Reverberant echo thresholds in ICC direction-sensitive cells should correlate with acoustical conditions yielding poor speech intelligibility – assuming that suprathreshold late reflections degrade speech intelligibility.

Note that hypotheses 2 and 3 may be logically inconsistent. If the response to the reverberant decay is suppressed in ITD-sensitive cells (hypothesis 3), then the noise floor during a late-arriving, highly-correlated reflection may be lower in these units than in ANFs or ILD-sensitive cells. Contrary to hypothesis 2, the effect of a reduced noise floor might be to lower the echo threshold in ITD-sensitive cells.

Experiment II will test hypotheses 1 - 4. Hypotheses 5 and 6 are currently untestable as behavioral data is unavailable on reverberant echo thresholds and speech intelligibility in reverberation + echoes.

2.2.1.2 Experimental Protocol

The techniques for single-fiber recordings of the auditory nerve are standard and have been described elsewhere (Kiang, Watanabe et al., 1965). The techniques for recording from single units in the ICC of anesthetized cats have been described by Yin (1994) and are commonly used in our lab. Experiments currently underway in our lab make use of a "virtual space" (VS) stimulus system, by which stimuli with multiple, realistic localization cues are delivered to the experimental animal through a calibrated closed-acoustic system. The VS system has the capability to incorporate the binaural impulse responses of a reverberant room, thus creating a realistic reverberant listening environment.

Using the VS stimulus system, we will record population responses from ANFs and ICC direction-sensitive cells to a reverberant click, speech-filtered noise burst, or short speech token (to be determined in preliminary experiments) in the presence of a simulated late-arriving reflection. (ANFs will be characterized by their CF, SR, and rate-level function, and ICC units by their CF, SR, and azimuth- and elevation-sensitivity prior to echoic data collection. ICC units not displaying directional-sensitivity will not be included in this study.) The simulated reflection will be varied between two locations (1 azimuth and 1 elevation) and up to three delay times – the specific times will depend on the reverberation time, but in all cases will be greater than 95 ms. Reverberation times will be 0.8, 1.5, and 2.1 seconds. Auditory-nerve data from approximately 10 cats and ICC data from approximately 15 cats will be obtained. In some animals, data may be collected from the auditory-nerve and ICC simultaneously.

Using a 2IFC adaptive up-down procedure with intensity level of the reflection as the variable (Relkin and Pelli, 1987), echo thresholds ($d' = 1$) will be obtained based on the neural response within the time window of the reflection. These echo thresholds will be compared between ANFs and ITD- and ILD-sensitive cells in the ICC to evaluate hypotheses 1 and 2.

PSTHs for the reverberation minus reflection condition will be generated in order to compare neural responses to the reverberant decay. According to hypothesis 3, the responses of ITD-sensitive cells should decay faster than the reverberant decay, and faster than ANFs and ILD-sensitive cells.

To evaluate hypothesis 4, a directional vector L_l for each echo location l (azimuth or elevation) will be calculated from the population of ICC cells: (Georgopoulos, Taira et al., 1993)

$$L_l = \frac{1}{m}\sum_{i=1}^{m} w_{i,l}C_{i,l} ,$$

where m is the total number of units in the population, $C_{i,l}$ is the preferred location of the ith cell (i.e., the azimuth or elevation at which the ith cell has the greatest response), and $w_{i,l}$ is a weighting factor inversely proportional to the ith cell's echo threshold, such that units with the lowest thresholds are weighted most heavily. According to hypothesis 4, the directional vector L_l should correspond to the actual location of the reflection.

Based on echo thresholds in the ICC, predictions regarding hypotheses 5 and 6 (behavioral echo thresholds and speech intelligibility) will be made.

III. Preliminary Results

Continuous speech shows an alternation between relatively intense vowels and weaker consonants. These alternations result in pronounced modulations of the amplitude envelope of speech near 3-4 Hz. Degradation of these modulations by noise or reverberation correlates with decreased speech intelligibility. We have begun to study the neural encoding of speech modulations by measuring modulation transfer functions (MTF) of ANFs in anesthetized cats. A functional model of ANFs incorporating tuning, compression and the MTF can simulate the envelope of ANF responses to both reverberant and non-reverberant speech utterances.

