Harvey Fletcher's role in the creation of communication acoustics[a)]

Jont B. Allen
Acoustics Research Department, AT&T Laboratories, Murray Hill, New Jersey 07974

(Received 15 June 1995; accepted for publication 29 January 1996)

The major contributions made by Harvey Fletcher and his colleagues to communication acoustics are summarized along with a brief account of the acoustical career of the first President of the Acoustical Society of America.

PACS numbers: 43.10.Ln, 43.05.Dr

The processes of speaking and hearing are very intimately related, so much so that I have often said that, We speak with our ears. We can listen without speaking but cannot speak without listening. People who are born without hearing learn to talk only with the greatest difficulty, and not one of them has yet succeeded in producing what most of us would call normal speech.—From the Introduction to Fletcher's 1953 book.

I. HISTORICAL PERSPECTIVE

A. A Brief Personal History

Harvey Fletcher was born on 11 September, 1884 in Provo, Utah, and died on 23 July, 1981 at the age of 96. He came from a Mormon background, and held strong religious beliefs throughout his life. Fletcher's father became profoundly deaf over a 3-day period at the age of 55, and remained deaf for 25 years until his death (Fletcher, 1929b). This deafness would help trigger young Harvey's career.

Just before Fletcher was to graduate from the University of Chicago, Dr. Frank Jewett, vice president of the Western Electric Company (who was also educated at the University of Chicago, and who came to AT&T as a professor from the Massachusetts Institute of Technology), asked his friend Dr. Robert Millikan, a Professor of Physics at the University of Chicago, to recommend a man trained in the ''new electron physics'' who could conduct research in electronic amplification. AT&T had been struggling with the difficult problem of network quality, and had come to realize they needed to harness the electron to amplify the weak network signals. Millikan recommended Dr. Fletcher, who was the first to convincingly demonstrate the existence of the electron by measuring its quantized charge while doing work towards the Ph.D. degree. In 1911 Fletcher received his Ph.D. from the University of Chicago as its first *summa cum laude*. He received a job offer from the Western Electric Company, but felt an obligation to return to his alma mater, Brigham Young University, where he served as Chairman of the Physics Department.

When Fletcher turned down the job at Western Electric, the position was offered to Dr. H. D. Arnold, a graduate student also working under Millikan on electron studies. Arnold went on to develop the vacuum tube (starting from the audion which had little or no vacuum) for the telephone network during 1913 and 1914.[1]

Fletcher's solid background in electron physics gave him a deep understanding of physical principles. His Mormon background made him want to focus his career on a worthy lifetime pursuit. The ''oil-drop'' experience greatly increased his self-confidence. The first electronic vacuum tube amplifiers developed by Arnold were introduced into the network on 1 January, 1915. Fletcher was intensely interested in these developments, because he understood the significance of electron physics and the role that electronics would play in human communications (Fletcher, 1929a).

Fletcher realized that his full potential was not being realized at Brigham Young University, and Jewett's offer to Fletcher to join AT&T was being renewed in January of every year. Thus in 1916, with this strong yearly encouragement from Jewett, Fletcher came to Western Electric Research Laboratories (Bell Telephone Laboratories beginning on 1 January 1925) and worked with I.B. Crandall, who was his acoustics mentor and colleague.

In 1918, during the first world war, Fletcher and Crandall independently developed acoustics research programs. After the war they compared notes and integrated their programs into one (Fletcher, 1992). Crandall passed away on 22 April, 1927, just before the second printing of his book on speech (Crandall, 1926), and Fletcher took over Crandall's work of improving intelligibility and quality in the telephone network.

In 1928, following the death of Crandall, Fletcher was promoted to *Director of Acoustical Research* in acknowledgment of his accomplishments, heading a new department formed to handle the increasingly important speech and hearing problems of the Bell System (Fletcher, 1992). During his 33-year career at AT&T Bell Labs (1916–1949), Fletcher introduced many seminal concepts about human auditory perception, which have defined the framework of contemporary hearing research and played an important role in many new technological developments. He was the first to accu-

[a)]Editor's Note: This is a slightly abridged version of the introductory chapter written by Jont B. Allen, serving as the editor of the Acoustical Society's reprinting of Harvey Fletcher's 1953 book *Speech and Hearing in Communication*. The section on the critical band was presented at the spring meeting of the ASA, Washington, DC, 1995. [J. Acoust. Soc. Am. **97**, 3378–3380 (A) (1995)], and has not been formally peer reviewed. The complete book is available from the Society (see the publication list and order form in this issue).

 0001-4966/96/99(4)/1825/15/$6.00

rately measure the auditory threshold, the first to measure the loudness-growth relationship now frequently called *Stevens' law*, and to identify the concepts of *loudness additivity* and *articulation additivity* (which later became the *articulation index* or AI). He helped develop the vacuum tube hearing aid, the commercial audiometer, the artificial larynx, and stereophonic sound.

His first book *Speech and Hearing*, published in 1929, (Fletcher, 1929a), brought him world acclaim. In the introduction to the 1929 edition of Fletcher's book, H. D. Arnold, then *Director of Research*, wrote

> The work that Doctor Fletcher discusses drew at the start on all the acoustic knowledge available in the literature and during its progress every effort has been made to use to the best advantage the information found by other experimenters. For the most part, however, he describes experiments performed and conclusions reached in Bell Telephone Laboratories during investigations, captained in their early stage by Doctor Crandall and himself, for which since Doctor Crandall's death he has had the full responsibility. No one can speak with better knowledge of the facts or with more complete authority for the opinions which he expresses.

In 1929, Fletcher was a key figure in the founding of the Acoustical Society of America, and served as its first president. He was also president that year of the American Federation of Organizations of the Hard of Hearing. He wrote his first paper on the theory of hearing[2] (Fletcher, 1930b), and published an important paper on articulation (Fletcher and Steinberg, 1929).

In 1933 Fletcher was promoted to *Director of Physical Research*. In the Preface to the 1953 book (Fletcher, 1953), Fletcher wrote

> About 35 years ago the Research Laboratories of the Bell Telephone System started a comprehensive research program on speech and hearing, and its relation to the design of telephone systems. It was apparent that great advantages would come if one could describe accurately every part of the system, namely (1) the talker, (2) the microphone, (3) the electrical transmission line, (4) the telephone receiver (head phone or terminating loudspeaker), and (5) the listener.

Today goals (2)–(4) have been well researched. This leaves (1) the talker and (5) the listener. It is here that Fletcher placed his effort, and it is what we read about in his book, which is devoted to the investigations of *speech* (Chapters 1–5), *hearing theory* (Chapters 7 and 14), *hearing acuity and loudness* (Chapters 8–12), and *measures of human speech recognition* (Chapters 15–20). AT&T's job was to build the best telephone system in the world, coast to coast! Understanding how speech was degraded by the network was an important piece of that problem.

Fletcher expected a great deal from the staff. The two who worked most closely with Fletcher were Munson and Steinberg. Many of the people in his group were physicists. A list of scientists working on speech and hearing at Bell Labs over this period includes Arnold, Biddulph, Crandall, Dunn, French, Fry, Galt, Gardner, Graham, Hartley, Kingsbury, Koenig, Kranz, Lane, MacKenzie, Munson, Riesz, Sacia, Shower, Sivian, Steinberg, J.Q. Stewart, Wegel, and Wente.

B. A brief technical history

In 1920 Fletcher and others in his group began by determining the absolute threshold of hearing which was only crudely known at that time (Fletcher, 1920; Fletcher and Wegel, 1922a, 1923b; Fletcher, 1922c; Fletcher, 1929a). These measurements required the development of an acoustical standard for pressure generation and measurement. A summary of their results is shown in Fig. 1 as the solid curve. Arnold and Crandall, and later Wente, developed the thermophone for this purpose Fletcher, 1929a, p. 134) and Wente the condenser microphone (Hunt, 1982). With these new instruments in hand they began making noise measurements on very noisy New York subway trains for the city (Fletcher, 1923a). Also in 1922 Fletcher started his work on audiometers, instruments for testing hearing thresholds over the entire population (Fletcher, 1922a).

In 1926 Fletcher analyzed the problem of ''telephone howl,'' namely positive feedback in an electroacoustic system (Fletcher, 1926; Millman, 1984). This seems to be the first time that network feedback had been mathematically analyzed. One year later Black proposed negative feedback to stabilize electronic amplifier network gain, and in 1932 Nyquist published his now famous test of network stability.

Speech quality. Starting in 1920 the main focus of Fletcher's work was to look at methods to solve the serious problem of evaluating speech quality in the telephone network. Among the tools Fletcher and his group developed for this task were procedures for measuring the *articulation* and the *loudness* of speech. In 1921 and 1922 Fletcher wrote several internal memoranda on *Telephone Quality* (Fletcher, 1921; Fletcher, 1922e) and on *Speech Transmission* (Fletcher, 1922b; Fletcher, 1922a). These papers were concerned with the problem of improving the quality and articulation of speech in the telephone network.

Articulation as defined by Fletcher is the probability of correctly transcribing speech phonemes and syllables modified by a telephone channel. This channel was specified in terms of its frequency response and noise spectrum. As used here, articulation is a joint property of the talker, the channel, and the listener. Under ideal conditions, the average articulation is about 98.5%. *Intelligibility* is the recognition with meaning, such as in words and sentences. Articulation does not require that any meaning be transmitted. One may achieve a high articulation score, for a language one does not know, with a zero intelligibility score (Allen, 1994).

