

Improving the Noise Performance of Communication Systems: 1920s to early 1930s

**Mischa Schwartz
Charles Batchelor Professor Emeritus
of Electrical Engineering
Columbia University
New York, NY 10027**

Abstract

This paper discusses the early pioneering work of both telephone and radio engineers in effecting improvements in the noise performance of communication systems. This work led ultimately to the explosive growth of communication activities following WWII. Radio engineers during the 1920s were most concerned with reducing the impact of externally-generated “static”, and showed this could be accomplished by the use of directional antennas and by moving to higher-frequency transmission. Telephone engineers during that period of time, most prominently John R. Carson of AT&T, were led to include the impact of “fluctuation noise” (shot and thermal noise) as well. Carson, using the novel concept of noise frequency spectrum, showed how the appropriate choice of bandwidth and frequency of transmission could be used to improve the signal-to-noise ratio, anticipating the concept of “matched filter” introduced 20 years later during radar developments of WWII. We conclude this discussion of this early period of improvement in system noise performance by describing the spectacular leap ahead made by Armstrong with his 1933 patent for noise-reducing FM

1. Introduction and Overview

The fields of communication theory and information theory developed explosively after World War II, following extensive radar development during that war. Critical to both of these fields is the study of system performance in the presence of noise. Communication system noise performance has its genesis in early work in radio telegraphy, telephony, and radio telephony, beginning circa 1920 and continuing on through the 1920s and 1930s, and then culminating in the radar research activities of WWII. Basically, the problem in all fields of communication, including radar, is that of appropriately detecting signals in the presence of noise. In this paper we focus on studies during the 1920s and early 1930s in radio (wireless) communications and radio telephony to determine how best to reduce the impact of noise on system performance. As we shall see, the term “noise” had two different meanings during that period of time: to radio engineers it meant atmospheric or static, arising from natural environmental causes and interfering with good radio reception, since that form of disturbance predominated at the time; to telephone engineers, aware of, and initially concerned about static as well, it became apparent in the early 1920s that fluctuation noise was the basic and more fundamental contributor to noise interfering with the appropriate reception of telephone signals.

The study of random or fluctuation noise and its impact on communication systems appears to have begun in 1918 with the publication of a now-classic paper by Walther Schottky, a German physicist working for the Siemens organization at the time

[1]. He was the first to describe the two types of fluctuation noise in electronic circuits, thermal or resistive noise (Wärmeeffekt in his terminology) and shot noise (his Schrotteffekt). The first is due to random motion of charge flowing through an electrical conductor in “response to the heat motion of the molecules making up the conductor” [2]; the second is due to the variations in electrical current as charges, whose motion gives rise to the current, are randomly emitted (hence the term “shot” effect). Note that Schottky’s formulations came shortly after the vacuum-tube amplifier came into prominence. Although Schottky’s paper describes both types of noise, its focus is on shot noise, since it was apparently much stronger than thermal noise in the amplifiers of the time. (As noted above, however, it was “atmospheric static” that dominated studies of radio noise for many years, at least until 1930.) Schottky’s paper stimulated a great deal of activity on shot noise, both experimental and analytical. It wasn’t until the work of J. B. Johnson of the Bell System in the 1920s, who initially began studying shot noise and then found that thermal noise was the more fundamental type of fluctuation noise, that thermal noise began to receive serious consideration. Johnson summarizes his work and that of others at the time in a brief retrospective paper on the first two decades of the studies of noise [2]. Johnson’s experimental work on thermal noise and Harry Nyquist’s theoretical study and explanation of thermal noise were presented in two classic papers, appearing back-to-back in the same issue of the 1928 Physical Review [3], [4].

A number of historical studies of fluctuation noise in electronic systems, with particular emphasis on the physical generation mechanisms of shot noise in vacuum tubes, and, later, in solid-state devices, have appeared in the recent literature. (See, for example, [5].) In this paper we avoid discussions of the physical mechanisms involved with noise generation and focus instead on the history of the impact of noise on communication systems, how engineers linked the performance of communication systems to the noise encountered in their use, and developed methods for reducing (never eliminating!) the effect of noise. We cover the initial period of investigation of the impact of noise on communication systems, from the early 1920s to the monumental work of Edwin H. Armstrong who, with his 1933 patent for noise-reducing wideband FM, provided a spectacular leap ahead in improving the noise performance of communication systems.