3.1 Modulation Transfer Functions of Auditory Nerve Fibers

The neural MTF relates (as a function of modulation frequency) the modulation index of an AM stimulus to that of the period histogram time-locked to the modulation cycle. We measured MTFs at modulation frequencies from 1 to 1500 Hz, using band-limited noise with sinusoidally modulated intensity. These MTFs were typically bandpass. Lower cutoffs were in the range from 2-25 Hz. Upper cutoffs were between 200-800 Hz, depending on CF and consistent with Joris and Yin (1991). MTF phase was nearly linear ($R^2 > 0.99$ typical), with slopes in the range from 2 to 8 ms depending on CF and roughly consistent with the group delay of ANF phase-locked responses to pure tones (Goldstein, Baer et al., 1971). The MTF step response, derived from the inverse discrete Fourier transform of the MTF magnitude and phase, describes the neural response to a step change in input envelope, as at the onset of a tone-burst (Yates, 1987). MTF step responses were consistent with the rapid and short-term adaptation characteristics of ANFs. MTFs had level-dependent behavior that correlated with fiber spontaneous rate (SR). Figure 21 and Figure 22 summarize the results for the different SR groups.

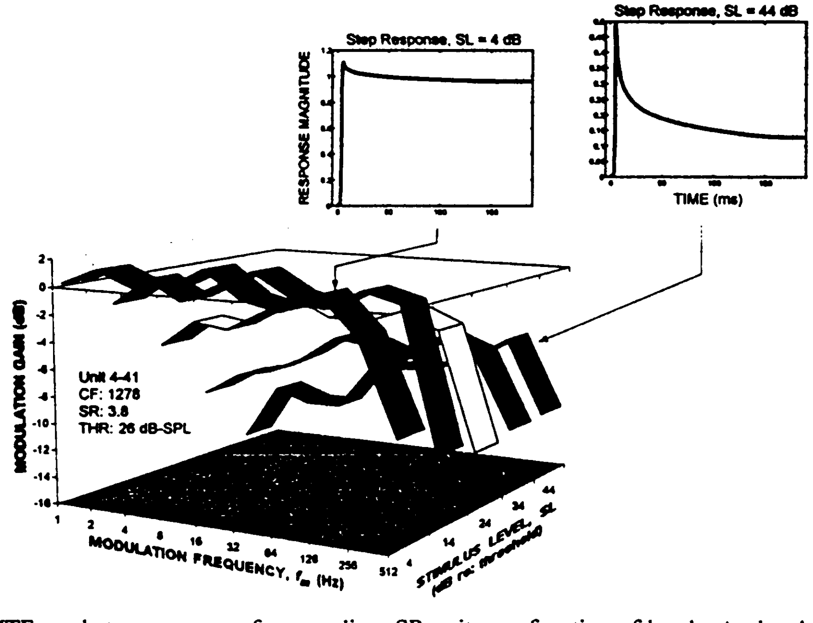

Figure 21. MTFs and step responses for a medium SR unit as a function of level. At threshold levels, the MTF is purely lowpass, and the MTF step response shows little adaptation. With increasing stimulus level, MTFs become increasingly bandpass, indicating a decrease in the ability to follow low frequency modulations, and the MTF step response shows increased rapid and short-term adaptation.

III. Preliminary Results

Continuous speech shows an alternation between relatively intense vowels and weaker consonants. These alternations result in pronounced modulations of the amplitude envelope of speech near 3-4 Hz. Degradation of these modulations by noise or reverberation correlates with decreased speech intelligibility. We have begun to study the neural encoding of speech modulations by measuring modulation transfer functions (MTF) of ANFs in anesthetized cats. A functional model of ANFs incorporating tuning, compression and the MTF can simulate the envelope of ANF responses to both reverberant and non-reverberant speech utterances.

3.1 Modulation Transfer Functions of Auditory Nerve Fibers

The neural MTF relates (as a function of modulation frequency) the modulation index of an AM stimulus to that of the period histogram time-locked to the modulation cycle. We measured MTFs at modulation frequencies from 1 to 1500 Hz, using band-limited noise with sinusoidally modulated intensity. These MTFs were typically bandpass. Lower cutoffs were in the range from 2-25 Hz. Upper cutoffs were between 200-800 Hz, depending on CF and consistent with Joris and Yin (1991). MTF phase was nearly linear ($R^2 > 0.99$ typical), with slopes in the range from 2 to 8 ms depending on CF and roughly consistent with the group delay of ANF phase-locked responses to pure tones (Goldstein, Baer et al., 1971). The MTF step response, derived from the inverse discrete Fourier transform of the MTF magnitude and phase, describes the neural response to a step change in input envelope, as at the onset of a tone-burst (Yates, 1987). MTF step responses were consistent with the rapid and short-term adaptation characteristics of ANFs. MTFs had level-dependent behavior that correlated with fiber spontaneous rate (SR). Figure 21 and Figure 22 summarize the results for the different SR groups.