Loudness, in *sones*, is the name given to the perceptual intensity and depends in a complex manner on a number of acoustical variables, such as intensity, frequency, spectral bandwidth, and on the temporal properties of the stimulus, as well as on the mode of listening (i.e., in quiet or in noise, binaural or monaural stimulation). The first loudness work at Bell Labs was done by MacKenzie, who worked with Cran-

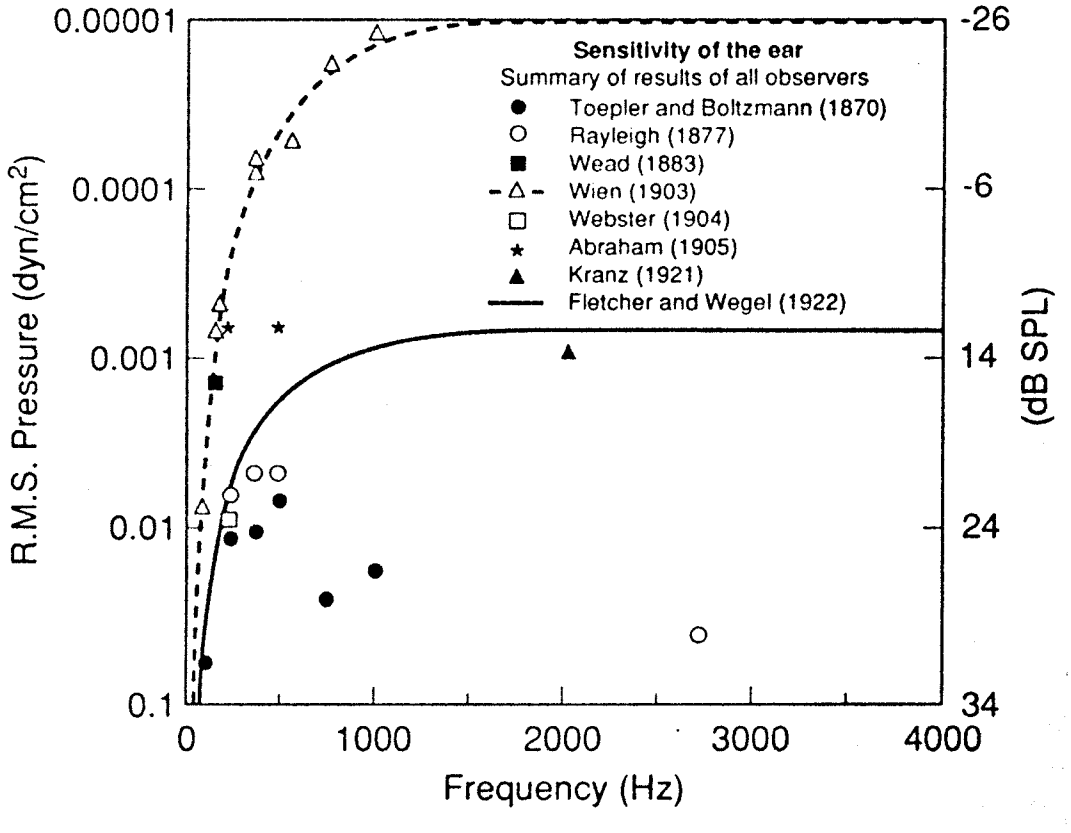

FIG. 1. This figure is redrawn from Fletcher and Wegel (1922a), Fletcher and Wegel (1922b), and Fagen (1975) and reviews all the hearing threshold data available at that time. The solid line shows Fletcher and Wegel's 1922 estimate of the hearing threshold. The dashed line is a summary of Wien's results. Because of the lack of calibrated acoustic sources before the development of the thermophone, there is a significant scatter in the estimated levels. The first measurements were by Boltzmann, the famous thermodynamicist. For a review of the early history of audiometry see the discussion by Sivian and White (1933), Fletcher's 1923 Ann Arbor lectures, Watson and Tolan (Watson and Tolan, 1967), Kranz (c.a., 1965), and Fletcher's 1929 book.

dall. Iso-loudness contours, which define the *phon* scale, were first determined in 1927 by Kingsbury (Kingsbury, 1927; Fletcher, 1929a, p. 227). These curves describe the relation between equally loud tones or narrow bands of noise at different frequencies. The importance of loudness to AT&T is made clear from the following quote (Fletcher, 1923b):

> In the telephone business, the commodity being delivered to the customers is reproduced speech. One of the most important qualities of this speech is its loudness, so it is reasonable to use a sensation scale to define the volume of the speech delivered.

In that 1923 paper he then goes on to show that 1 dB is the just noticeable change in intensity.

II. ARTICULATION

In 1922 Fletcher published his first paper on articulation in the *Journal of the Franklin Institute*[3] (Fletcher, 1922c). This paper was on the effects of filtering on speech articulation. Additivity of articulation was first formally suggested by Crandall in 1917. However, Crandall's formulation of additivity turned out to be incorrect. By the application of low- and high-pass filters to the speech,[4] Fletcher showed that the high and low articulation bands could be made additive only after a nonlinear transformation. Fletcher, working with real articulation data, showed in 1921 that the nonlinearly transformed articulation defined an articulation density over frequency. Integration, or summation, over this density gave the *articulation index*. This theory "formed a basis for introducing into the telephone plant [the network] engineering (about 1926) transmission penalties which took volume [level], frequency distortion and noise into account, where formerly only volume was given consideration" (Steinberg *et al.*, 1937).

After about 1919 the databases they used were formed from statistically balanced nonsense CVC, CV, and VC syllable lists, which we shall refer to as {C,V} units, where C represents a consonant and V a vowel. The syllable lists were spoken and the listeners recorded what they heard. Empirical probabilities c and v for the sound units were computed, where c is the probability of the correct identification of consonants (consonant articulation) and v is the probability of correct identification of vowels (vowel articulation). After being emperically verified as accurate, an average {C,V} unit articulation probability s was used from the composition of {C,V} units in the database to simplify computations. For example, for CVC's, $s=(2c+v)/3$, and for CV's, $s=(c+v)/2$.

Fletcher showed that the CVC syllable articulation probability S is accurately predicted from the sound unit articulation s by the relation $S=c^2v\approx s^3$. This formula suggests that the three sound units are perceived independently, and

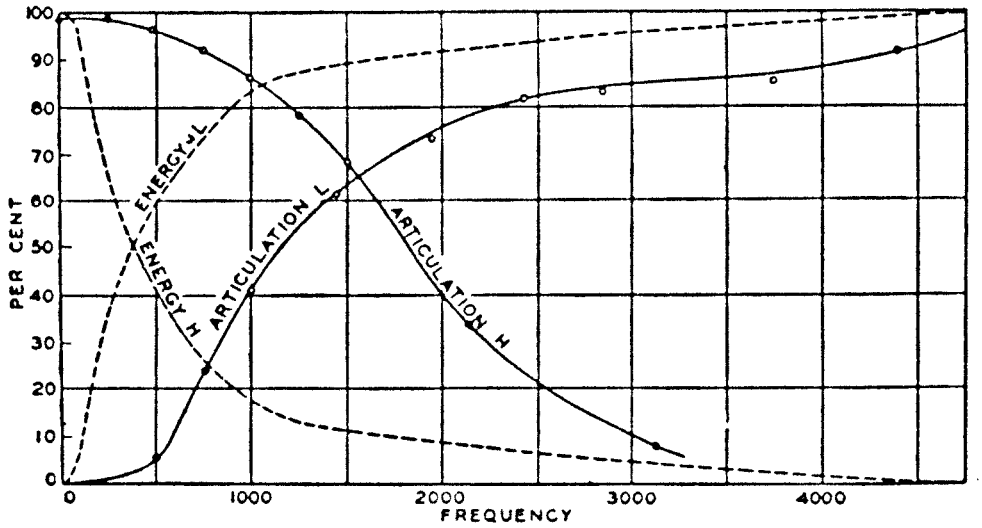

FIG. 2. This figure is from Fletcher (1929a) and Fletcher (1922c, p. 280). It shows how the articulation and the energy of speech depend on low- and high-pass filtering. Note how the energy curves for the low- and high-pass speech cross at an ordinate of 0.5. This means that the two have equal energy at the corresponding cross-over frequency of 400 Hz. The articulation curves, however, do not add to one at the cross-over point. Thus he showed that articulation does not sum. Based on these data, Fletcher proposed a nonlinear transformation of the articulation bands that sums. This transformation was called the articulation index.

that to correctly identify the syllable, all three sound-units must be correctly identified. For example, suppose the CVC is *wif*. If a listener responds with *mif* that would be one error of the first C sound unit and the syllable would be scored *incorrect*. As discussed in the 1953 book (Fig. 175, page 285) this formula for S is surprisingly accurate. It implies that speech coarticulations are decoded into C and V units before the meaning is extracted, early in the auditory nervous system (Allen, 1994; Allen, 1996).

A. The articulation index

By 1921 Fletcher had shown that the sound-unit articulation s may be accurately predicted from the channel frequency response and the channel noise (Fletcher and Galt, 1950). He did this by introducing the concept of *articulation additivity*, which in turn led to the *Fletcher–Stewart multichannel model of speech recognition* and the *articulation index*.

By 1921 Fletcher had low- and high-pass filtered the speech sounds with sharp filters. We designate these articulations of the low- and high-pass filtered sounds (see Fig. 2) as $s_L(f_c,\alpha)$ and $s_H(f_c,\alpha)$. The cutoff frequency of the filters is given by f_c, and the parameter α is the gain applied to the speech. By varying the speech level, the probe to masker ratio of the speech was varied. Fletcher (Fletcher, 1922c) showed that $s_L(f_c,\alpha)+s_H(f_c,\alpha)$ did not sum to $s(\alpha)$. He proposed that the articulation s could be made additive by a nonlinear transformation $\mathscr{A}(s)$, which he called the *articulation index*. He was seeking a transformation $\mathscr{A}(s)$ such that

$$\mathscr{A}(s_L(f_c,\alpha))+\mathscr{A}(s_H(f_c,\alpha))=\mathscr{A}(s(\alpha)) \tag{1}$$

for all values of f_c and α. There was no guarantee, of course, that such a transformation might exist. He proceeded by finding the cross-over frequency $f_c=f_c^*$ where

$$s_L(f_c^*,\alpha)=s_H(f_c^*,\alpha), \tag{2}$$

which is the frequency where the curves s_L and s_H cross.

For example, in Fig. 2 the two articulation curves cross at 1550 Hz where they have a value of 65%. He then argued that the two transformed articulations must be equal at f_c^*, and therefore must each be 1/2,

$$\mathscr{A}(s_L(f_c^*,\alpha))=0.5\mathscr{A}(s(\alpha)). \tag{3}$$

By repeating this procedure as a function of α, he determined $\mathscr{A}(s)$ as a function of s.

Under the special condition that the word corpus consisted of nonsense CVC's, Fletcher found that

$$\mathscr{A}(s)=\log(1-s)/\log(1-0.985), \tag{4}$$

which from Eq. (1) leads to

$$s=s_L+s_H-s_Ls_H, \tag{5}$$

or in terms of the articulation errors $e=1-s$, $e_L=1-s_L$, and $e_H=1-s_H$,

$$e=e_Le_H. \tag{6}$$

Fletcher gave an example that helps to clarify these ideas (Fletcher and Galt, 1950; Fletcher, 1953, p. 281). Suppose that the speech is filtered into a low band and a high band. When the subjects listen to the low band alone, the average articulation error is 10% ($e_L=0.1$), and when they listen to the high band alone, the error is 20% ($e_H=0.2$). Then when they listen to both bands, the error will be 2%, since $e=0.1\times0.2=0.02$. One may summarize these results [Eq. (6)] as follows: You listen to (detect) the sound features in each band independently of the other bands. For example, errors due to noise in the high band will not effect errors in the low band.