Both radio and telephone engineers contributed to this early work on effecting improvements in system noise performance, although, as noted earlier and as we shall shortly see in more detail, radio engineers in the 1920s focused almost exclusively on reducing the impact of externally-generated “static”, while telephone engineers led by John R. Carson of the Bell System, broadened the scope of their investigations to include the impact of shot and thermal noise, generically called “fluctuation noise”, as well. It was only in the early 1930s, especially in the case of Armstrong’s invention, that radio engineers began to take fluctuation noise into account in the design of systems. It was during the period of the 1920s that engineers, principally those working in the telephone industry, began to recognize that the noise power in a given system was proportional to the bandwidth of the system. (Nyquist, for example, in the 1928 paper [4] discussing thermal noise referenced above, came up with the famous equation showing the noise power was directly proportional to the bandwidth.) Spectral considerations, variations of noise and signal power density with frequency, also began to be developed and clarified in the 1920s. Since noise power played such a significant role in system performance,

much effort went into system design considerations for reducing the noise. Radio engineers recognized that a measure of the performance of radio receivers was given by the “signal-to-static” ratio, and attempted to design systems to improve this quantity. Carson and his colleagues broadened this measure of noise performance to include fluctuation noise and they, and other engineers, began working with the performance measure more generally called the “signal-to-noise ratio”, or SNR. It was soon found and understood that the noise in the first stages of an amplifier was most critical in determining the SNR in a receiving system: both noise and signal were equally amplified in moving through a system; hence it was most important to reduce the noise at the input to the system. In fact, as already noted, in radio receivers the noise determining performance was, for many years, atmospheric noise or “static” picked up by the antenna. Radio engineers thus worked to reduce the static picked up by antennas, as well as designing receiver front-end circuitry to reduce the impact of atmospheric noise.

In the next section we briefly describe the work of radio engineers in attempting to understand the phenomenon of static, as well as developing methods of designing systems to reduce the impact of static on radio receivers. In the third section we focus on the work of Carson and his colleagues in enlarging the scope of the investigation of noise to include fluctuation noise as well. In the fourth and concluding section of this paper we move to the work of Armstrong in attempting to reduce noise in radio receivers, culminating in his discovery of wideband FM as a noise-reducing method of communication signal transmission and reception.

2. Radio engineering and “static” reduction

As noted in the previous section, early radio engineers worked to reduce natural interference, “static”, being picked up by receiving systems in addition to the desired signals. This type of interference, generated, for example, in electrical storms or in emanations from the Sun, was found to vary with the seasons, with location, and frequency. It was also noted previously that radio engineers focused almost exclusively on the phenomenon of static. Although work on understanding the generation of shot noise in vacuum tubes was going on throughout the 1920s in a number of laboratories in the USA, England, and Germany, results of these studies being published in a variety of journals, it is interesting to note that very little of this work appeared in the radio engineering literature. A comprehensive perusal by this author of the Proceedings of the Institute of Radio Engineers (IRE) from its first issue in January, 1913 through 1929, for example, shows only occasional reference in passing to vacuum-tube shot noise.¹ Monthly issues throughout much of that period feature a Monthly List of References to Current Radio Literature prepared by the (U.S.) Bureau of Standards, which covers the literature from throughout the world, including the USA, England, France, Germany, and Japan, among other countries. There were no references at all found in these listings to shot or thermal noise. The only references to natural noise of any type were to atmospheric noise or static, as already noted. It wasn’t until 1930, with the publication of a paper by F. B. Lewellyn of AT&T [8] that shot and thermal noise, generically called fluctuation noise, received full recognition in the radio literature. This paper was followed soon thereafter, later in 1930, and in 1931, by a flock of papers on fluctuation noise. This lack of technical information on fluctuation noise in the radio engineering literature until 1930 was presumably due to the fact that static or “atmospherics”, as this

type of noise was often called, was the dominant form of noise appearing in radio receivers throughout the 1920s. The noise problem was of course very different in the wired telephone field, and, as we shall see in the next section, telephone engineers did study the effects of fluctuation noise appearing in the wired transmission receivers of the time.

Radio engineers, in focusing on reducing the impact of static on radio reception during the 1920s, both tried to determine the origin of the static encountered during radio reception, and to design systems to improve receiver performance in its presence. As part of this process, many experiments were carried out to characterize the types of static they found appearing in the radio receivers as well as determining their origin. For example, a series of papers by G. W. Pickard, beginning in 1920, and summarizing experiments he carried out, proposed a solar origin for static [9], [10]. This conjecture in 1920 led him to propose as well a form of directional loop antenna to reduce the static entering a radio receiving system [9]. Such an antenna pointed in the direction of the desired signal picks up a maximum amount of signal energy while at the same time reducing the static coming from other directions. Later work on improving reception of trans-Atlantic radio telephone signals in the presence of static referred specifically to Pickard's directional loop antenna [9] as pioneering in this area. Harold H. Beverage at RCA developed the so-called "Wave Antenna" or Beverage Antenna with marked directivity soon thereafter [11], [12]. An even earlier example of an improved receiving system developed by the Marconi Corporation for the purpose of reducing static was one incorporating multiple receiving antennas [13]. An early paper by AT&T and Western Electric engineers focusing on radio telephony notes that the "amount of noise [static] received is dependent upon the directional and selectivity characteristic of the receiving system." [14] Here "selectivity" means that the receiver bandwidth should be kept as small as possible consistent with clarity of voice reception. The implication is that the amount of static energy introduced into the receiver increases with bandwidth. Carson of AT&T would shortly thereafter quantify this concept, as explained in the next section.