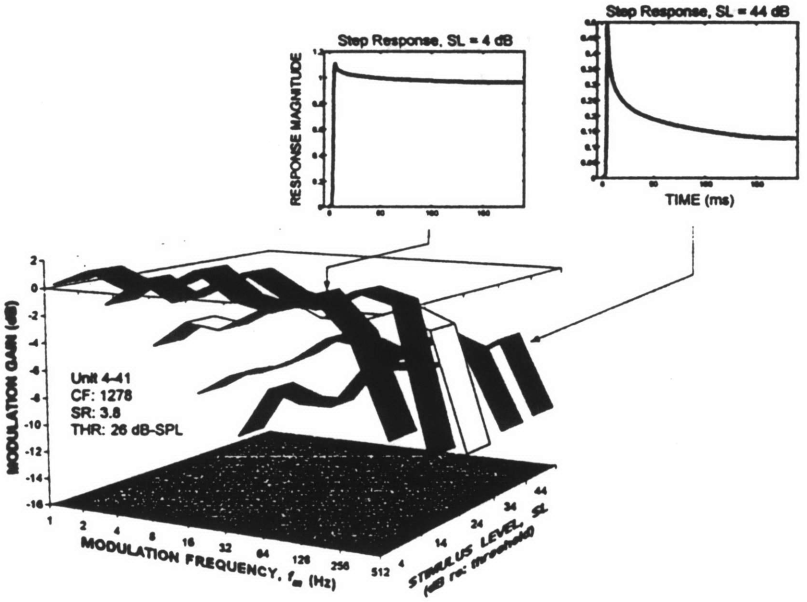

Figure 21. MTFs and step responses for a medium SR unit as a function of level. At threshold levels, the MTF is purely lowpass, and the MTF step response shows little adaptation. With increasing stimulus level, MTFs become increasingly bandpass, indicating a decrease in the ability to follow low frequency modulations, and the MTF step response shows increased rapid and short-term adaptation.