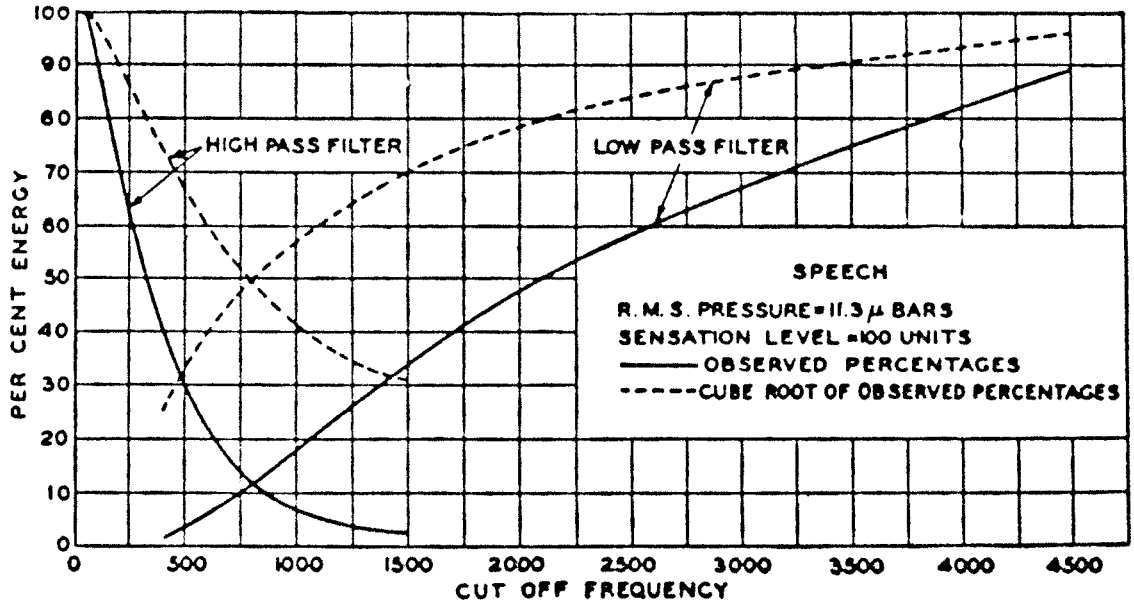

FIG. 3. This figure is from Fletcher and Steinberg (1924) and Fletcher (1929a, p. 236). It shows the effect of low- and high-pass filtering on speech loudness. The wide band speech level is varied until it is equal in loudness to low-pass filtered speech. This is done as a function of the filter cutoff frequency. The same experiment is repeated for the high-pass speech. Next the energy of the equally loud wide band speech is expressed as a percentage of the energy of the low pass speech. An identical calculation is performed on the equally loud high-pass-filtered speech. For example, if wide band speech is to be equal in loudness to speech that has been low-pass filtered to 1 kHz, it must be reduced in level to 17% of its original energy. The corresponding relative level for 1 kHz high-pass-filtered speech is 7%. These functions are shown as the solid lines in the figure. We see that the banded loudnesses do not add to one because the two solid lines cross at about 11%. After taking the cube root, however, the loudness curves cross at 50%, and therefore sum to 100%. A level of 11.3 μbars corresponds to 1.13 Pa, or about 94 dB SPL. It is interesting to note that this loudness argument is nearly identical to Fletcher's articulation argument described in Fig. 2.

Fletcher then generalized Eq. (6) to K bands with the formula

$$e = e_1 e_2 \cdots e_K . \tag{7}$$

Given e we may compute s, and then $S(s) = s^3$, as shown in many publications (Fletcher and Steinberg, 1929, p. 35); (Fletcher, 1929a, Fig. 125); (Fletcher and Galt, 1950, p. 93); (Fletcher, 1953, p. 285). This model was tested with hundreds of combinations of channel parameters and is impressively accurate over a large range of channel conditions (Fletcher and Galt, 1950; French and Steinberg, 1947; Fletcher, 1953).

When words instead of nonsense syllables are used, the sound-unit errors must be further transformed to determine the word intelligibility $W(\mathscr{A})$. This case represents a decrease in the speech entropy. These methods were partially worked out as well (Fletcher and Steinberg, 1929, Figs. 11 and 13 and Eqs. 2 and 15). For a more detailed review of this work, see (Allen, 1994).

B. The History of AI

A major articulation study was begun in about 1921 and was proprietary within AT&T. These ideas and Fletcher's method were not publicly released but were internally documented in an AT&T memorandum (Fletcher, 1921). Thus this work remained largely unpublished until 1947. The internal records in the Bell Labs files are now marked "declassified."[5] The company private records were held by J.C. Steinberg, and only a handful of people had access to them (namely Fletcher, Steinberg, Snow, Munson, Galt, and Dunn).

The method that Fletcher developed was first publicly described in the French and Steinberg article (French and Steinberg, 1947). Other earlier references that touched on related topics are Fletcher (1922c); Fletcher (1922d); Fletcher and Steinberg (1929); Fletcher (1929a); and Steinberg (1936). The original 1921 Fletcher report (no. 21839 dated 4/8/21) is in the Fletcher archives at BYU (Fletcher, 1921). The method used by Fletcher in the 1924 *loudness* paper with J.C. Steinberg (Fletcher and Steinberg, 1924), as shown in Fig. 3, is almost identical to the method described in the AI work (Fig. 2) published in 1922 (Fletcher, 1922c), and again in greater detail in his 1929 book (Fletcher, 1929a). Based on the similarity of the loudness mapping and the articulation mapping, it is now clear how Fletcher was thinking about these two problems.

The articulation index was used at the Harvard Acoustics Laboratory during World War II for high noise and narrow-band systems. In 1943 Fletcher[6] gave Beranek an internal report describing the articulation index to evaluate pilot–navigator communications. Beranek headed the Electroacoustics Lab at Harvard, which was funded by the National Defense Research Committee (NDRC), to develop an improved pilot headset. Fletcher was the Chairman of the NDRC (1940–44). Beranek used the method to solve the problem of British pilot–crew communications by improving the signal-to-noise ratio in the pilot's ear. At low pressures

communications broke down because of the reduced speaking level, the reduction in the receiver (earphone) damping, which caused it to become resonant, and the high noise level. The earphone problem was fixed with a Bell System earphone that was magnetically damped, produced at the Hawthorne plant in Chicago. This earphone worked in a rarefied atmosphere. Pilot communications may have been the first non-Bell application of the articulation theory developed at Bell Labs for telephone purposes. As a result of these WWII studies, Beranek and Fletcher became friends. If it had not been for Beranek's insistence after the war that the method be published, the articulation index might still be unknown. A final internal memorandum published in 1947 (Fletcher, 1947; Fletcher and Galt, 1950) and again in his book (Fletcher, 1953), were his final words on the subject.

III. LOUDNESS

On 8 February 1924, the American Telephone and Telegraph Company and the Western Electric Company created *The Bell Telephone Laboratories*. By September of that year Fletcher and Steinberg published an important paper on the measurement of the loudness of speech signals (Fletcher and Steinberg, 1924). In this paper, when describing the growth of loudness, the authors state

> the use of the above formula involved a *summation of the cube root of the energy rather than the energy*.

This cube-root dependence was first described by Fletcher the year before (Fletcher, 1923b; Fletcher, 1923c). Figure 3 is from the 1924 work where they studied the loudness of low- and high-pass filtered speech. There was a fundamental question as to which physical variable best described loudness. Fletcher began by assuming that this variable was the acoustic intensity. This view was initially encouraged by his finding that loudness was related to log intensity (i.e., Fechner's law). Acoustic intensity dominated Fletcher's early view, resulting in his heavy use of the decibel. Later Fletcher and Munson used pressure as the measure. The modern view is to work directly with the linear variables, such as pressure and velocity. For example, dB SPL is defined in terms of intensity, but thought of as a pressure. This question, of the relevant physical variable, is still unresolved. For example, we now know that cochlear hair cells respond to their cilia displacements. We still do not know what controls the basilar membrane nonlinearity, although we know that it is related to outer hair cell membrane voltage.

Today any power-law relation between the intensity of the physical stimulus and the psychophysical response is referred to as *Stevens' law* (Rosenblith, 1959). Fletcher's 1923 loudness equation establishes the important special case of loudness for Stevens' approximate, but more general, psychological ''law.'' Weber's ''law'' states that $\Delta I/I$ is constant, where I is the intensity and ΔI is the just-noticeable difference (jnd) in the intensity. Weber's law is known to be only approximately correct for pure tones (Riesz, 1928). Fechner's ''law'' is based on the idea that the jnd may be integrated to obtain a psychophysical scale, which in this case is the loudness growth relationship (Boring, 1929). Stevens argued that Fechner's law was fundamentally incorrect. The relation between the loudness growth law and the jnd has been a longstanding problem[7] (Scharf, 1978).

In 1933 Fletcher and Munson published their seminal paper on loudness (Fletcher and Munson, 1933), which detailed the relation of iso-loudness across frequency (loudness level or phons), their loudness growth argument (in loudness units, or sones), a model of masking, and the idea behind the critical band (critical ratio). The arguments they used were elegant, and the results are important. They used ''the method of constant stimuli, wherein the observer listens to fixed levels of the two sounds and estimates which sound is the louder . . . .'' They then computed the percent of the observations that reported the test sound to be louder than the reference tone, as a function of the sound level of the test sound. In other words, they measured loudness psychometric functions.

A. Loudness additivity

Instead of thinking directly in terms of loudness growth, they tried to find a formula describing how the loudnesses of multiple stimuli combine. From loudness experiments with low- and high-pass speech and complex tones (Fletcher and Steinberg, 1924; Fletcher, 1929a), and other unpublished experiments over the previous 10 years, loudness had been found to add.

Today this model concept is called *loudness additivity*. The hypothesis was that when two equally loud tones were presented together, the result would be ''twice as loud.'' This method is now referred to as the *indirect method*. Fletcher verified his additivity hypothesis by use of the *direct method* in which subjects are asked to turn up the sound until it is ''twice as loud.'' A further verification of this assumption lies in the predictive ability of this additivity assumption. For example, they showed that 10 tones that are all equally loud, when played together, are 10 times louder, as long as they do not mask each other. Fletcher and Munson found that loudness additivity held for signals ''between the two ears'' as well as for signals ''in the same ear.'' When the tones masked each other (namely when their masking patterns overlapped), additivity still holds, but over the overlapping set of patterns (Fletcher and Munson, 1933).