Many other investigations of the origin and characterization of atmospheric static over a period of years were reported on in the radio literature of the 1920s. Examples include work by L. W. Austin of the US Naval Radio Research Laboratory (see, for example, [15], and papers in subsequent issues of the Proceedings of the IRE); by other workers in the radio field; and, more specifically, work by AT&T engineers reporting on measurements of static jointly with RCA engineers as to variation with frequency, time of day, season of the year, and location [16]. One of their conclusions was that the "Major source of long-wave static, as received in England and the US [was] of tropical origin." ² As noted earlier, the Proceedings of the IRE (Institute of Radio Engineers) of the 1920s had a regular section in each issue called References to Current Radio Literature in which many references to papers on atmospherics in the Proceedings itself and in radio journals throughout the world may be found.

One of the conclusions of these various measurements and studies of static was that these atmospherics decreased with increasing frequency [6], [15], [16]. In particular, experiments carried out on short-wave transmission (3-30 MHz) in the early and mid-1920s [17] indicated that there was "very little trouble due to strays [static]" and that the "actual limit for reception [was] fixed by set [shot] noise". (See also [12], pp. 641-643, published years later, and a leading textbook on radio engineering, in which the author

specifically states “the static intensity at short waves...is much less than at lower frequencies, and during a good part of the time is of the same order of magnitude as the noise level of a typical radio receiver”)

It was recognized early on that it was the “signal-to-static ratio” that determined the performance of a receiving system, as measured by its ability to detect an incoming signal. For as static entering the receiver increased, the received signal power had to be correspondingly increased to keep the signal detectable [13]. We find the authors of [14] noting that the basic performance parameter for radio telephony was the ratio of speech and noise [static] “volume”, as measured at the audio output. References [6] and [16] similarly used the ratio of signal to noise (the term noise here referring to static) as a performance measure to be made as large as possible. They use the generic acronym SNR to represent the *signal to noise ratio*. By the late 1920s and early 1930s, with the knowledge that both static and inherent fluctuation noise played a role in signal detectability, engineers generalized this concept, using the ratio of signal power to noise power in a communication system to determine the performance of a communications system, whether radio or telephony. Here *noise power* refers to the power of any interfering signal, whether static, fluctuation noise, or a combination of the two. The term SNR began to be adopted as the basic objective to be maximized, the performance measure of the communication system. As was the case with static, if the noise power increases in a given environment the signal power must increase correspondingly to maintain the same performance.

As noted in the Introduction to this paper, John R. Carson of the Bell System was among the first to recognize the significance of both static and fluctuation noise in determining the performance of communication systems, and to use SNR or related concepts in attempting to improve performance. We therefore move on in the next section to describe his work and that of his colleagues in the Bell System on noise performance during the 1920s.

3. Carson’s work on noise in communication systems

John R. Carson, who had received his electrical engineering training at Princeton University, graduating with a B.S. in 1907, an EE degree in 1909, and an M.S. in 1912, joined AT&T in 1914. He is best known for his invention of single sideband transmission in 1915 and his 1922 analysis of FM bandwidth. That he did ground-breaking work on noise in communication systems is less well-known. His first published paper on this subject appeared in the Proceedings of the IRE in 1923 [18]. This paper compares the signal-to- static interference ratio in single- and double sideband transmission. This measure of system performance is precisely the one cited earlier as having been used as a performance measure for radio systems as early as 1919. (*See* [12].) The difference, however, is that by “static interference” here is meant not only interference such as atmospherics, but, presumably, random noise as well. (This point is somewhat ambiguous, Carson stating, in footnote 1 of his paper, [18], that “The word *static* is used in a generic sense to cover random and irregular interference, such as atmospherics, as distinguished from the interference from another station.” [Emphasis added].) This ambiguity is cleared up at about the same time in a paper Carson co-authored by Otto J. Zobel, also with the Bell System, [19], in which the authors state specifically that random disturbances or interference can be “ ‘static’ in radio

transmission and ‘noise’ in wire transmission”. There is thus a distinction made between atmospheric *static* and the *noise* introduced in the system circuitry. The analysis carried out by these authors groups the two phenomena together, treating them both as examples of random disturbances.

