U.D.C. 621.376.3(091)

The Invention of Frequency Modulation in 1902

By

Professor D. G. TUCKER,
D.Sc., C.Eng., F.I.E.E., F.I.E.R.E.†

It is shown that the accounts in the literature of the early history of frequency modulation are very confused and contradictory. After demonstrating that Fessenden in his 1901 patent application almost certainly did not intend to have or to utilize f.m., but was concerned only with a.m., it is shown that the real first inventor of f.m. was C. D. Ehret with his very clearly stated patent application of 1902.

1. Introduction

Frequency modulation (f.m.) is now a very important process in radio communication, and is widely used in preference to amplitude modulation (a.m.) at the higher radio frequencies (e.g. v.h.f.) where the availability of adequate bandwidth enables better signal/noise discrimination to be obtained by the use of f.m. Its introduction as an important process in the communication of speech and music is generally, and correctly, credited to Armstrong,[1] who published his comprehensive paper in 1936 following extensive pioneering experiments from 1933 onwards. In this paper he described technical means for achieving f.m. communication, he reviewed the state of knowledge on noise interference in radio and on thermal agitation (or 'Johnson') noise[2] as a fundamental limitation on communication, and he described for the first time the special noise-discriminating properties of f.m.

It is also well known that Carson[3] in 1922 had analysed for the first time the sideband structure of f.m., and had shown that, unlike a.m., f.m. had an infinite set of sidebands. Above all he showed that:

'...the transmission of the signal by frequency modulation requires the transmission of a band of frequencies *at least* $2p/2\pi$ in width; *that is* a band of width equal to twice that of the signal itself.'

Here p is the signal frequency in rad/s.

He alleged, and subsequent writers, e.g. Heising,[4] have repeated, that earlier workers had believed that f.m. would permit speech to be transmitted in a *narrower* bandwidth than that required by a.m. His work certainly should have disposed of this fallacy, but Heising states that there was subsequent work done by others in apparent ignorance of Carson's conclusions.

When, however, we try to discover the earlier history of f.m., that is, before Carson, we find a very confused picture in the literature. Contradictions abound, as will be demonstrated below. The purpose of the present paper is to clarify the origins of f.m. It will be shown that f.m. was invented in 1902 by C. D. Ehret.

2. The Confusion

Various writers who have discussed the history of f.m. give various accounts of its early development, and attribute the basic invention differently. To illustrate this, it will be best to let the various writers speak for themselves. Heising, who was himself an early worker in the field of communications,[5] has recently written:[6]

> 'The earliest discussion on f.m. in telephony is a short item in a book by Dr. A. N. Goldsmith in 1918.'

Yet in the text-book by Ruhmer[7] (probably the first ever published on wireless telephony, appearing in German in 1907 and in English translation by Erskine-Murray in 1908) we read:

> 'The conversion of the sound waves may of course take place in any of the ways already described, and may depend on the variation of either the intensity or frequency of the oscillations. In both groups of methods the transmission of speech is by means of varying electrical waves, in the first case with constant, and in the second with varying frequency, which act with varying intensity on the detector in the receiving circuit and thus reproduce the corresponding sounds in the telephone attached to it.
>
> 'When a constant wave-length is used the variations of the intensity produce similar variations in the receiver, while if transmission depends on variation of the wavelength, the receiver must be a persistent oscillator which only responds when acted upon by waves which are of its own natural frequency, and falls out of resonance when the frequency of the impressed waves varies.'

This looks very much like a description of f.m., although the second paragraph has a suggestion that Ruhmer may have had his thoughts rather more on frequency-shift keying in telegraphy than on telephony. In this connexion it is worth stating that the keying of telegraph signals by frequency-shift was an established practice in those days,[8] and may be regarded

† Department of Electronic and Electrical Engineering, University of Birmingham.

as a very crude kind of f.m.

In his classic paper, already referred to, Armstrong[1] says:

> 'The subject of frequency modulation is a very old one. While there are some vague suggestions of an earlier date, it appears to have had its origin shortly after the invention of the Poulsen arc, when the inability to key the arc in accordance with the practice of the spark transmitter forced a new method of modulation into existence.'

This clearly is equating f.m. with frequency-shift telegraphy and erroneously placing the date of origin after 1903, which was the year in which the Poulsen arc was introduced.[9]

Some uncertainty in the exact meaning of the wording in the patent specification filed in 1901 by the famous early wireless engineer Fessenden,[10] who is usually and probably incorrectly credited with the invention of a.m., has led to some suggestion that he also had f.m. in mind. Heising[11] says:

> 'Fessenden, in 1902, proposed using a condenser type of microphone in circuits that appear to produce f.m., but in his published remarks about tests he seemed interested in using the frequency variation to produce amplitude variation by throwing the frequency in and out of resonance with the antenna or other tuned circuit to modulate the amplitude.'