In the 1933 paper the argument was refined to account quantitatively for the difficult problem of masking. The masking M is defined as the ratio of the threshold sound pressure measured in the presence of a masking signal to the threshold sound pressure measured in quiet. Each tone has a masking pattern that spreads away from the tone (Fletcher, 1923b; Wegel and Lane, 1924). They first showed that the loudness of two equally loud tones adds to twice the loudness if, and only if, the tones are far enough apart that they do not mask each other. They also showed how to account for the masking in the loudness calculation when masking occurs, by defining a *neural excitation pattern* on the basilar membrane, and account for the overlap of the partial excitation patterns for each of the tones. Masking reduces the partial loudness of the tones, but the additivity law continues to hold. Today this model is fundamental to our understanding of auditory sound processing.

Fletcher's working hypothesis was that each signal is *nonlinearly compressed* by the cochlea, neurally coded, and the resulting cochlear nerve neural rates are added. The 1933 experiment clearly showed how loudness (i.e., the neural rate, according to Fletcher's model) adds. Fletcher and Munson also determined the *compression function* $G(p)$. Their experiment does not prove that $G(p)$ must result from the nonlinear action of the cochlea, but it is consistent with it. Their argument is now carefully reviewed.

1. The argument

Let $L(p(t))$ be the perceived loudness in sones corresponding to pressure waveform $p(t)$. When the signal consists of two pure tones, the pressure $p(t)$ is $p_1 \sin(\omega_1 t)+p_2 \sin(\omega_2 t)$. We say that *the tones do not mask each other* when they are sufficiently far apart in frequency that their masking patterns do not overlap. This condition was first established by Fletcher in 1923 (Fletcher, 1923b) and Wegel and Lane in 1924 (Wegel and Lane, 1924). Let $G(p_1,p_2,f_1,f_2)$ be the nonlinear compression function that maps the pressure amplitudes p_1 and p_2 in Pascals into the loudness in sones, *under the condition that the tones do not mask each other*. We shall suppress the last two arguments f_1 and f_2 except when they are necessary and will refer simply to $G(p_1,p_2)$. When the tones mask each other, Fletcher and Munson found that the loudness L was always less than G, namely $L \lesssim G$. When each tone is presented alone there is no masking, so $L=G$. It also follows that $L_1=G(p_1,0)$ and $L_2=G(0,p_2)$. We assume that $G(0,0)=0$. The problem is to find $G(p_1,p_2)$.

2. Choice of p_1 and p_2 level

Pressure p_1 is taken as the reference level for the experiment with $f_1=1$ kHz. The level of pressure p_2, at frequency f_2, was next determined by requiring that its loudness be equal to that of p_1. We call this pressure $p_2^*(p_1,f_2)$ since it is a function of both p_1 and f_2. In other words, p_2^* is defined by $L(p_2^* \sin(\omega_2 t))=L(p_1 \sin(\omega_1 t))$ (the two tones have the same loudness in sones, and p_2^* is on the same iso-loudness contour as p_1, namely they have the same phon value). In terms of the compression function G, p_2^* is defined by

$$G(0,p_2^*)=G(p_1,0). \tag{8}$$

3. The experiment

Next, Fletcher and Munson scaled pressure p_1 by scale factor α and defined α^* such that the loudness of $\alpha^* p_1$, which we write as $L(\alpha^* p_1 \sin(\omega_1 t))$, is equal to the loudness of p_1 and p_2^* when played together, namely $L(p_1 \sin(\omega_1 t)+p_2^* \sin(\omega_2 t))$. In terms of G this condition is

$$G(\alpha^* p_1,0)=G(p_1,p_2^*). \tag{9}$$

This equation says that the loudness of $\alpha^* p_1$ is equal to the loudness of p_1 and p_2^* presented together, as long as they do not mask each other. This equation defines α^*.

4. Results

For f_1 between 0.8 and 8.0 kHz, and f_2 far enough away from f_1 (above or below) so that there is no masking, $20 \log_{10}(\alpha^*)$ was found to be 9 dB for p_1 above 40 dB SPL. This value decreased linearly to 2 dB for p_1 at 0 phons, as shown in Fig. 4. They found that the loudness $L(p_1 \sin(\omega_1 t)+p_2^* \sin(\omega_2 t))=G(p_1,p_2^*)$ does not depend on p_2^* as f_2 is varied (α^* was the same for any p_2 on the p_1 iso-loudness contour).[8] Thus we may write $\alpha^*(p_1)$ to show its dependence on p_1, and its independence of p_2^*.

5. Discussion of the experiment

Since the loudness $G(p_1,p_2^*(p_1,f_2),f_1,f_2)$ is constant along the iso-loudness contour $p^*(f_2)$, it is independent of f_2. A mathematical summary of their experimental result is

$$\frac{d}{df_2} G(p_1,p_2^*(p_1,f_2),f_1,f_2)=0. \tag{10}$$

Fletcher and Munson found an elegant (but not unique) solution to this differential equation. They tested the assumption that

$$G(p_1,p_2)=G(p_1,0)+G(0,p_2), \tag{11}$$

namely that the loudness of the two tones adds. Equation (11) is a solution to Eq. (10) because the first term on the right-hand side is independent of f_2, and the second term is independent of f_2 on the iso-loudness contour p_2^*.

Using Eq. (8), Eq. (11) becomes

$$G(p_1,p_2^*)=2G(p_1,0). \tag{12}$$

Combining Eq. (9) and Eq. (12) gives the nonlinear difference equation

$$G(\alpha^*(p_1)p_1,0)=2G(p_1,0), \tag{13}$$

which determines G once $\alpha^*(p_1)$ is specified. $G(p)$ may be found by graphical methods, or by numerical recursion, as shown in Fig. 136 (Fletcher, 1953, p. 190).

Fletcher and Munson found that at 1 kHz, and above 40 dB SPL, the pure tone loudness G is proportional to the cube root of the signal intensity [$G(p)=(p/p_{\text{ref}})^{2/3}$, since $\alpha^*=2^{3/2}$, or 9 dB]. This means that if the pressure is increased by 9 dB, the loudness is doubled. Below 40 dB SPL, loudness was frequently assumed to be proportional to the intensity [$G(p)=(p/p_{\text{ref}})^2$, $\alpha^*=2^{1/2}$, or 3 dB]. Figure 4 shows the loudness growth curve and α^* given in Fletcher (1953, Table 31, page 192). As may be seen from the figure, the measured value of α^* at low levels was not 3 dB, but was closer to 2 dB. Fletcher's statement that loudness is proportional to intensity (α^* of 3 dB) was an idealization that was appealing, but not supported by the actual results. The basic idea, and the cube root dependence on intensity above 40 dB SPL, was first published in Fletcher (1923b).

6. Discussion of loudness additivity

The conclusion that loudness of tones adds follows from the ''generalizability'' of Eq. (11), which was found to hold *even when the tones mask each other* (Fletcher and Munson, 1933). Most of the 1933 paper deals with this case of loud-

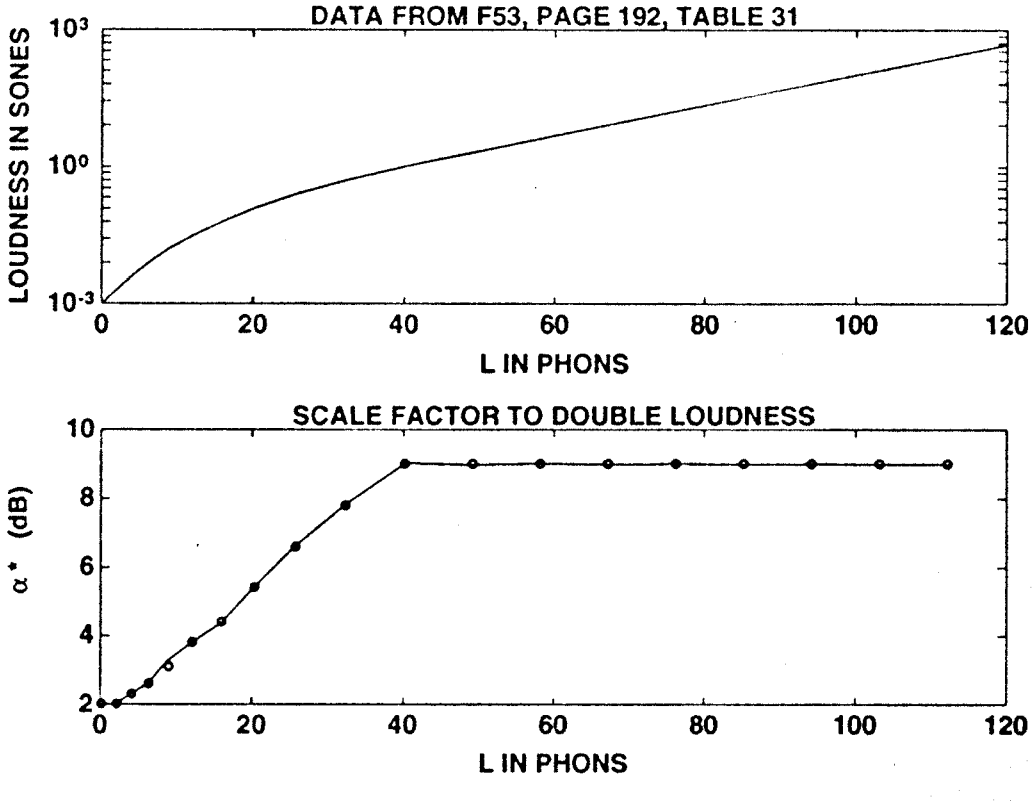

FIG. 4. This figure shows the loudness growth and α^* from p. 192, Table 31 (Fletcher, 1953) as a function of the loudness level, in phons. When α^* is 9 dB, loudness grows as the cube root of intensity. When α^* is 3 dB, loudness is proportional to intensity.

ness additivity when masking is present. In 1933 Fletcher and Munson found that additivity of loudness holds for two tones of the same frequency in different ears, and for as many as 2 to 11 tones in the same ear, as long as the tones were at least a critical band apart. The loudness growth function G and *loudness additivity*, when generalized, form the foundation for our present-day theory of loudness.

There are two exceptions to Fletcher's loudness additivity ''law.'' The first led Fletcher to the discovery of the critical band. The second remains unstudied: additivity seems to break down for two tones of different frequencies in different ears (Stevens and Davis, 1983), for speech in noise (Jeng, 1992), and for certain ambiguous musical *auditory streams*, namely complex tones that fuse across time and frequency in unpredictable ways [unpublished observations of the author]. It may be that loudness additivity only holds for a single *auditory stream*.