In this second paper by Carson and Zobel, [19], the authors study the transmission of signals, including random disturbances, through “[frequency-] selective networks” or “wave-filters”, models for receiving systems which are tuned to specific frequency ranges. They use their study to derive a general expression for a quantity closely related to the signal-to-noise ratio, specifically the “statistical signal-to-random interference energy ratio”. They term this quantity the “selective figure of merit of the network with respect to random interference”. Although this paper is a theoretical one, providing general formulas only, it obviously arose out of work at AT&T and presumably elsewhere on means for reducing the effect of noise and static on communication systems. For the authors state directly on p.1 of the paper, that there has been a “uniform failure of wave-filters to suppress irregular and transient interference, such as ‘static’, in anything like the degree with which they discriminate against steady-state currents outside the transmission range. This limitation is common to all types of selective networks and restricts the amount of protection it is possible to secure from transient or irregular circuits.” The work described in this paper is therefore a direct attempt to provide a better understanding of the quantitative impact of noise on electrical systems, thereby leading to specific design considerations.

To develop the figure of merit, the authors first provide a model for the random disturbance at the input to a receiving system. To this end, they use a shot-noise-like model for the random interference, a model used much earlier, in 1909, by Norman Campbell in England [20] in studying the random emission of particles such as α particles (He^{++} ions), β particles (electrons), and light particles (photons), all of which had been discovered or postulated by the early 1900s. This model was also used by Schottky [1] and later investigators in studying the shot effect. Quoting Carson and Zobel, the model consists of “a large number of individual impressed forces...which are unrelated and varying in intensity and in waveform in an irregular, indeterminate manner and thus constitute what will be called *random interference*.” [Their emphasis.] Since they are studying the transmission of signals and noise through frequency-selective electrical networks, they must convert the shot-noise-like model for the noise to an equivalent formulation involving the variation of the random interference energy with frequency. This variation of energy with frequency is called the *random interference energy spectrum*. (“Spectrum” specifically refers to the variation of a physical quantity, here the random interference or noise energy, with frequency.) This frequency-based approach uses Fourier analysis, after the French mathematical physicist Fourier who, in the 1820s, first introduced the concept of a frequency spectrum for signals varying in space or time.

Carson and Zobel in this paper simply write down the defining equation for this energy spectrum of the random interference, with no indication of how it was obtained. There are no references provided; it is stated as a given.³ (The expression is now well-known, but this clearly was not the case at the time their work was done.) This expression, using, in their notation, the symbol $R(\omega)$ for the energy spectrum, is given by

$$R(\omega) = (1/T) |F(\omega)|^2$$

Here ω is the *radian* frequency, 2π times the frequency measured in Hertz (Hz), and $F(\omega)$ is the Fourier transform, generally a complex quantity, of the sum of the “individual impressed forces” taken over an interval T seconds long. The two vertical bars represent the absolute value of the Fourier transform. As noted, no reference is given for this equation. It is simply stated as a given. The authors use this equation for the energy spectrum of the random interference to develop a measure of the “mean energy absorbed per unit time by [the electrical network] from the random interference”. Using this quantity they go on to define a “figure of merit of a selective network with respect to random interference”, essentially the signal-to-noise ratio, to be maximized. They point out that, in order to proceed with a rigorous evaluation of the effect of random interference on signal transmission, $R(\omega)$ must be completely specified, i.e., known, over the entire frequency range (spectrum) of interest. $R(\omega)$ is generally *not* known. They note, however, that “it would appear that all frequencies are equally-probable in the spectrum $R(\omega)$...[making it] a constant, independent of [the frequency] ω . This inference, however, has not been theoretically established.” (Note that this corresponds to saying the noise has a flat or constant spectrum over the frequencies of interest. This assumption of what is now called *bandlimited white noise* was quite a remarkable conjecture at the time!) Their conclusion (p.2): “This formula leads to general deductions of practical importance regarding the relative merits of selective networks...[as well as] a method for experimentally determining the spectrum of random interference.” They then go on to state that “fortunately,... a complete specification of $R(\omega)$ is not at all necessary for a practical solution of the problem.” They show that, for a selective network, i.e., one involving, as noted earlier, frequencies in a range (bandwidth) about a specific center frequency, one can come up with a formula for the figure of merit. In their words, “this formula...furnishes...a means of estimating the comparative merits of the very large number of circuits which have been invented for the purpose of eliminating ‘static’ in radio communication, and leads to general deductions of practical value [for solving] the ‘static’ problem”. They follow with examples of selective circuits or filter types. The implication, therefore, is that one may evaluate and compare the signal-to-noise noise performance of many different types of practical circuits, using simple models of these circuits, without actually building them or analyzing them in detail. This represented quite a stride forward in the understanding of the noise performance of communication systems and how one might design systems, using simple models, to increase the signal-to-noise ratio.