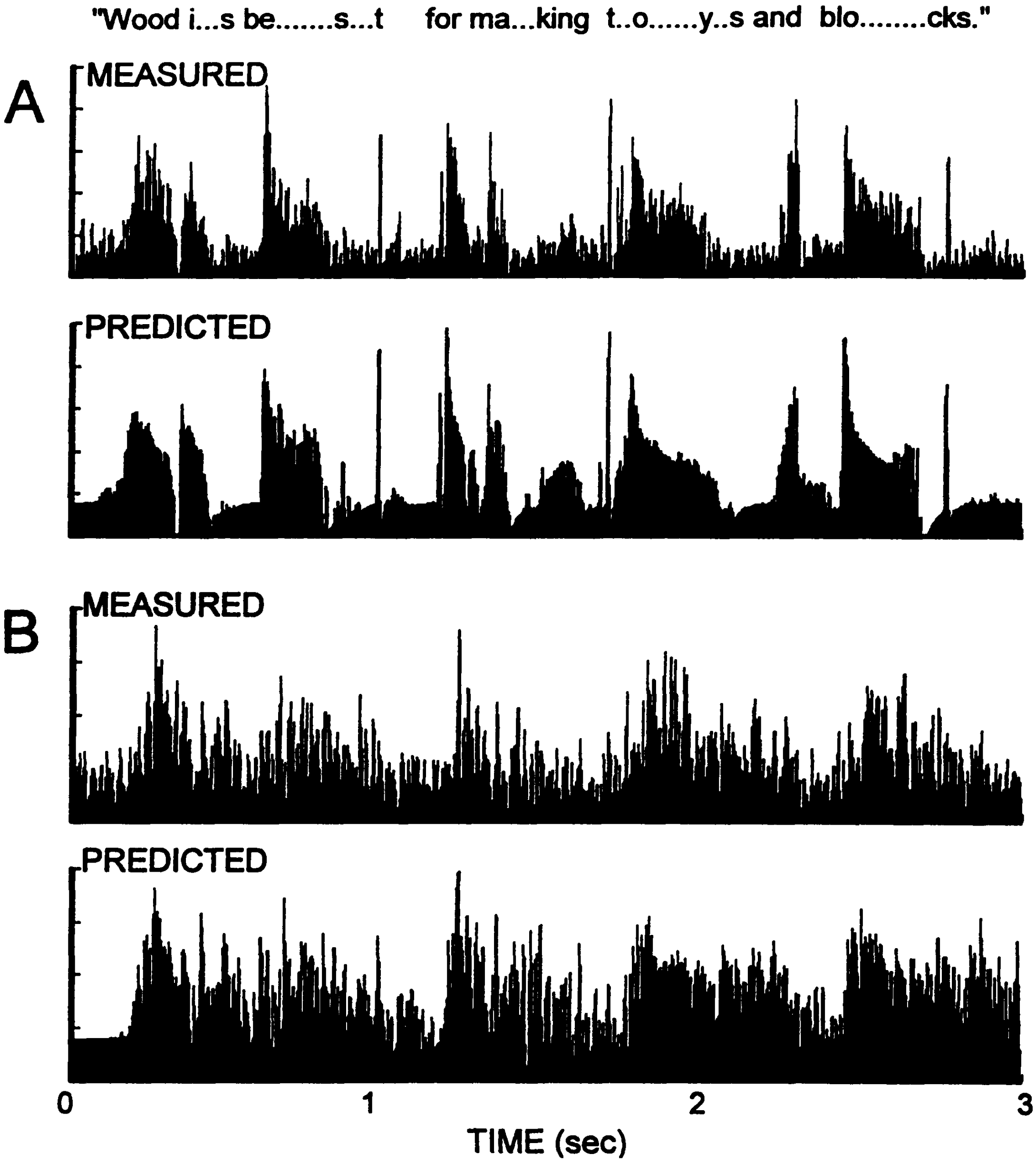

Figure 23. A. Measured and predicted PSTH responses to non-reverberant sentence. B. Measured and predicted PSTH responses to reverberant sentence (RT_{60} = 1.5 sec.). CF 1150 Hz; Threshold 18 dB-SPL; SR 75 sp/sec.

IV. Summary

In 1906, Wallace Clement Sabine was appointed Dean of the new Graduate School of Applied Science at Harvard University. He served at this post for nine years, until the position dissolved in a short-lived merger between Harvard and the Massachusetts Institute of Technology (1915 - 1917) (Hunt, 1964 p. xv). It is, therefore, fitting that the research presented in this proposal, which owes so much of its theoretical groundwork to Sabine's work in room acoustics, should be carried out under the auspices of the Harvard – M.I.T. Division of Health Sciences and Technology, a modern collaboration of the two institutions that is attempting to make good the previous failed effort that cost Sabine his deanship.

The goal of the proposed research is to bring about another type of merger, that between two fields. The fields of auditory physiology and architectural acoustics have evolved to the point where we may now ask specific questions about how speech coding is influenced by the natural acoustical conditions in which communication takes place. Experiment I will quantify the effects of reverberation both on "rate-place" and "temporal" representations of the speech signal and will attempt to define the conditions under which these representations may be lost beyond recovery by central processing schemes. Experiment II will study a central processing area devoted to localization in an attempt to define the conditions under which late-arriving sound energy in a reverberant field can be assigned to an unambiguous location, precisely those conditions that have been avoided by acoustics consultants ever since Sabine.

The motivating subject in this thesis is speech intelligibility, and the knowledge obtained from the proposed research may have practical applications in fields such as audiology and speech recognition. It is expected that the results of this study will also be of general interest to those working in the fields of architectural acoustics, music perception and auditory scene analysis, as the acoustical factors that affect speech intelligibility have a profound influence on music and may also affect auditory stream segregation. Finally, it is not unreasonable to expect that by studying the mechanisms underlying the degradation of speech intelligibility, we may come to a better understanding of the physiological basis of speech perception.

Appendix A: Proposed Schedule

The research described in this proposal will be carried out over a period of approximately three years, beginning in September, 1996, and commencing in August, 1999. All experimental work and data analysis will be carried out at the Eaton-Peabody Laboratory at the Massachusetts Eye and Ear Infirmary. Based on previous work in the auditory-nerve and ICC of the cat, approximately 45 animals will be required for sufficient data to answer the proposed experimental questions. The three year time estimate is based largely on the expected man-hours to analyze experimental data from this number of animals (see Appendix C: Vertebrate Animals, p. 33). The electrophysiology for each experimental protocol will be performed on an ongoing basis during the initial year so that preliminary results may obtained and data analysis routines developed. In general, however, the experiments will be carried out roughly in the order in which they have been presented, with Experiment I occurring in the first 1.5 years and Experiment II occurring in the second 1.5 years. Preliminary model development for Experiment I-a will take place during the summer of 1996.

Appendix B: Resources

Budget

Table 1 lists the itemized budget for the proposed three year period of research. All costs have been increased at a rate of 4% per year to cover inflation. The total cost for the entire three year period is estimated at $122,089. Anticipated funding sources for this cost include additional fellowship supplements from the Harvard – M.I.T. Division of Health Sciences and Technology and Speech and Hearing Sciences Program and the Eaton-Peabody Lab NIH Project Grant, Project 5.