Fletcher and Munson's unique formulation of the loudness question was in terms of loudness additivity rather than loudness growth. They never asked the subjects to estimate the loudness directly, only to report which of two sounds was louder.[9] This meant that the role of the subject's ''gray-matter'' was reduced to the binary task ''which sound is louder?'' They found that a nonlinear transformation of the pressure reduced loudness combination to an algebraic sum. These results are consistent with a qualitative model of tonal loudness, which assumes that (a) the nonlinearity $G(p)$ is in the cochlea, (b) masking occurs at the neural level, (c) loudness is a function of the neural spike rate (i.e., its integral), and (d) the additivity of loudness results from a summation of the neural spike rate by the central nervous system (CNS).

Today we know that the cochlea *is* nonlinear, and we know that recruitment (abnormal loudness growth) occurs there (Kiang *et al.*, 1970; Liberman and Kiang, 1978). In fact, we know from noise trauma experiments on animals and humans that recruitment results from outer hair cell (OHC) loss (Liberman, 1984). The loss of OHC's causes a loss of the basilar membrane compression (e.g., a change in G) described by Rhode in 1971 (Allen and Fahey, 1992; Pickles, 1982, p. 291). This model of hearing and hearing loss, along with the results of Fletcher and Munson (1933), are basic to an eventual quantitative understanding of cochlear signal processing and the cochlea's role in detection, masking, and loudness in normal and impaired ears. The work by Fletcher (Fletcher, 1950) and Steinberg and Gardner (Steinberg and Gardner, 1937), and work on modeling hearing loss and recruitment (Allen, 1990) support this view. It is sad that, as a result of our preoccupation with models of statistical detection, intensity discrimination, and masking, we have tended to ignore these important early nonlinear loudness results — and for this reason, many of these recent models seem to be very limited in their scope.

B. The relation between masking and loudness

Fletcher then introduced a nice insight. The masking $M(f)$ is the ratio of the raised threshold due to a masker, to the threshold in quiet, expressed in dB. He argued that if the masking depends on the activity at the auditory nerve, then the loudness must depend on the masking, since the loudness

also depends on the stimulus to the nerve. He found that once the masking $M(f)$ is measured for a given noise stimulus condition, then the loudness is uniquely determined (Fletcher and Munson, 1937; Fletcher, 1938a; Stevens and Davis, 1983, p. 128). By carefully measuring the masking $M(f)$ (which is a function of frequency) and the loudness N (which is a single number) for many stimulus conditions, he found the functional relation between them. This relation is

$$N = \int F(M(x))dx, \tag{14}$$

where x is the place in the cochlea corresponding to masker frequency f. Thus they had reduced the loudness calculation to one of experimentally determining the masking M. *Accurate models for $M(x)$ remain an unsolved problem to this day*. Today we call $F(M(x))$ the *neural excitation pattern*, or the *partial loudness*. The integral over x represents the additivity of the partial loudness F along the basilar membrane. A more complete model of the cochlea should be able to relate $G(p)$ to $F(M(x))$. This connection remains as a challenge to auditory models.

C. Stevens and Davis

An important book appeared in 1938 entitled *Hearing, its Psychology and Physiology*, written by S.S. Stevens and H. Davis, two eminent Harvard scientists. Stevens earned his Ph.D. under Boring in 1934, and Davis received a B.A. in chemistry (*summa cum laude*) in 1918, and an M.D. in 1922, both from Harvard. This followed by a postdoctoral year at Cambridge University (England) under E.D. Adrian (later Lord Adrian), where he studied electrophysiology.

In the 1983 Preface to the ASA's reprint of *Hearing* (Stevens and Davis, 1983) Hallowell Davis wrote

> The Bell Telephone Laboratories, particularly the group headed by Harvey Fletcher, had led the way, and their data, assembled by ''Smitty'' Stevens in the first section of *Hearing*, are still valid. Additions have been made, of course, but few modifications of the basic ideas have been necessary. Stevens' lucid presentation of the psychological section makes the present reprint [of *Hearing*] worthwhile.

1. Stevens and loudness

Chapter 4 of the Stevens and Davis book is entitled *Loudness*. On page 114 Stevens discusses methods of loudness judgment. He defines the *direct* method as one that requires the subject to adjust a sound to one-half or twice as loud as a reference sound. The alternative method, the one used in 1933 by Fletcher and Munson (Fletcher and Munson, 1933) was tacitly therefore the *indirect* method.

While this chapter relies heavily on Fletcher and Munson's (1937) (Fletcher and Munson, 1937) work, it discounts the indirect method and recommends the ''direct'' method. On page 114 they say

> This method of direct estimate is necessarily the most fundamental (although not the most reliable) one under the criteria previously laid down for the nature of the loudness scale. The other methods are also valid in so far as they offer alternative ways of getting the same result.

The direct method is ''the most fundamental'' because ''twice as loud'' is, *by definition*, what people heard when you asked them to ''double the loudness.''

The indirect and direct methods are two different methods for measuring loudness. Both methods give approximately the same loudness growth curve (according to both Stevens and Fletcher). Thus there is no substantial controversy about the shape of the loudness growth function. The indirect method gives this curve with smaller variance, a point also made by both Stevens and Fletcher.

This then leads to the following question: *If the direct method is the most fundamental, why did Fletcher and Munson rely so heavily on the indirect method in their 1933 paper?* I believe that the answer to this question highlights a serious longstanding misunderstanding regarding measurement methods of loudness, which is important to openly discuss and clarify.

2. The direct versus the indirect method

The direct and indirect methods differ in the nature of the task required of the CNS. With the direct method the CNS must estimate the relative magnitude of the two signals, and judge when the two signals are related by a 2:1 magnitude (loudness) ratio. With the indirect method, the CNS must judge when the two signals (which for the monaural case differ in pitch) have equal intensity. In this case the CNS is operating as a null detector along the perceptual dimension called loudness. This is, in my opinion, a ''low level'' binary subjective task. Consider, for comparison, the ''mental computation'' required by the direct method, namely that required to estimate ''how many times louder'' one sound is relative to another. Because of the reduced involvement of the CNS in binary tasks, the results from the indirect method say more about the signals in the auditory nerve (the output of the cochlea). Comparisons between the direct and indirect method tell us something about how the CNS makes the ''mental computation.'' Fletcher realized that the best way to model loudness at the auditory nerve level was to use the indirect method, which uses the CNS as a null detector.

Stevens was looking for a loudness growth function, and not looking for relations between loudness and other variables. For example, he did not describe loudness in terms of masking, nor did he develop a model of loudness additivity.

Fletcher, on the other hand, was *not* looking for the loudness growth function $G(p)$. Rather he was studying the relations between equally loud sounds. These data can be *modeled* because they are only weakly influenced by the subjective nature of the task. While the direct method can properly claim that the signals are, on the average, twice as loud, *it can never lead to a model without a deeper understanding of the CNS processes responsible for the neural computation.*

With Fletcher's approach we conclude that loudness is additive when masking is not present, and that partial loud-

ness is additive when masking is present [Eq. (14)]. As a result of the breakdown of loudness additivity when the signals were close in frequency, he discovered the critical band. Fletcher's results are consistent with the idea that loudness growth function results from the cochlea and that loudness is a function of the total firing activity of the auditory nerve. In my view, Fletcher showed that a model of the loudness of tones requires a model of the mechanics of the cochlea, and not a model of CNS processes.

IV. THE CRITICAL BAND

When the 1938 book *Hearing* appeared, Fletcher and Munson were busy formulating the critical band. The resulting critical band papers (Fletcher, 1938a; Fletcher, 1938b) are important sources of information that are virtually unreferenced in the literature. The critical band paper that *is* commonly referred to in the literature is a tutorial/review paper, originally presented as a dinner address to the American Association of Physics Teachers in December of 1938 (Fletcher, 1940). Perhaps the failure to reference the 1938 papers has led to some misunderstanding and misrepresentation of Fletcher's model of the critical band.

The critical band idea was first published in 1933 (Fletcher and Munson, 1933, pp. 97–98); estimated bandwidths are given on page 98. By 1933 the concept of the critical band was ''obvious'' to Fletcher because of the following simple logic: When several tones are far apart in frequency, their loudnesses (the output of the cochlea) add; but when they are very close together, the acoustic intensities (the input) add.[10] Since the loudness $G(p)$ is $p^{2/3}$ above 40 dB SPL, the sum of the loudnesses for the two component case is $p_1^{2/3}+p_2^{2/3}$. This is not the same as the loudness of the sum of component intensities, namely $(p_1^2+p_2^2)^{1/3}$. Since below 40 dB SPL $G(p)$ is approximately p^2, the loudness should be approximately independent of the frequency difference between the tones, as was shown by Zwicker (Zwicker *et al.*, 1957). Since α^* is not exactly equal to 3 dB at low levels (see Fig. 4), systematic deviations are observed. Fletcher realized that this meant that for high levels there must be a transition between the two values of loudness, as $|f_2-f_1|$ increases, defining a filter bandwidth. This effect is shown in Fig. 15 of Fletcher (1938a) for a 10-component complex tone, and much later by Zwicker for a two component tone and Scharf for a four component tone.

Between 1934 and 1938 the details of the critical band were systematically worked out (Fletcher, 1934; Fletcher, 1935; Fletcher and Munson, 1937; Fletcher, 1938a; Fletcher, 1938b). The motivation of these studies was to find a loudness formula for signals having continuous spectra, such as speech and noise, where masking necessarily occurs. These results extended their 1933 results for (a) unmasked signals and (b) masked tonal signals (i.e., complex tones), where the excitation patterns overlap but are separated by more than a critical bandwidth.

(i) *Experiment I*. Fletcher and Munson first determined the spectrum of a noise that raised the pure-tone hearing thresholds by 50 dB uniformly across frequency. They found (Fletcher and Munson, 1937, Fig. 3) (Fletcher, 1938b, Fig. 2) that masking grew linearly for this noise, namely the masking increased 1 dB for a 1-dB increase in masker level.[11] This is an important generalization of Weber's law.