Carson and Zobel’s paper was followed up a year later by another paper by Carson in the Transactions of the AIEE (American Institute of Electrical Engineers [24] essentially repeating and clarifying the sections on random interference of the earlier paper. (The same paper was reprinted in the 1925 Bell System Technical Journal [25].)⁴ Several significant new points are made and discussed, however. One deduction is that, “even with absolutely ideal selective circuits, an irreducible minimum of interference will be absorbed”, and that the resultant noise output power will increase linearly with bandwidth. Hence the object is to choose a bandwidth as narrow as possible as required to transmit the signal, but no narrower. For, as Carson points out, too narrow a bandwidth results in “sluggishness of [signal] response, with consequent slowing down of the

possible speed of signaling.” Implicit, therefore, in the sense of maximizing the signal – to-noise ratio, is the idea that there exists an optimum receiving system bandwidth, neither too narrow nor too wide. Moreover, as Carson points out, “The only way in which the interference can be reduced, assuming an efficiently designed band filter and a prescribed frequency range [bandwidth], is to select a carrier frequency at which the spectrum $R(\omega)$ is low”. Both ideas, that of choosing an optimum bandwidth and selecting signaling frequencies at which the noise is low, are early versions of the *matched filter* concept, which was introduced 20 years later in the study and design of radar systems during WWII. Carson’s comments on what we now call a matched filter are best summarized in his own words: “Discrimination between signal and interference by means of selective circuits depends on taking advantage of differences in their wave forms and hence on differences in their *frequency spectra*. It is therefore the function of the selective circuit to respond effectively to the range of frequencies essential to the signal while discriminating against all other frequencies.” [Original emphasis.] These thoughts of Carson are prescient and quite remarkable, despite their apparent simplicity.

Summarizing our discussion thus far, by the late 1920s we have the following status of understanding of the impact of noise on communication system performance:

1. The concept of designing systems to maximize a signal-to-noise ratio was well-recognized, with the term “noise” being used interchangeably for static (atmospherics) and fluctuation noise.
2. Workers in the fields of radio and telephony generally knew that noise at the first stage in an electronic system was paramount in determining the signal- to-noise ratio, although radio engineers were most concerned with “atmospherics” or static. Carson and his co-workers in the telephone industry included both static and circuit noise in their discussions of fluctuation noise and its impact on system performance.
3. Measurements of radio noise indicated that the noise decreased with frequency, essentially becoming comparable to fluctuation noise at short-wave frequencies. This point was particularly noted by Armstrong, as we shall see in the next section.
4. The spectrum concept as applied to noise was beginning to be used, despite the unease expressed by some workers as to its applicability to random phenomena. There was a recognition that, at least at radio frequencies, the spectrum of fluctuation noise was flat;
5. Carson had early on implicitly recognized the concept of what was later called the “matched filter” in maximizing the signal-to-noise ratio.

4. Early 1930s: Armstrong and wideband FM

In this section we move ahead to the early 1930s, focusing on Armstrong’s invention of wideband FM . Armstrong’s work was followed a few years later by Alex H. Reeves’ invention of pulse-code modulation (PCM). These two inventions changed the study of the impact of noise on system performance dramatically. Armstrong had recognized, as we shall see shortly, that purposeful widening of the bandwidth of the FM transmission signal, by increasing the frequency deviation, results in a reduction in noise at the communication receiver output. Reeves recognized a similar phenomenon with PCM. Within 10 years of Armstrong’s invention, with the PCM invention coming soon after, it began to be understood by the telecommunications community that both of these

systems were in a class of communication systems for which one could trade increased bandwidth off for improved signal-to-noise ratio. The invention of wideband FM, in particular, resulted in a flock of technical papers attempting to explain the noise improvement obtained, leading in turn to an increased understanding of the modeling of fluctuation noise in communication systems. Both FM and PCM had a revolutionary impact on communications technology as well. Reeves' achievement as well as subsequent advances throughout the 1930s and into the 1940s, including immense advances in the understanding of the impact of noise on systems made during the development of radar during World War II, are not further discussed here. Instead we now turn to the groundbreaking work of Armstrong on wideband FM.

Edwin H. Armstrong had been experimenting for years in his laboratory at Columbia University with various methods of reducing static in radio reception [27], [28]. By 1927 he thought he had come up with a solution, based on many experimental studies carried out in his laboratory. His proposal was to "cancel" static noise, thinking that atmospheric disturbances in a "crash or burst of static" are highly-correlated at closely-adjacent frequencies. His paper describing the technique and including results of some of his experiments was published in the 1928 Proceedings of the IRE [29] and was promptly (and properly) critiqued by John Carson in a paper appearing six months later in the same journal [30]. Carson, in replying to Armstrong's paper, showed, using a simple, single-sinusoidal, model to represent random noise (static), that noise cannot be canceled out. In setting up this model for random noise and carrying out the analysis, Carson relied on his previous work of 1923 to 1925, described here in the previous section. He did, however, make a regrettable and now-famous comment in this paper. In replying to Armstrong's error in thinking noise can be cancelled, Carson stated unequivocally

"Static, like the poor, will always be with us."