Budget Item	**Cost**		
	FY 97	**FY 98**	**FY 99**
Personnel	$24,920	$25,472	$39,170
Supplies	$6,482	$8,482	$8,822
Travel	$1,600	$1,664	$1,731
Other Expenses	$1,200	$1,248	$1,298
Total Cost	$34,202	$36,866	$51,020

Table 1. Three year itemized budget. FY: fiscal year from September to August.

Budget Justification

Personnel The largest components of this budget item are the cost of M.I.T. tuition, which is expected to average $33,000 per year during the next three years, and a cost-of-living stipend, which is approximately $16,000 per year. M.I.T. does not permit students who are working solely on their thesis to waive any portion of tuition costs. Thus, the budget includes the entire cost of tuition for all three years. For two of the three years (FY 97 and FY 98), an NIH Training Grant will offset the tuition and stipend costs by approximately $25,000 per year. For an additional semester, a teaching assistance fellowship for HST 718 (Acoustics of Speech and Hearing) will offset the cost of tuition and stipend (the effect of this fellowship is included in FY 99, although the actual T.A. appointment will occur in FY 97).

Supplies include general lab and office supplies, histology supplies for FY 98 and FY 99 to cover the cost of post-experiment ICC histology to verify electrode positions, and the cost of 15 cats per year plus cat care (see Appendix C: Vertebrate Animals, p. 33).

Travel includes the cost of attending two conferences per year to present work completed on this research project. The highest priority conferences for this project are those of the Acoustical Society of America and the Association for Research in Otolaryngology.

Other Expenses includes estimated publication costs and the cost of purchasing and maintaining computer software programs required for stimulus generation and data analysis. Such programs might include Matlab statistics and optimization toolboxes, the Matlab C compiler, a UNIX-based Klatt synthesis program, and a room acoustics modeling system such as Renkus-Heinz EASE or Bose Modeler.

Appendix C: Vertebrate Animals

This section is copied with permission from the 1994 RO1 grant application entitled "Neural Coding of Speech" by Bertrand Delgutte.

1. Healthy adult cats of either sex and free of middle-ear infection will be used in this series of experiments. About 15 animals will be used each year. Briefly, the surgical procedures are as follows. After induction of anesthesia by Dial in urethane, a tracheal canula is inserted, skin and muscles overlying the back of the skull are reflected. Ear canals are severed for insertion of closed acoustic systems. Tympanic bullae are opened to allow recording from the round-window area. The skull overlying the cerebellum is removed to allow for cerebellar retraction in the case of auditory-nerve experiments or aspiration for direct viewing of the inferior colliculus. The animal is placed in a Horsley-Clark stereotaxic apparatus, and single units are recorded with microelectrodes for durations of 24-48 hours.
2. Study of the neural coding of speech requires a living preparation. The cat is chosen as the animal model for four reasons:
   - The large number of stimulus conditions that need to be studied in these experiments requires stable preparations over long periods of time, and cats are much better than rodents for such experiments.
   - A great deal of knowledge is already available on the anatomy and physiology of brainstem auditory neurons for the cat.
   - Cats hear better than most rodents at low frequencies, which convey the most important information for speech discrimination.
   - Cats can be trained to discriminate speech sounds (Dewson, 1964). Such information is not available for most rodents.

   The number of animals requested (15/year) is based on estimates of the total number of experiments that can be thoroughly analyzed given the allotted time and manpower. Our experimental procedures are highly computerized, allowing a maximal amount of data to be collected from each animal.
3. The animals will be housed in the USDA-approved animal care facility of the Massachusetts Eye and Ear Infirmary. This facility is under the supervision of a veterinarian who is consulted about animal health and experimental protocols.
4. Prior to any surgical procedures, animals will be anesthetized by intraperitoneal injections of Dial in urethane (75 mg/kg). Animals remain anesthetized at all times until euthanasia. Throughout the procedures, animals are administered anesthetic boosters when needed as assessed by the toe-pinch reflex. Thus the animals feel no unrelieved pain at any time.
5. At the end of the experiments, the animals will be euthanized either by intracardiac injection of anesthetic overdose, or by exsanguination followed by intravascular perfusion of aldehyde fixatives while under deep anesthesia. The methods are consistent with the recommendations of the Panel of Euthanasia of the American Veterinary Medical Association.

References

Anderson, B. W. and J. T. Kalb (1987). "English verification of the STI method for estimating speech intelligibility of a communications channel." *J Acoust Soc Am* 81(6): 1982-1985.

Bergman, M. (1980). *Aging and the Perception of Speech*, University Park.

Blauert, J. (1983). *Spatial Hearing*. Cambridge, MA, MIT Press.