The same idea that Fessenden was proposing frequency modulation, although merely to obtain amplitude modulation, appears in Ruhmer's book[12] of 1907, already referred to. He says:

> 'In another of Fessenden's senders the control of the radiation is effected by alteration of the frequency through variation of the inductance or capacity of the aerial system.'

The present author has long held the view[13] that Fessenden did not actually intend the frequency of the radiated signal to vary significantly as the speech amplitude varied, and that the inductance (or capacitance) which varied in value as the speech amplitude varied was intended as what we would now call a magnetic (or a dielectric) amplitude-modulator. This was also the view of Pungs[14] in 1923. Referring to Fessenden's patent[10] he says:

> 'Für die Steuerung von Hochfrequenzströmen in das Verfahren zuerst von Fessenden (1902) vorgeschlagen worden, jedoch in einer praktisch nicht ausführbaren Form (Ringspule mit kurzgeschlossener Steuerwicklung).'†

† 'For the control of high-frequency currents the arrangement proposed by Fessenden in 1902 was the first, although in a form not workable in practice (ring-coil with short-circuited control winding).'

As it is rather important to settle this matter, the relevant drawings from Fessenden's patent and the relevant part of his description are reproduced in the Appendix.

The important words here are those put into italics: 'sustained oscillations of practically constant frequency'. It seems clear that Fessenden intended to produce a.m. (not f.m.) by causing the speech amplitude to control the division of current between the low-loss tuned circuit (of constant parameters) and the more lossy tuned aerial (of variable parameters). Of course, there would have been some frequency variation, but this was merely a stray undesired effect. It was not an efficient arrangement, quite apart from the fault mentioned in Pungs' critical comment, quoted above.

If this interpretation, that Fessenden had no idea of using f.m., is accepted, then the field is left clear for the true inventor, Cornelius D. Ehret of the U.S.A., who filed his patents in 1902. The clue to Ehret's status in this regard was given by Guy,[15] who, however, gave no reference to the documents concerned.‡

3. Ehret's Patents on F. M., 1902

Ehret's patent on f.m. was divided into two almost identical specifications,[16,17] one dealing with the method and the other with the systems. There can be no doubt whatever that he describes f.m. as now understood:

> 'It comprises, further, a method of modifying and varying the frequency of the electroradiant energy in a manner corresponding and in accordance with the signal to be transmitted.
>
> 'It resides also in an additional step of modifying the energy to be transmitted and received by and in accordance with sound-waves, such as speech.
>
> 'It comprises, further, a method of receiving the modified transmitted energy and causing the reproduction of speech and other signals by the effects of variations or changes in the frequency of the received energy.'

His claim is a very concise definition of f.m.:

> 'The method of transmitting intelligence, which

‡ The author is grateful to one of the referees for pointing out that Ehret's patents on f.m. are also cited, albeit very casually and without discussion, by Hammond and Purington,[30] and by Miessner[31] who incorrectly states that Ehret was concerned only with frequency-shift telegraphy. Unfortunately these additional references, while interesting in their own right as personal statements by authors who themselves contributed greatly to the development of radio, nevertheless tend to increase the confusion already commented on by the present author.

consists in generating electroradiant energy, modifying the frequency of said energy in accordance with the signal to be sent and receiving the energy in a device responsive to changes in the frequency of the transmitted energy.'

The process of modulation was to modify the frequency of the second stage of a two-stage spark transmitter, e.g. by connecting the microphone across a tuning inductance. The receiver used the variation of voltage across an inductance connected in series with the aerial or across a shunt-tuned circuit coupled to it. It is perhaps unlikely that the system could have worked well with Ehret's proposed circuits, but the idea of the system is perfectly clear and correct. The patent also very clearly covers frequency-shift telegraphy. It is, moreover, so worded as to suggest that Ehret thought he was the first to discover any way of achieving the transmission of speech by wireless, although we have seen above that Fessenden had filed the patents for his method the year before. Ehret's words are:

'. . .it is apparent that I have disclosed a method of transmitting speech electrically without the employment of conductors joining the transmitting and receiving stations; that speech is transmitted by this method by the agency of electroradiant energy, as employed heretofore for telegraphy only; . . .'