Armstrong was to demolish this unfortunate comment by Carson just a few years later with his invention of noise-reducing wide deviation/wideband FM.

How and when did Armstrong arrive at wideband FM as a way to reduce noise in radio systems? These questions are difficult to answer and may, in fact, never be answered exactly. The problem is that Armstrong never kept notes of his conceptions and experiments based on them. All that is generally available are system and circuit diagrams, usually prepared for purposes of proving dates of inventions. In the case of FM specifically, he kept no detailed journals or notebooks of his work in this area, writing down only those results of experiments that were ready for patenting. One can make a good surmise, however, based on a study of what papers are available, that he came to the conclusion that wideband FM would reduce noise in about September of 1931 [31]. He applied for a US patent on his wideband FM invention January 24, 1933, some 16 months later, and the patent, with the simple title *Radiosignaling*, was granted December 26, 1933 with the number 1,941,069.

Armstrong, while working over the years on finding a means to reduce noise, had been heavily involved with the design of FM systems as well. Much of this work on FM systems was carried out in conjunction with RCA engineers, in his capacity as a consultant to RCA [32]. Correspondence between RCA engineers and managers over the years he worked with them indicates that Armstrong worked closely with RCA personnel in perfecting his FM systems, and provided demonstrations for them at his Columbia laboratories [31] [32]. But these demonstrations were, until he was granted the '069

patent, of narrowband FM systems only. His '069 wideband FM patent was one of four FM patents granted to Armstrong the same day. It was the only one dealing with noise suppression, however, The others dealt with such issues as improved means of generating FM (now called the Armstrong system), the use of a limiter in FM, and FM as a way to reduce signal fading. The other patents had application dates ranging from May 18, 1927 to the same date as '069 of January 24, 1933.

The wideband FM patent is quite specific on the noise-suppression property of wide-deviation FM, indicating that this is the essence of the invention. The patent begins with the words “This invention relates to a method of increasing the distance of transmission which may be covered in radio signaling with very short waves. It is well known that waves of the order of ten meters or lower are limited in the distance of transmission by tube noise alone as the amount of static in that part of the spectrum is negligible.” So it is clear that he knows that this method of transmission will reduce tube noise, which he states is “due mainly to the irregularities of the electron emission from the filaments of the vacuum tubes”, i.e., mainly shot noise, but, in the jargon of the day, encompassing thermal noise as well. He does not seem to be sure of its effect on static, however, hence the constraint to operate at higher frequencies for which measurements had shown, as noted in section 2, static would be much lower. He then describes the specific invention as follows: *“I have discovered that by imparting greater swing to the frequency of the transmitted wave than can exist in the disturbances due to tube irregularities and providing means for selecting these large swings of frequency which are at the same time substantially not responsive to the lesser swings due to the tube disturbances or to the variations in amplitude due to these disturbances, that a very great improvement in transmission can be produced.”* [Emphasis added]. This wide “swing”, or deviation, of the transmitted or carrier frequency about its nominal value, the essence of Armstrong’s invention, results in a wider required transmission bandwidth as compared, for example, with amplitude modulation (AM). Armstrong was, of course, aware of the resultant increase in the FM bandwidth. Such an FM system is now referred to as wideband or wide-deviation FM.

Note was made in the previous sections of the fact that much work had been published during the 1920s on shot noise, with knowledge of thermal noise as a fundamental deterrent to communications following along in 1928, after publication of the Johnson and Nyquist papers [3], [4]. As noted earlier, the radio engineering literature had, throughout the 1920s, focused on static, i.e., atmospheric disturbances. It wasn’t until 1930 that papers on shot and thermal noise began to appear in the radio literature. It is clear from patent '069, however, that Armstrong, at the time of the patent application in January 1933, was well-aware of the properties of tube or fluctuation noise. He writes of the “irregularities of the electron emission from the filaments of the vacuum tubes”. He notes the radio frequency noise current, as a result, “consists of irregular variations in amplitude”. The patent further notes that “the limit of reception is... determined by tube noise or the disturbances which arise usually in the first tube in the receiving system”, this “interference manifest[ing] itself as a steady hiss in the telephones or receiver”, and exhibiting “a continuous spectrum of substantially constant amplitude...” These properties of random or fluctuation noise are precisely those described in the work in the 1920s of Carson and other investigators, as noted in the previous section. The only

ambiguity appearing in the patent appears to be on the differentiation made between the properties of static and fluctuation (tube) noise.