(ii) *The idea behind experiment I*. The idea was that a region of the noise was grouped by a cochlear filter. The tone also passed through this filter. They assumed that the probe to masker ratio (PMR) at the output of the filter, integrated over the basilar membrane, determines the masked threshold. This global PMR at the masked threshold was used as the *detection criterion* C. Fletcher argued that the difference in spectral level, in dB, between the tone (the probe, or maskee) at the masked threshold and the noise (the masker) is proportional to the cochlear filter bandwidth Δ_f at the probe frequency. He called this dB level difference the *critical band* κ. From this point on we shall refer to κ as the *critical ratio*, and the corresponding *critical bandwidth* Δ_f as the *equivalent rectangular bandwidth*, or ERB, as defined in Table II.

(iii) *Experiment II*. The bandwidth of a masking noise that had constant spectral level was increased and the masked threshold for a tone centered in that noise was measured. This experiment was first reported in a tutorial paper (Fletcher, 1940). The methods were not specified, as the results were intended to support conceptually the results of experiment I. This experiment is frequently referred to as the ''band widening'' or the ''direct'' method.

(iv) *The idea behind experiment II*. As the noise masker spectrum is increased in bandwidth from zero, the power at the tone probe (maskee), at place x, due to the masker must increase. At some point, however, the power seen by the hair cell will saturate as the cochlear filter bandwidth is reached. Experiments I and II lead to independent estimates of the cochlear filter bandwidth. The critical band experiment (I), which defines the ERB, κ, and C has frequently been confused with the band widening experiment (II).

A. Fletcher's energy model

The relation between the critical bandwidth and the critical ratio was derived using an energy analysis in the frequency domain of the tone–probe to noise–masker ratio at the output of the array of cochlear filters, at the signal detection threshold. This is a theoretical analysis of experiment I described above. I have reconstructed Fletcher's argument in greater detail than his original work using a modified notation for improved clarity.

Tables I and II provide a brief summary of abbreviations and definitions of the basic measures used by Fletcher. The *critical ratio* (CR) is a measure of the threshold probe to masker power spectrum ratio (PMR). The *basilar membrane power spreading function* $|H(f,x)|^2$ describes how the power along the BM is distributed for a tone at frequency f. The ERB $\Delta_f(x)$ is a measure of cochlear bandwidth at each place, while the ERS $\Delta_x(f)$ is a measure of the effective spread of tonal signal power at frequency f along the basilar membrane. The *cochlear map function* $F(x)$ describes the CF frequency f_0 of the maximum excitation on the basilar membrane for a given place x. When a probe tone is present along with the wide band masking noise, the *single hair cell tone probe to masker ratio* (PMR) is defined as $c(f,x)$. Integrating this PMR along the length of the BM

TABLE I. Summary of abbreviations.

Abbreviation	Description	Symbol
AI	articulation index	$\mathscr{A}$
BM	basilar membrane	
CNS	central nervous system	
CF	characteristic frequency	$f_0(x)\equiv F(x_0)$
CP	characteristic place	$x_0(f)$
CR	critical ratio	$\kappa(f)$
CVC	consonant–vowel–consonant	
ERB	equivalent rectangular bandwidth	Δ_f
ERS	equivalent rectangular spread	Δ_x
jnd	just-noticeable difference	ΔI
PMR	probe-masker ratio	c
SPL	sound pressure level	$20\log_{10}(p)+94$

gives the *gobal PMR* $C(f)$ which was Fletcher basic measure of signal detection. In the following paragraphs we summarize the basic relations between Fletcher's power-flow measures and derive the formula for the threshold level of a tone masked by wide band noise.

If a tone probe of frequency f and amplitude T is applied to the ear, the *power excitation pattern* along the basilar membrane is defined as

$$|\mathscr{T}(f,x)|^2\equiv|T(f)|^2|H(f,x)|^2 \quad \text{(Watts/m)}, \tag{15}$$

where $|\mathscr{T}(f,x)|^2$ is in Watts/m, $|H(f,x)|^2$ is the cochlear spatial power density ratio in units of m^{-1}, f is the frequency of the tone in Hz, and $|T(f)|^2$ is the level of the tone in Watts. This excitation pattern is maximum at one point $x_0(f)$ along the basilar membrane called the CP, as described by the inverse of the cochlear map function $x_0(f)=F^{-1}(f)$. The cochlear map is defined by the family of cochlear filters $H(f,x)$ and depends on the physics of the cochlea. An integral over x of $|\mathscr{T}(f,x)|^2$ gives the total acoustic power of the basilar membrane motion in response to the tone.

TABLE II. Summary of definitions.

Name	Symbol	Definition
Critical Ratio (CR)	$\kappa(f)$	$\left\|\frac{T^*}{N}\right\|^2\left(\frac{\text{Tone power @ threshold}}{\text{Noise spectral level}}\right)$
BM/Ear–canal power ratio	$\|H(f,x)\|^2$	$\left(\frac{\text{BM power density}}{\text{ear canal power}}\right)$
Single hair cell PMR	$c(f,x)$	$\left(\frac{\text{BM threshold}-\text{tone power}}{\text{BM Noise power density}}\right)$
Gobal PMR	$C(f)$	$\int_0^L c(f,x)dx$
PMR per critical band	$\bar{c}$	C/Δ_x
Equivalent rectangular bandwidth (ERB)	$\Delta_f(x_0)$	$\frac{\int\|H(f,x_0)\|^2\,df}{\|H(f,x_0)\|^2}$
Equivalent rectangular spread (ERS)	$\Delta_x(f)$	$\frac{\int\|H(f,x)\|^2\,dx}{\|H(f,x_0)\|^2}$
Cochlear map	$f_0=F(x)$	place to frequency map
Cochlear map	$x_0=F^{-1}(f)$	frequency to place map
Length of BM	L	

If a band of masking noise is applied to the ear, it is transformed to place x in a manner similar to that for the tone. In this case the *masker power excitation pattern* is defined as

$$|\mathscr{N}(x)|^2\equiv\int_0^\infty|N(f)|^2|H(f,x)|^2df \quad \text{(Watts/m)}, \tag{16}$$

where $|N(f)|^2$ is the spectral level of the noise, in Watts/Hz, at the input to the cochlea, and $|\mathscr{N}(x)|^2$ is the noise level *per unit length* along the basilar membrane in Watts/meter.

We may calculate Fletcher's threshold detection measure C from the *single hair cell* PMR $c(f,x)$ by substitution of the definitions of the probe and noise masker distributions along the BM

$$\begin{aligned} c(f,x)&\equiv\frac{\text{BM Tone power @ threshold}}{\text{BM Noise spectral level}}\\ &=\frac{|T^*(f)|^2|H(f,x)|^2}{\underbrace{\int_0^\infty|N|^2}_{\hookleftarrow}|H(f,x)|^2\,df}\\ &=\frac{\kappa(f)}{\Delta_f(x_0)}\underbrace{\left|\frac{H(f,x)}{H(f,x_0(f))}\right|^2}_{\text{excitation pattern}}. \end{aligned}$$

The * label on T indicates that it has been adjusted to threshold. The left hook ($\hookleftarrow$) indicates that the noise power spectrum is removed from the integral under the assumption that it is constant with respect to frequency over the cochlear filter bandwidth at location $x_0(f)$. The definitions for the critical ratio $\kappa(f)\equiv|T^*(f)/N(f)|^2$ and of the ERB are then substituted into the expression, given the final expression for $c(f,x)$.

The tone energy is spread over the ERS $\Delta_x(f)$ which, as discussed below, is about 40 hair cells in width (i.e., there are about 40 hair cells in one critical band). Fletcher summed up all the PMRs along the BM in a linear manner to account for the many hair cell channels (i.e., 40) having similar PMRs. Thus we must integrate c along the BM length to find $C(f)$, the gobal PMR. Referring to our definitions, $C(f)$ is

$$\begin{aligned} C(f)&\equiv\int_0^L c(f,x)dx\\ &=\int_0^L\underbrace{\frac{\kappa(f)}{\Delta_f(x_0)}}_{\hookleftarrow}\frac{|H(f,x)|^2}{|H(f,x_0)|^2}dx\\ &=\kappa(f)\frac{\Delta_x(f)}{\Delta_f(f)}. \end{aligned}$$

Finally, we define $\bar{c}$ as the *PMR per critical band*

$$\overline{c(f)}\equiv\frac{C}{\Delta_x(f)} \tag{17}$$

$$=\frac{\kappa(f)}{\Delta_f(x_0(f))}. \tag{18}$$

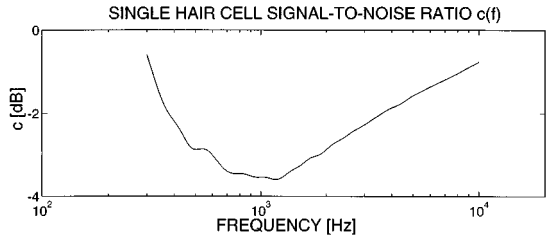

FIG. 5. Empirical estimate of $\overline{c(f)}$ for the cat based on the data of Costalupes (1983) and Liberman's neural tuning curves.

The measure $\bar{c}$ is important because it is the conversion factor between the critical ratio κ to the critical band Δ_f. Normalization by $\Delta_x(f)$ accounts for variations in the ERS with probe frequency.

We have empirically estimated $\overline{c(f)}$, the *PMR per critical band*, for the cat using κ from Costalupes (1983) and Δ_f computed from Liberman's neural tuning curve data, as shown in Fig. 5 (Garbes, 1994). The PMR per critical band has been a longstanding controversial ''constant,'' uncertain by a factor of about 2.5 (4 dB), creating uncertainty in estimates of the critical band. We see that $\overline{c(f)}$ for the cat is not independent of f and varies between -4 and -1 dB.

Fletcher assumed that C, the gobal PMR, is independent of the probe frequency f. We may test this question in the cat using the Costalupes (1983) and Liberman cat data. From Fig. 6 we see that C is a measure of detectability that is not constant below 1 kHz, but is constant above 1 kHz, as Fletcher assumed. $C=\bar{c}\Delta_x$ is constant above 1 kHz because $\bar{c}$ and $\Delta_x(f)$ inversely covary, with a constant product.

In summary, (a) C is approximately constant and $\bar{c}$ is not. (b) There are many hair cells with similar PMRs within a critical band (i.e., $\approx$ 40 = 0.5 mm/12μm). (c) We conclude that the *sum* of the signals from many hair cells must be used for tone detection (e.g., when we detect a tone we count the increase in the total number of spikes coming from those cells within a critical band).