4. Some Additional Notes on Ehret's Work

As far as can be ascertained from the U.S. Patent files, Cornelius D. Ehret was awarded 35 U.S. patents, from 1902 to 1933, with eight preceding his f.m. patents. Of the 35, four were for inventions not connected with electricity or communications: centrifugal machine, 1908; making wire-glass, 1914; optical pyrometry, 1920; and method and apparatus for compressing fluids, 1928. None of the patents preceding the f.m. disclosure were particularly related to it, except that one[21] which was filed around the same time shows a system in which the transmitted frequency is changed in synchronism with the tuning of the receiver, at quite high speed, to make a secrecy system. In the same year as the f.m. patents were issued, Ehret also had patents[22,23] granted covering telephone line repeaters and frequency-division multiplex.

It is interesting that at intervals other inventors with the name Ehret occur in the U.S. Patent Indexes, and as some of these give Philadelphia as their address (as does C. D. Ehret in many of his patents), one cannot help wondering if there was a whole family of inventors. C. D. Ehret himself took the assignment of several non-electrical patents, so he must have varied business interests. Guy[15] describes him as a Philadelphia patent attorney.

It is interesting that Ehret has almost completely escaped the attention of those writing on the history of radio. Apart from books and papers already cited, there are several others which give a substantial historical review of the early days of radio and communications generally (e.g., Refs. 24–28), and as far as the present author can detect, no reference to Ehret or his ideas is given in any of them.

It is probable that Ehret did not publish his ideas in any form other than patent specifications, as the only contribution he made to the American Institute of Electrical Engineers during the years 1902–1910 inclusive was a group of three short questions[29] on the audion at a meeting in Philadelphia. The author has found no evidence that Ehret carried out any experimental work on his ideas. It was to be thirty years before practical significance was to be attached to f.m., and it is perhaps understandable that during this time Ehret's patent should be forgotten. Even at the time of issue it may well have had no impact, as it was clearly premature; the technology of 1902–5 did not give it a real chance of application then.

5. The Word 'Modulation'

Guy[15] seems to suggest that Ehret actually used the word 'modulation' in his patents. Guy's words are:

'So far as is known these were the first disclosures to describe any system of modulation by name.'

But the word 'modulation' does *not* appear in Ehret's patents, except once (in the plural) in the traditional meaning of the fluctuations of the voice itself: never in connexion with the electrical system.

Heising[18] attributes the introduction of the word 'modulation' to Fessenden in an article of 1907.[19] The relevant sentence by Fessenden is:

'The difficult problem in wireless telephony is, of course, the modulation of the large amount of energy used for transmission.'

No explanation of the word is given, and one cannot help feeling that the author was under the impression that the word was generally understood in this sort of context.

6. References

1. Armstrong, E. H., 'A method of reducing disturbances in radio signalling by a system of frequency modulation', *Proc. I.R.E.*, **24**, pp. 689–740, May 1936.
2. Johnson, J. B., 'Thermal agitation of electricity in conductors', *Phys. Rev.* **32**, p. 97, July 1928.
3. Carson, J. R., 'Notes on the theory of modulation', *Proc. I.R.E.*, **10**, pp. 57–64, February 1922.
4. Heising, R. A., 'Modulation methods', *Proc. I.R.E.*, **50**, pp. 896–901, May 1962 (see p. 899).
5. For short biography see *Proc. I.R.E.*, **50**, p. 1430, May 1962.
6. Heising, R. A., *loc. cit.*, p. 897.