Armstrong's invention of wide-deviation/wideband FM and its ability to suppress noise was well-received by the engineering community. Very soon after being awarded patent '069, Armstrong demonstrated his wide-deviation FM system for the first time to RCA engineers [31], [32] with impressive results. Armstrong's later demonstration of his system before an IRE audience on November 5, 1935, and its very positive impact on the engineers attending, has been well-documented [27]. The demonstration accompanied a formal paper presented at the meeting, which was later published as Armstrong's now-classic paper "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation"[33]. What is particularly remarkable about this paper is that Armstrong was able to analytically demonstrate the noise-reduction effect of wide-deviation FM, given a carrier-to-noise threshold had been exceeded, using a vector approach. Soon thereafter, M. G. Crosby of RCA, who had worked for years on FM systems, published a more mathematical paper on the noise-suppression effect of FM [34], followed by papers by Hans Roder of GE [35], as well as Carson and Thornton C. Fry of Bell Labs [36]. Other papers by various investigators soon followed. It is of interest to note that through the years, even to this day, papers by various investigators have continued to be published using different techniques to explain the wide-deviation FM noise-suppression property. Armstrong had created quite a cottage industry on explanations for this property of wide-deviation FM! (The reception of FM signals in the presence of noise is a highly non-linear process, involving, for example, the required use of a limiter. Analysis thus does not lend itself to a tidy mathematical approach.)

Notes

¹ An example appears in [6] in which the authors, all from AT&T, state that at radio transmission frequencies where radio noise is often "practically absent", measurements of noise "tended to approach the minimum determined by the set [tube] noise." The non-radio-specific Electrical Engineering literature does show that engineers were familiar with shot noise even as early as 1920. In the January 1920 issue of the British Journal of Electrical Engineers, for example, in a paper on amplifier design using triode tubes, a number of references are made to "valve noise", the British term for tube noise [7]. The author here notes that atmospheric disturbances are the prime source of interference, but that valve noise "constitute[s] one of the most formidable obstacles to the attainment of very high amplification" [7, p.69] It is interesting to note that this paper was published only two years after Schottky's paper [1], the latter written in German and the work for it done presumably during World War I.

² The term "long-wave" refers to measurements made in the frequency range 17 kHz to 57 kHz, for which the corresponding transmission wavelengths ranged from 18,000 meters down to 5300 meters. Note that, in the radio world, the product of wavelength and frequency is the speed of light, 300 million meters/sec. These "long-wave" wavelengths compare to "short-wave" wavelengths of 100 meters down to 10 meters, corresponding to a frequency range of 3-30 MHz. (M stands for "mega", or million, while k stands for "kilo", or thousand.)

³A Fourier integral representation of random impulses, the “individual impressed forces”, such as adopted for the random interference model here, must have been controversial at the time: Thornton Fry, a mathematician and colleague of Carson and Zobel at AT&T, stated, in a 1925 paper on the shot effect [21], that the Fourier representation of a *random* disturbance does not exist. Schottky, however, in a 1926 paper [22], dismissed Fry’s objection and indicated that the Fourier representation of shot noise is valid. Researchers at GE, among others, used a Fourier representation of shot noise in 1925 [23] in comparing theory with experiment in shot noise studies. Johnson and Nyquist used the spectral representation of thermal noise without qualms in their 1928 papers as well.

⁴G. W. Kenrick of the University of Pennsylvania, in a 1929 paper [26] “communicated by Carson” to the journal in which it appeared, referenced the 1925 BSTJ paper by Carson in calculating the spectrum of various models for static and random telegraph signals. It is thus clear that the concept of a Fourier representation for random signals was generally accepted by this time despite earlier misgivings.

References

- [1] W. Schottky, Uber spontane Stromschwankungen in verschiedenen Elektrizitätsleitern, Annalen der Physik, verte folge, Band 57, 1918, 541-567.
- [2] J. B. Johnson, Electronic Noise: the first two decades, IEEE Spectrum, Feb. 1971, 42-46.
- [3] J. B. Johnson, Thermal Agitation of Electricity, Physical Review, vol. 32, July 1928, 97-109.
- [4] H. Nyquist, Thermal Agitation of Electric Charges in Conductors, Physical Review, vol. 32, July 1928, 110-113.
- [5] L. Cohen, The History of Noise, IEEE Signal Processing Magazine, Nov. 2005, 20-45.
- [6] Austin Bailey et al, The Receiving System for Long-Wave Trans-Atlantic Radio Telephony, Proc. IRE, vol. 16, no. 12, Dec. 1928, 1645-1705.
- [7] C. L. Fortesque, The Design of Multiple-Stage Amplifiers Using Three-Electrode Thermionic Valves, J. IEE, vol. 58, no. 287, Jan. 1920, 65-82.
- [8] F. B. Llewellyn, A Study of Noise in Vacuum Tubes and Attached Circuits, Proc. IRE, vol. 18, no. 2, Feb. 1930, 243-265.
- [9] G. W. Pickard, Static Elimination by Directional Reception, Proc. IRE, vol. 8, no. 5, Oct. 1920, 358-394.