Bolt, R. H. and A. D. MacDonald (1949). "Theory of speech masking by reverberation." *J Acoust Soc Am* 21: 577-580.

Bradley, J. S. (1986a). "Predictors of speech intelligibility in rooms." *J Acoust Soc Am* **80**(3): 837-845.

Bradley, J. S. (1986b). "Speech intelligibility studies in classrooms." *J Acoust Soc Am* **80**(3): 846-854.

de Boer, E. (1967). "Correlation studies applied to the frequency resolution of the cochlea." *J Aud Res* 7: 209-217.

Delgutte, B. (1980). "Representations of spech-like sounds in the discharge patterns of auditory-nerve fibers." *J Acoust Soc Am* **68**(3): 843-857.

Delgutte, B. (1986). Analysis of French stop consonants with a model of the peripheral auditory system. *Invariance and Variability of Speech Processes*. J. S. Perkell and D. H. Klatt. Hillsdale, NJ, Erlbaum: 163-177.

Delgutte, B. (in press). Auditory neural processing of speech. *The Handbook of Phonetic Sciences*. W. J. Hardcastle and J. Laver. Oxford, Blackwell.

Delgutte, B. and N. Y. S. Kiang (1984a). "Speech coding in the auditory nerve: I. Vowel-like sounds." *J Acoust Soc Am* **75**(3): 866-878.

Delgutte, B. and N. Y. S. Kiang (1984b). "Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds." *J Acoust Soc Am* **75**(3): 879-886.

Delgutte, B. and N. Y. S. Kiang (1984c). "Speech coding in the auditory nerve: III. Voiceless fricative consonants." *J Acoust Soc Am* **75**(3): 887-896.

Delgutte, B. and N. Y. S. Kiang (1984d). "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics." *J Acoust Soc Am* **75**(3): 897-907.

Delgutte, B. and N. Y. S. Kiang (1984e). "Speech coding in the auditory nerve: V. Vowels in background noise." *J Acoust Soc Am* **75**(3): 908-918.

Dewson, J. H. I. (1964). "Speech sound discrimination by cats." *Science* **144**: 555-556.

Durlach, N. I. and H. S. Colburn (1978). Binaural Phenomenon. *Handbook of Perception*. E. C. Carterette and M. P. Friedman. New York, Academic Press, Inc. **Volume IV**: 365-466.

Fay, R. D. (1936). "A method of obtaining natural directional effects in a public address system." *J Acoust Soc Am* 7: 131-132.

Georgopoulos, A. P., M. Taira, et al. (1993). "Cognitive neurophysiology of the motor cortex." *Science* **260**: 47-52.

Goldstein, J. L., T. Baer, et al. (1971). A theoretical treatment of latency, group delay, and tuning characteristics for auditory-nerve responses to clicks and tones. *Physiology of the Auditory System*. M. B. Sachs. Baltimore, National Educational Consultants: 133-156.

Greenberg, S. (1996). Auditory Processing of Speech. *Principles of Experimental Phonetics*. N. J. Lass. St. Louis, MO, Mosby: 362-407.

Haas, H. (1951). "Uber den Einfluss eines Einfachechoes auf die Horsamkeit von Sprache [The influence of a single echo on the audibiliity of speech]." *Acustica* 1: 49-58.

Hall, W. M. (1936). "A method for maintaining in a public address system the illusion that the sound comes from the speaker's mouth." *J Acoust Soc Am* 7: 239.

Harris, G. G., J. L. Flanagan, et al. (1963). "Binaural interaction of a click with a click pair." *J Acoust Soc Am* 35(5): 672-678.

Helfer, K. S. (1994). "Binaural cues and consonant perception in reverberation and noise." *J Speech Hear Res* 37: 429-438.

Helfer, K. S. and R. A. Huntley (1991). "Aging and consonant errors in reverberation and noise." *J Acoust Soc Am* 90(4): 1786-1796.

Houtgast, T. and H. J. M. Steeneken (1972). *Envelope spectrum and intelligibility of speech in enclosures.* Conference of Speech Communication and Processing, Newton, MA, IEEE-AFCRL.

Houtgast, T. and H. J. M. Steeneken (1973). "The modulation transfer function in room acoustics as a predictor of speech intelligibility." *Acustica* 28: 66-73.

Houtgast, T. and H. J. M. Steeneken (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics." *Acustica* 46(1): 60-72.

Houtgast, T. and H. J. M. Steeneken (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria." *J Acoust Soc Am* 77(3): 1069-1077.

Humes, L. E. (1993). Some temporal factors affecting speech recognition. *Acoustical Factors Affecting Hearing Aids.* G. A. Studebaker and I. Hochberg. Needham Heights, MA, Allyn and Bacon: 205-220.