7. Ruhmer, Ernst, 'Wireless Telephony in Theory and Practice', translated from the German by J. Erskine-Murray. (Crosby Lockwood & Son, London, 1908). (Published in German 1907). See pp. 136–7.
8. Weldon, J. O., 'Transmitters', *Proc. I.R.E.*, **50**, pp. 901–906, May 1962 (see p. 903).
9. Poulsen, V., 'Method of producing alternating currents with a high number of vibrations', U.S. Patent No. 789,449, issued 9th May 1905, filed 19th June 1903.
10. Fessenden, R. A., 'Wireless signaling', U.S. Patent No. 753,863, filed 28th September 1901, patented 8th March 1904.
11. Heising, R. A., *loc. cit.*, pp. 896–7.
12. Ruhmer, E., *loc. cit.*, p. 140.
13. Tucker, D. G., 'Modulators and Frequency Changers', p. 93. (Macdonald, London, 1953).
14. Pungs, L., 'Die Steuerung von Hochfrequenzströmen durch Eisendrosseln mit überlagerter Magnetisierung', *Elektrotechnische Zeitschrift*, **4**, pp. 78–81, 1923 (see p. 78).
15. Guy, R. F., 'A.M. and F.M. Broadcasting', *Proc. I.R.E.*, **50**, pp. 811–817, May 1962 (see p. 816).
16. Ehret, C. D., 'Art of transmitting intelligence', U.S. Patent No. 785,803, filed 10th February 1902, patented 28th March 1905.
17. Ehret, C. D., 'System of transmitting intelligence', U.S. Patent No. 785,804, filed 10th February 1902, patented 28th March 1905.
18. Heising, R. A., *loc. cit.*, p. 896.
19. Fessenden, R. A., 'Wireless Telephony', *Electrical Review*, **60**, pp. 251–253, 15th February 1907. (N.B. This was Part 1 of a 3-part article; the other parts were pp. 327–329 and 368–370 of the volume).
20. Palmer, L. S., 'Wireless Principles and Practice', p. 400. (Longmans, Green, London, 1928).
21. Ehret, C. D., 'Wireless selective signalling system', U.S. Patent No. 717,978, issued 6th January 1903, filed 3rd July 1902.
22. Ehret, C. D., 'Art of electrical wave transmission', U.S. Patent No. 789,087, issued 2nd May 1905, filed 3rd December 1901.
23. Ehret, C. D., 'System of electrical wave transmission', U.S. Patent No. 789,124, issued 2nd May 1905, filed 3rd December 1901.
24. Armstrong, E. H., 'Evolution of frequency modulation', *Electrical Engineering*, **59**, p. 485, December 1940.
25. Eccles, W. H., 'Wireless'. (Thornton Butterworth, London, 1933).
26. Blake, G. G., 'History of Radio Telegraphy & Telephony', (Radio Press, London, 1926).
27. Colpitts, E. H. and Blackwell, O. B., 'Carrier current telephony and telegraphy', *Trans. Amer. Inst. Elect. Engrs.* **40**, pp. 205–300 (1921).
28. Dunlap, O. E., 'Radio's 100 Men of Science'. (Harper, New York and London, 1944).
29. Ehret, C. D., *Trans. Amer. Inst. Elect. Engrs.* **25**, pp. 775, 776, 778 (1906).
30. Hammond, J. H. and Purington, E. S., 'A history of some foundations of modern radio-electronic technology', *Proc. I.R.E.*, **45**, pp. 1191–1208, 1957 (see p. 1205).
31. Miessner, B. F., 'On the Early History of Radio Guidance'. (San Francisco Press, 1964) (see p. 60).

7. Appendix

Extract from Fessenden's Patent Specification

'In the practice of my invention I provide at the sending-station a conductor 1 of suitable construction and arrangement and connect the same to one terminal of a coil 2, surrounding a core 3, preferably annular in shape and preferably formed of fine iron wire. The other terminal of the coil is connected to one of the knobs or terminals 4 of an induction-coil or other suitable generator 6 capable of producing practically continuous and rapid oscillations in the conductor. The systems described in connection with Fig. 10 of Patent No. 706,742 or in Patent No. 730,753 are well adapted for this purpose. The other terminal, 5, of the generator is connected to ground. A second coil 7, forming a part of the circuit for the battery 8, is placed on the core 3, and a transmitter 9, preferably microphonic in construction, or other mechanism capable of modifying the current in the circuit is included in the circuit of the battery and coil 7. A capacity 18 and inductance 19 are connected in shunt to the spark-gap for the purpose of maintaining *sustained oscillations of practically constant frequency*. [Italics due to present author.] The capacity 18 and inductance 19 should be arranged to have the same period of oscillation as the receiving-conductor 10 and the sending-conductor 1. It will be seen that the circuit containing capacity 18 and inductance 19 being connected across the spark-gap, forms a parallel circuit in the sending conductor 1, whose aerial and grounded sections are also connected across the spark-gap. On account of the fact that the circuit 18, 19, and the sending-conductor 1 are in parallel and not in series the difference of potential across these two circuits is the same, while the currents in the two circuits are different, this construction being thus differentiated from a series connection, in which the circuit 18, 19, would be connected electrically between the aerial portion of the sending-conductor 1 and the ground.