- [10] G.. W. Pickard, The Correlation of Radio Reception with Solar Activity and Terrestrial Magnetism, Proc. IRE, vol. 15, no.2, Feb. 1927, 83-97.
- [11] Harold H. Beverage et al, The Wave Antenna- A New Type of Highly Directional Antenna, Trans. AIEE, vol. 42, 1923, p.215.
- [12] Frederick E. Terman, Radio Engineering, 2nd ed., McGraw-Hill, NY, 1937.
- [13] R. A. Weagant, Reception through Static and Interference, Proc. IRE, vol. 7, issue 3, 1919, 207-244; Discussion, 245-266.
- [14] Ralph Brown et al, Radio Transmission Experiments, Proc. IRE, vol. 11, no. 3, April 1923, 115-152.
- [15] L.W. Austin, The Relation Between Atmospheric Disturbances and Wave Length in Radio Reception, Proc. IRE, vol. 9, no. 1, Feb. 1921, 28-35; Discussion, 36-40; The Reduction of Atmospherics in Radio Reception, same issue, 41-55.
- [16] Lloyd Espenscheid et al, Transatlantic Radio Telephone Transmission, Proc. IRE, vol. 14, no. 1, Feb. 1926, 7-56.
- [17] R.A. Heising et al, Some Measurements of Short-Wave Transmission, Proc. IRE, vol. 14, no. 5, Oct. 1926, 613-647.
- [18] John R. Carson, Signal-to-Static-Interference Ratio in Radio Telephony, Proc. IRE, vol. 11, June 1923, 271-274.
- [19] John R. Carson and Otto J. Zobel, Transient Oscillations in Electric Wave-Filters, Bell System Technical Journal (BSTJ), vol. 2, July 1923, 1-29.
- [20] Norman Campbell, The study of discontinuous phenomena, Math. Proc. Cambridge Philo. Soc., vol. 15, 1909, 117-136; Discontinuities in light emission, *ibid.*, 1909, 310-328.
- [21] Thornton C. Fry, The Theory of the Schrotteffekt, J. Franklin Inst., vol. 199, Feb. 1925, 203-220; *see* pp. 204-206 specifically.
- [22] W. Schottky, Small-Shot Effect and Flicker Effect, Physical Review, vol. 28, July 1926, 74-85; *see* pp. 84, 85 specifically.
- [23] A. W. Hull and N. H. Williams, Determination of Elementary Charge e from Measurements of Shot-Effect, Physical Review, vol. 25, Feb.1925, 147-173.
- [24] John R. Carson, Selective Circuits and Static Interference, Trans. AIEE, vol. 43, June 1924, 789-797.

- [25] John R. Carson, Selective Circuits and Static Interference, Bell System Technical Journal, vol. 4, April 1925, 265-279.
- [26] G. W. Kenrick, The Analysis of Irregular Motions with Applications to the Energy-Frequency Spectrum of Static and of Telegraph Signals, London, Edinburgh, and Dublin Philosophical Magazine and J. of Science, vol. VII, series 7, Jan. 1929, 176-196.
- [27] Lawrence Lessing, Man of High Fidelity, J.B. Lippincott Co., Phila., 1956; Bantom Paperback, 1969.
- [28] Report from Edwin H. Armstrong to Michael Pupin, April 3, 1933, Box 115, Armstrong papers, Rare Book Library, Columbia University.
- [29] Edwin H. Armstrong, Methods of Reducing the Effect of Atmospheric Disturbances, Proc. IRE, vol. 16, no. 1, Jan. 1928, 15-26.
- [30] John R. Carson, Reduction of Atmospherics, Proc. IRE, vol. 16, no. 7, July 1928, 966-975.
- [31] M. Schwartz, Armstrong's Invention of Noise-Suppressing FM, submitted for publication.
- [32] Gary Lewis Frost, The Evolution of Frequency Modulation Radio, 1902-1940, Ph. D. dissertation, University of North Carolina, Chapel Hill, 2004.
- [33] E. H. Armstrong, A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, Proc. IRE, vol. 24, no. 5, May 1936, 689-740.
- [34] M. G. Crosby, Frequency Modulation Noise Characteristics, Proc. IRE, vol. 25, no. 4, April 1937, 472-514.
- [35] Hans Roder, Noise in Frequency Modulation, Electronics, May 1937, 22-25; 60-64.
- [36] John R. Carson and Thornton C. Fry, Variable Frequency Electrical Circuit Theory with Application to the Theory of Frequency Modulation, BSTJ, vol. 16, Oct 1937, 513-540.