Humes, L. E., S. Boney, et al. (1987). "Further validation of the Speech Transmission Index (STI)." *J Speech Hear Res* 30(3): 403-410.

Humes, L. E., D. D. Dirks, et al. (1986). "Application of the articulation index and the speech transmissino index to the recognition of speech by normal-hearing and hearing-impaired listeners." *J Speech Hear Res* 29: 447-462.

Hunt, F. V. (1964). Preface. *Collected Papers on Acoustics.* W. C. Sabine. New York, Dover Publications, Inc.

Jacob, K. (1989). "Correlation of speech intelligibility tests in reverberant rooms with three predictive algorithms." *J Audio Eng Soc* 37: 1020-1030.

Joris, P. X. and T. C. T. Yin (1991). "Responses to amplitude-modulated tones in the auditory nerve of the cat." *J Acoust Soc Am* 91(1): 215-232.

Kaltenbach, J. A., R. J. Màleca, et al. (1993). "Forward maskin properties of neurons in the dorsal cochlear nucleus: Possible role in the process of echo supression." *Hear Res* 67: 35-44.

Kiang, N. Y. S. and E. C. Moxon (1974). "Tails of tuning curves of auditory-nerve fibers." *J Acoust Soc Am* 55(3): 620-630.

Kiang, N. Y. S., T. Watanabe, et al. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Neve.* Cambridge, MA, MIT Press.

Knudsen, V. O. (1929). "The hearing of speech in auditoriums." *J Acoust Soc Am* 1: 56-82.

Kuttruff, H. (1991). *Room Acoustics.* Amsterdam, Elsevier Applied Science.

Latham, H. G. (1979). "The signal-to-noise ratio for speech intelligibility - an auditorium acoustics design index." *Appl Acoust* 12: 253-320.

Lochner, J. P. A. and J. F. Burger (1958). "The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech." *Acustica* 8(1): 1-10.

Lochner, J. P. A. and J. F. Burger (1964). "The influence of reflections on auditorium acoustics." *J Sound Vib* 1(4): 426-454.

Meyer, E. and G. R. Schodder (1952). "On the influence of reflected sound on directional localization and loudness of speech." *Nachr Akad Wiss Gottingen, Math Phys Klasse IIa* **6**: 31-42.

Miller, M. I. and M. B. Sachs (1983). "Representation of stop consonants in the discharge patterns of auditory-nerve fibers." *J Acoust Soc Am* **74**(2): 502-517.

Müller, M. and D. Robertson (1991). "Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibers in the guinea pig." *Hear Res* **57**: 63-70.

Nabelek, A. (1994). "Cues for perception of the diphthong /ai/ in either noise or reverberation. Part I. Duration of the transistion." *J Acoust Soc Am* **95**(5): 2681-2693.

Nabelek, A. and A. M. Donahue (1984). "Perception of consonants in reverberation by native and non-native listeners." *J Acoust Soc Am* **75**(2): 632-634.

Nabelek, A., A. Ovchinnikov, et al. (in press). "Cues for perception of synthetic and natural diphthongs in either noise or reverberation." *in press*.

Nabelek, A. K., Z. Czyzewski, et al. (1993). "Vowel boundaries for steady-state and linear formant trajectories." *J Acoust Soc Am* **94**(2): 675-687.

Nabelek, A. K. and T. R. Letowski (1988). "Similarities of vowels in nonreverberant and reverberant fields." *J Acoust Soc Am* **83**(5): 1891-1899.

Nabelek, A. K., T. R. Letowski, et al. (1989). "Reverberant overlap and self-masking in consonant identification." *J Acoust Soc Am* **86**: 1259-1265.

Nabelek, A. K. and P. K. Robinson (1982). "Monaural and binaural speech perception in reverberation for listeners of various ages." *J Acoust Soc Am* **71**: 1242-1248.

Nakijama, T. and Y. Ando (1991). "Effects of a single reflection with varied horizontal angle and time delay on speech intelligibility." *J Acoust Soc Am* **90**(6): 3173-3179.

Payton, K. L., R. M. Uchanski, et al. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing." *J Acoust Soc Am* **95**(3): 1581-1592.

Plomp, R., H. J. M. Steeneken, et al. (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. II. Mirror image computer model applied to rectangular rooms." *Acustica* **46**(1): 73-81.

Reichart, W., O. Abdel Alim, et al. (1974). *Appl Acoust* **7**: 243.

Relkin, E. M. and J. R. Doucet (1991). "Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons." *Hear Res* **55**: 215-222.