(i) *Determining the cochlear map and C*. The *ERB density* over frequency is $1/\Delta_f(F^{-1}(f))$ and the *ERS density* over place is $1/\Delta_x(F(x))$. The number of critical bands between $x_1(f_1)$ and $x_2(f_2)$ is therefore

$$N_{12}=\int_{x_1}^{x_2}\frac{dx}{\Delta_x(F(x))}=\int_{F(x_1)}^{F(x_2)}\frac{df}{\Delta_f(F^{-1}(f))}.$$

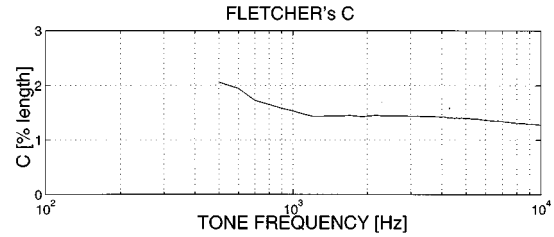

FIG. 6. We may estimate $C=\kappa\Delta_x/\Delta_f$ since we may directly estimate Δ_x from Liberman's neural data, or alternatively Δ_x/Δ_f from Liberman's cochlear map.

Since $\Delta_f=\kappa\Delta_x/C$, the number of critical bands between the stapes ($x_1=0$) and place X is given by

$$N_{0X}\equiv\int_0^X\frac{dx}{\Delta_x(x)} \tag{19}$$

$$=C\int_\infty^{F(x)}\frac{df}{\kappa(f)\Delta_x(f)}. \tag{20}$$

Given measured values of $\kappa(f)$, and assuming that $\Delta_x(f)$ is independent of frequency, this equation determines the inverse cochlear map function (the Bark scale)

$$x_0(f)\equiv F^{-1}(f)=C\int_\infty^f\frac{df}{\kappa(f)},$$

and determines L/C, and thus C, if we set $x_2=L$

$$\frac{L}{C}=\int_\infty^0\frac{df}{\kappa(f)}.$$

Fletcher initially assumed that Δ_x is constant. Based on von Bekesy's data, he soon established that $\Delta_x(x)$ is proportional to the width of the basilar membrane $W_{bm}(x)$, which varies by about a factor of 3 over its length. He also showed that it was necessary to account for the density of nerve fibers per unit length (Fletcher, 1953).

Fletcher computed the value of C and determined the cochlear map [Eq. (20)] from $\int df/\kappa(f)$ (Fletcher, 1953, p. 170). The value of $100C/L$ was found to be level dependent at low levels. At typical masker levels it was estimated[12] to be 1.1 (Fletcher, 1938b, p. 273 and Fletcher, 1938a, p. 284). Thus he had estimates for L, C, κ, the slope of the cochlear map $\Delta_f/\Delta_x=\kappa/C$, and the cochlear map $x_0(f)$.

To estimate $\bar{c}$, Δ_x, and Δ_f, he used the observation that the ERS and the frequency jnd are related by a factor of 20. As I understand it, he estimated the ERS as 20 times the length of a frequency jnd on the BM. On pages 172 and 410 of the book Fletcher argues that Δ_x is 1.5% of the basilar membrane length (Fletcher, 1953) because (1) the number of frequency jnd's is 1500, and (2) there are 20 frequency jnd's per critical band. Using the cochlear map, this provided him with everything he needed to calculate Δ_f and Δ_x. This gives a value of Δ_x of 0.5 mm (Fletcher, 1940, p. 56), which is about 40 hair cells in length, and that one frequency jnd is about 2 hair cells in length. Finally, he *verified* this estimate of the bandwidth using the band widening experiment (experiment II described above) using the data of Fig. 124 of the book (Fletcher, 1953, p. 171).

Is $\Delta_x\propto W_{\rm bm}(x)$ as Fletcher suggested (Fletcher, 1953, p. 172)? In Fig. 7 we have investigated Fletcher's statement that the ERS as a function of place covaries with the BM width $W_{\rm bm}(x)$. From these plots we see that for the cat, the ERS is approximately 2.35 times the width of the BM. The total variation in the ERS over the region plotted (i.e., from L/3 to L) is approximately a factor of 2.

B. Discussion

The basic assumption is that each group of hair cell of width Δ_x, taken as a group (what he called a ''nerve patch'' along the basilar membrane), has the same sensitivity. Fre-

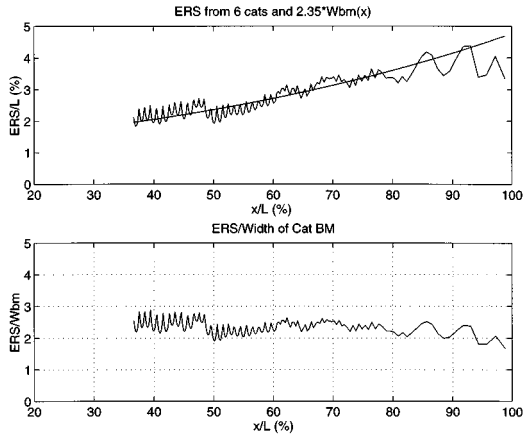

FIG. 7. The wavey line is $\Delta_x(x)$ (*re*: % of L) computed from Liberman's cat neural tuning curve data. The solid line is $2.75W_{\rm bm}(x)$, where $W_{\rm bm} = 0.11\exp(1.4\ x/L)$ is the width of the BM for cat. The lower plot is $\Delta_x(x)/W_{\rm bm}(x)$. When computing the estimates of the ERS, prototypical tuning curves were shifted over a small frequency range to generate neighboring tuning curves. The tuning curves were then transformed into place using Liberman's cochlear map, giving neural excitation patterns. The roughness in the ERS curve is due to smoothing artifacts present in the neural tuning curve data due.

quently he used $\overline{c(f)}=1$ for rough calculations (Fletcher, 1953, for example, Eqs. 10–13, p. 172). Later, Hawkins and Stevens (Hawkins and Stevens, 1950, pp. 8–9) were the first of many to question the value of $\bar{c}$ derived by Fletcher.[13] From the data shown in Fig. 5 we see that for the cat, $\bar{c}$ depends on frequency.

(i) *Conservation of cochlear energy*. The term $|H(f,x_0)|^2$, which is the power ratio at the place $x_0(f)$ of maximum response corresponding to frequency f, plays a critical role in the theory of hearing. The integral of $|H(f,x)|^2$ over x is, by definition, the power gain on the basilar membrane. Assuming that the basilar membrane wave propagation is lossless (a power gain of one) and all of the power of the tone is delivered to the detector (e.g., the cilia of the $\approx$40 hair cells), it follows that $\overline{|\mathcal{T}(x_0)|^2} = |T(f)|^2$, and therefore

$$\Delta_x(f)=1/|H(f,x_0)|^2. \tag{21}$$

Since the term $|H(f,x_0)|^2$ is a measure of the power sensitivity of the cochlear filter at its maximum, its reciprocal is a measure of the threshold of hearing. Therefore, conservation of cochlear energy implies that the ERS is proportional to the hearing threshold; namely, a higher threshold implies a larger ERS. Since the hearing thresholds are approximately constant over the midrange of hearing, $\Delta_x(f)$ is expected to be approximately constant. Today there is experimental evidence to support the validity of the energy conservation assumption (Allen and Fahey, 1992). He could have argued that Δ_x is constant based on conservation of energy, as given by Eq. (21), along with the fact that the hearing threshold, re the ear canal intensity, is approximately constant between 1 and 8 kHz.

(ii) *The rectangular filter assumption*. Nowhere did Fletcher assume that the filters were rectangular, as has sometimes been stated in the literature. In fact, he thought in terms of a continuum of filters along the basilar membrane, as described in his own work (Fletcher, 1930a; Fletcher, 1951), and in the work of his close colleagues (Wegel and Lane, 1924, Fig. 10). He did, however, describe the bandwidth in terms of the ERB, even though he did not use this terminology. Unfortunately he was not particularly consistent in his notation for the critical bandwidth, even in his 1953 book! I believe his poor notation has contributed to the confusion. He also could have done a lot more to help correct and clarify the errors people were making after 1953. On the other hand, it was clear that these people had not bothered to understand his papers. For example, they did not even reference the relevant papers in their studies.

(iii) *The term critical band*. Eventually the term *critical band* lost almost all of its original meaning as a cochlear filter bandwidth (the ERB) as researchers departed further from its original masking definition. This story does not yet have a happy ending since this set of issues has not yet been resolved.

V. SUMMARY

As the reader might appreciate after reading Fletcher's 1953 views, in 1918 Fletcher had taken on the toughest problem of all: to quantify and model how we hear and understand speech. This understanding allowed AT&T Bell Labs engineers to develop the necessary specifications of what was to become the largest telephone network in the world. The problems that Fletcher and his colleagues studied were so complicated, and took so many years, that it has been difficult to appreciate the magnitude of their accomplishments.

It is therefore understandable why his work has had such a great impact on our lives. Almost single-handedly[14] he created the fields of *communication acoustics* and *speech and hearing* as we know them today. Everyone who has ever used the telephone has reaped the benefit provided by this man and his genius. von Békésy, Davis, Stevens, and Zwicker are some of the names that come to mind when we think of hearing. Bell invented the telephone, and Edison made it into a practical device. Harvey Fletcher may not be as well known as these men today, but his scientific contributions to the fields of telephony, hearing, and human communication are absolutely unsurpassed.

Given this present opportunity to reflect back on this great man, *I would describe Harvey Fletcher as the singular intellectual force in the development of present-day communication acoustics and telephony.*

ACKNOWLEDGMENTS

I would like to thank the many people that have contributed to this history of Harvey Fletcher and his work.

[1]In January of 1912 Fritz Lowenstein demonstrated to Bell officials De Forest's audion amplifier. Within a year a project was organized under Arnold to study the device to see if it could be made practical. Arnold saw the blue glowing gas inside the tube, and realized that the tube would work better when evacuated. A vacuum pump was imported from Germany for this purpose (Fagen, 1975, p. 260).

[2]Wegel and Lane (1924) had previously published a key paper where they described the theory of the cochlea using a tapered electrical transmission line model.

[3]This paper won a Gold Medal from the Franklin Institute. The same paper was also published in the *Bell System Technical Journal* (Fletcher, 1992d).

[4]During this effort, the first low- and high-pass filters having sharp cutoff frequencies were designed by Crandall, Wente, Johnson, and McKown, based on the theory of G.A. Campbell, the first mathematician at Bell Labs. This theory and practice became the basis for network theory. Vacuum tube amplifiers were just coming into limited use. Microphones and receivers required entirely new designs. Campbell was also the first to suggest experimental articulation methods (Campbell, 1910; Fletcher, 1947). Thus a great deal of effort was devoted to these early articulation studies.

[5]The term ''declassified'' was not an official AT&T designation. ''Company private'' would have been a more appropriate term than ''classified.''