Relkin, E. M. and D. G. Pelli (1987). "Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures." *J Acoust Soc Am* **82**: 1679-1691.

Relkin, E. M. and C. W. Turner (1988). "A reexamination of forward masking in the auditory nerve." *J Acoust Soc Am* **84**(2): 584-591.

Rietschote, H. F. v. and T. Houtgast (1983). "Predicting speech intelligibility in rooms from the modulation transfer function. V: The merits of the ray-tracing model versus general room acoustics." *Acustica* **53**: 72-78.

Rietschote, H. F. v., T. Houtgast, et al. (1981). "Predicting speech intelligibility in rooms. IV: A ray-tracing computer model." *Acustica* **49**(3): 245-252.

Sabine, W. C. (1993). *Collected Papers on Acoustics*. Los Altos, CA, Peninsula Publishing.

Sachs, M. B. and P. J. Abbas (1974). "Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli." *J Acoust Soc Am* **55**(6): 1835-1847.

Sachs, M. B., R. L. Winslow, et al. (1988). Representation of speech in the auditory periphery. *Auditory function: Neurobiological bases of hearing*. G. M. Edelman, W. E. Gall and W. M. Cowan. New York, Wiley: 747-774.

Sachs, M. B. and E. D. Young (1979). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate." *J Acoust Soc Am* **66**(2): 470-479.

Sachs, M. B. and E. D. Young (1980). "Effects of nonlinearities on speech encoding in the auditory nerve." *J Acoust Soc Am* **68**(3): 858-875.

Santon, F. (1976). "Numerical prediction of echograms and of the intelligibility of speech in rooms." *J Acoust Soc Am* **59**(6): 1399-1405.

Schmidt-Nielsen, A. (1987). "Comments on the use of physical measures to assess speech intelligibility." *J Acoust Soc Am* **81**(6): 1985-1987.

Schroeder, M. R. (1981). "Modulation transfer functions: definition and measurement." *Acustica* **49**: 179-182.

Sinex, D. G. and C. D. Geisler (1983). "Responses of auditory-nerve fibers to consonant-vowel syllables." *J Acoust Soc Am* **73**(2): 602-615.

Smith, R. L. (1977). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects." *J Neurophys* **40**(5): 1098-1112.

Smith, R. L. and J. J. Zwislocki (1975). "Short-term adaptation and incremental responses of single auditory-nerve fibers." *Biol Cybernetics* **17**: 169-182.

Snow, W. B. (1936) Sound Reproducing System. U. S. Patent.

Takata, Y. and A. K. Nabelek (1990). "English consonant recognition in noise and in reverberation by Japanese and American listeners." *J Acoust Soc Am* **88**(2): 663-666.

Thiele, R. (1953). *Acustica* **3**: 291.

Vitruvius, M. (1st cent. BC). *De Architectura.*

Wallach, H., E. B. Newman, et al. (1949). "The precedence effect in sound localization." *Am J Psychol* **52**(3): 315-336.

Wattel, E., R. Plomp, et al. (1981). "Predicting speech intelligibility in rooms from the modulation transfer function. III: Mirror image computer model applied to pyramidal rooms." *Acustica* **48**(5): 320-324.

Wickesberg, R. E. and D. Oertel (1990). "Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression." *J Neurosci* **10**(6): 1762-1768.

Yates, G. K. (1987). "Dynamic effects in the input/output relationship of auditory-nerve." *Hear Res* **27**: 221-230.

Yin, T. C. T. (1994). "Physiological correlates of the precedence effect and summing localization in the inferior colliculus of thecat." *J Neurosci* **14**(9): 5170-5186.

Yin, T. C. T. and J. C. K. Chan (1988). Neural mechanisms underlying interaural time sensitivity to tones and noise. *Auditory Function: Neurobiological Bases of Hearing*. G. M. Edelman, W. E. Gall and W. M. Cowan. New York, Wiley: 385-430.

Young, E. D. and M. B. Sachs (1979). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers." *J Acoust Soc Am* **66**(5): 1381-1403.

Zurek, P. M. (1979). "Measurements of binaural echo suppression." *J Acoust Soc Am* **66**(6): 1750-1757.

Zurek, P. M. (1980). "The precedence effect and its possible role in the avoidance of interaural ambiguities." *J Acoust Soc Am* **67**: 952-964.

Zurek, P. M. (1987). The precedence effect. *Directional Hearing*. W. A. Yost and G. Gourevitch. New York, Springer-Verlag: 85-105.

Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. *Acoustical Factors Affecting Hearing Aid Performance*. G. A. Studebaker and I. Hochberg. Needham Heights, MA, Allyn and Bacon: 255-276.