[6]In Fletcher (1992) it states that Fletcher was under the general supervision of the Office for Scientific Research and Development (OSRD) and charged with groups at Harvard and other universities. He received a citation signed by President Harry S. Truman for this work.

[7]See Hellman and Hellman (1990) for an interesting discussion of the relation between the intensity jnd and the loudness growth.

[8]The repetition of this experiment would make a nice lab problem for students.

[9]One must learn to ignore the large difference in pitch, and concentrate on the difference in loudness.

[10]Fletcher knew that the loudness of a subcritical–band random signal depends only on its energy, as shown in their narrow-band loudness data (Fletcher and Munson, 1937, Fig. 11).

[11]At high levels the masking power slightly decreases with increasing level (Fletcher and Munson, 1937, Fig. 17), (French and Steinberg, 1947, Fig. 8). This change could either be a degradation in signal processing ability, or an increase in the filter bandwidth with increased level. The latter seems more likely.

[12]The value of C as defined here may be computed from Fletcher's $C=0.5$ dB (Fletcher, 1938a, p. 284) using the expression $(L/100)10^{C/10}$. In all of Fletcher's work, C was expressed in terms of a percent of L.

[13]Why did they not discuss or reference Fletcher's argument or derivation, or most of his relevant papers?

[14]J.C. Steinberg and W. Munson played critical roles in this history, but Fletcher was the intellectual force.

Allen, J. B. (**1990**). ''Modeling the noise damaged cochlea,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Giser, J. W. Matthews, M. A. Russero, and C. R. Steele (Springer-Verlag, New York) pp. 324–332.

Allen, J. B. (**1994**). ''How do humans process and recognize speech?,'' IEEE Trans. Speech Audio Process. **2**(4); 567–577.

Allen, J. B. (**1996**). ''How do humans process and recognize speech?,'' *Modern Methods of Speech Processing*, edited by R. Ramachandran, and R. Mammone (Kluwer, Boston), pp. 251–275.

Allen, J. B., and Fahey, P. F. (**1992**). ''Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,'' J. Acoust. Soc. Am. **92**, pp. 178–188.

Boring, E. (**1929**). *History of Psychophysics* (Appleton–Century, New York).

Campbell, G. (**1910**). ''Telephonic intelligibility,'' Philos. Mag. **19**, 152.

Costalupes, J. A. (**1983**). ''Broadband masking noise and behavioral pure tone thresholds in cats,'' J. Acoust. Soc. Am. **74**(3), 758–764.

Crandall, I. (**1926**). *Vibrating Systems and Sound* (Van Nostrand, New York).

Fagen, M., Ed. (**1975**). *A History of Engineering and Science in the Bell System — The Early Years (1875–1925)* (AT&T Bell Laboratories, Murray Hill, NJ).

Fletcher, H. (**1920**). ''The relative difficulty of interpreting the spoken sounds of English,'' Phys. Rev. **15**, 513–516.

Fletcher, H. (**1921**). ''An empirical theory of telephone quality,'' AT&T Internal Memorandum 101(6).

Fletcher, H. (**1922a**). ''Discussion of the proposed transmission efficiency units,'' AT&T Internal Memorandum 338(5).

Fletcher, H. (**1922b**). ''Fundamental studies in the transmission of speech,'' AT&T Internal Memorandum 198(6).

Fletcher, H. (**1922c**). ''The nature of speech and its interpretation,'' J. Franklin Inst. **193**(6), 729–747.

Fletcher, H. (**1922d**). ''The nature of speech and its interpretation,'' Bell Syst. Tech. J. **1**, 129–144.

Fletcher, H. (**1922e**). ''The study of telephone quality outlines case 32–311,'' AT&T Internal Memorandum 249(5).

Fletcher, H. (**1923a**). ''Noise measurements on subway trains,'' AT&T Internal Memorandum 780(6).

Fletcher, H. (**1923b**). ''Physical measurements of audition and their bearing on the theory of hearing,'' J. Franklin Inst. **196**(3), 289–326.

Fletcher, H. (**1923c**). ''Physical measurements of audition and their bearing on the theory of hearing,'' Bell Syst. Tech. J. **2**(4), 145–180.

Fletcher, H. (**1926**). ''The theory of the operation of the howling telephone with experimental confirmation,'' Bell Syst. Tech. J. **5**, 27–49.

Fletcher, H. (**1929a**). *Speech and Hearing* (Van Nostrand, New York).

Fletcher, H. (**1929b**). ''What can I do,'' Volta Rev. **31**, 637–642.

Fletcher, H. (**1930a**). ''The progress of hearing tests in the public schools of the United States,'' Volta Rev. **32**, 1–4.

Fletcher, H. (**1930b**). ''A space-time pattern theory of hearing,'' J. Acoust. Soc. Am. **1**, 311–343.

Fletcher, H. (**1934**). ''Loudness, pitch, and the timbre of musical tones and their relation to the intensity, the frequency, and the overtone structure,'' J. Acoust. Soc. Am. **6**, 59–69.

Fletcher, H. (**1935**). ''Newer concepts of the pitch, the loudness, and timbre of musical tones,'' J. Franklin Inst. **220**(4), 405–429.

Fletcher, H. (**1938a**). ''Loudness, masking and their relation to the hearing process and the problem of noise measurement,'' J. Acoust. Soc. Am. **9**, 275–293.

Fletcher, H. (**1938b**). ''The mechanism of hearing as revealed through experiments on the masking effect of thermal noise,'' Proc. Natl. Acad. Sci. (USA) **24**, 265–274.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Fletcher, H. (**1947**). ''A modern version of the old theory of telephone quality,'' AT&T Internal Memorandum **110**(50), 1–124.

Fletcher, H. (**1950**). ''A method of calculating hearing loss for speech from an audiogram,'' J. Acoust. Soc. Am. **22**, 1–5.

Fletcher, H. (**1951**). ''On the dynamics of the cochlea,'' J. Acoust. Soc. Am. **23**, 637–645.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Krieger, Huntington, NY).

Fletcher, S. H. (**1992**). ''Harvey Fletcher 1884-1981,'' Biographical Memoirs **61**, 165–193.

Fletcher, H. and Galt, R. (**1950**). ''Perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.

Fletcher, H. and Munson, W. (**1933**). ''Loudness, its definition, measurement, and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H. and Munson, W. (**1937**). ''Relation between loudness and masking,'' J. Acoust. Soc. Am. **9**, 1–10.

Fletcher, H., and Steinberg, J. (**1924**). ''The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound,'' Phys. Rev. **24**(3), 306–317.

Fletcher, H., and Steinberg, J. (**1929**). ''Articulation testing methods,'' Bell System Tech. J. **8**, 806–854.

Fletcher, H., and Wegel, R. (**1922a**). ''The frequency-sensitivity of normal ears,'' Phys. Rev. **19**, 553–565.

Fletcher, H., and Wegel, R. (**1922b**). ''The frequency-sensitivity of normal ears,'' Proc. Natl. Acad. Sci. (USA) **8**(1), 5–6.

French, N., and Steinberg, J. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.

Garbes, P. (**1994**). ''Evaluating human neural tuning curves from a mechanical model of the cochlea,'' Master's thesis, MIT, Cambridge, MA.

Hawkins, J., and Stevens, S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.

Hellman, W., and Hellman, R. (**1990**). ''Intensity discrimination as the driving force for loudness. Application to pure tones in quiet,'' J. Acoust. Soc. Am. **87**, 1255–1271.

Hunt, H. (**1982**). *Electroacoustics* (Acoustical Society of America, Woodbury, NY).

Jeng, P. S.-F. (**1992**). ''Loudness prediction using a physiologically based auditory model'' Ph.D. thesis, The City University of New York (UMI, Ann Arbor, MI 48106).

Kiang, N., Moxon, E., and Levine, R. (**1970**). ''Auditory nerve activity in cats with normal and abnormal cochleas,'' in *Sensorineural Hearing Loss*, edited by G. Wolstenholme and J. Knight (J. & A. Churchill, London), volume CIBA Symposium, pp. 241–273.

Kingsbury, B. (**1927**). ''A direct comparison of the loudness of pure tones,'' Phys. Rev. **29**, 588–600.

Kranz, F. W. (**c.a. 1965**). ''The audiometer: Principles and history,'' Technical report, Riverbank Laboratories.

Liberman, M. (**1984**). ''Single-neuron labeling and chronic cochlear pathol-

ogy III. Stereocilia damage and alterations of tuning curve thresholds,'' Hear. Res. **16**, 55–74.
Liberman, M., and Kiang, N. (**1978**). ''Acoustic trauma in cats,'' Acta Otolaryngol. Suppl. **358**, 1–63.
Millman, S., editor (**1984**). *A History of Engineering and Science in the Bell System — Communication Sciences (1925–1983)* (AT&T Bell Laboratories, Murray Hill, NJ).
Pickles, J. (**1982**). *An Introduction to the Physiology of Hearing* (Academic, London).
Riesz, R. (**1928**). ''Differential intensity sensitivity of the ear for pure tones,'' Phys. Rev. **31**(2), 867–875.
Rosenblith, W. (**1959**). ''Sensory performance of organisms,'' Rev. Mod. Phys. **31**, 485–491.
Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception*, edited by E. Carterette, and M. Friedman (Academic, New York), Vol. IV, Chap. 4, pp. 187–234.
Sivian, L. J., and White, S. D. (**1933**). ''On minimum audible sound fields,'' J. Acoust. Soc. Am. **4**, 288–321.
Steinberg, J. (**1936**). ''Title unknown,'' in *Communication and Electronics*, edited by H. Pender and K. McIlwain (Electrical Engineers' Handbook), Vol. 5, p. 9:27.
Steinberg, J., and Gardner, M. (**1937**). ''Dependence of hearing impairment on sound intensity,'' J. Acoust. Soc. Am. **9**, 11–23.
Steinberg, J., Galt, R., and Anderson, A. (**1937**). ''Study of speech and hearing — case 20871-3,'' Bell Syst. Tech. Memorandum **328**(67), 1–36, I:1–17, II:1–16, III:1–9, Figs: 1–115.
Stevens, S., and Davis, H. (**1983**). *Hearing, Its Psychology and Physiology*. (Acoustical Society of America, Woodbury, NY).
Watson, L., and Tolan, T. (**1967**). *Hearing Tests and Hearing Instruments*. (Hafner, New York).
Wegel, R., and Lane, C. (**1924**). ''The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear,'' Phys. Rev. **23**, 266–285.
Zwicker, E., Flottorp, G., and Stevens, S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.