'At the receiving-station I employ a conductor 10, connected to one terminal of the mechanism capable of responding to oscillations in the conductor 10. A form of mechanism adapted to the purpose consists of a coil 11, having one terminal connected to the conductor and the other terminal grounded. A telephone-diaphragm 12, formed of metal or consisting of insulating material having a metal plate or coil of wire secured thereto, or any other suitable construction adapted to vibrate in unison with changes of current or voltages produced by the waves radiated from the sending-station is suitably supported in operative relation to the coil 11. The apparatus at the receiving-station is tuned or made resonant by any suitable means known in the art to the sending-conductor 1. The terms "tuned" and "resonant" are used herein, one to include the other. When an alternating current is set up in the conductor 10, as by waves or impulses from the sending-station, such current acts to repel or attract the diaphragm, according to the time constant of the metal part of the diaphragm, through induced currents set up in the diaphragm. When the generator is operated, the diaphragm 12 will take up a mean position relative to coil 11, the distance of such position from the coil varying with the intensity of the oscillations in the sending-conductor; but

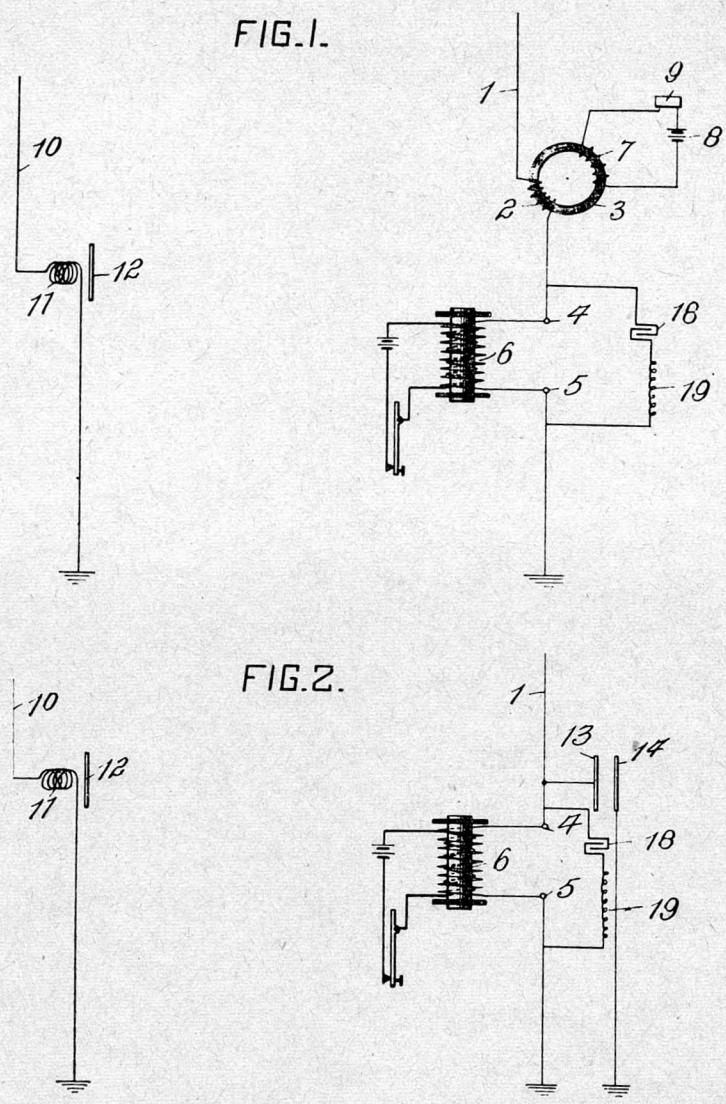

Fig. A. From R. A. Fessenden's U.S. Patent of 1904.[10]

when the current in the circuit of the coil 7 is modified or changed by speaking into the transmitter the permeability of the core 3 is correspondingly changed or modified, thereby producing a corresponding change or modification in the self-inductance and a change in the frequency of the natural period of vibration of the sending-conductor 1, which is thereby thrown out of resonance with a resonating-circuit 18, 19, connected in parallel to said sending-conductor 1, and due to this failure in resonance producing a corresponding change or modification in the intensity of the waves or impulses given off by the conductor 1 and in the intensity of the oscillations produced in the receiving-conductor. The changes in the intensity of the oscillations will produce corresponding changes in the mean position of the diaphragm, such changes corresponding to the vibrations of the diaphragm of the transmitter, exactly reproducing any of the waves or impulses which affected the transmitter. The same result may be effected by changing the capacity of the conductor 1, as shown in Fig. 2. To this end the conductor 1 is connected to a fixed plate 13 of a condenser, while the other plate, 14, is formed by or connected to a diaphragm capable of responding to waves or impulses. As the plate 14 in vibrating moves toward or from the other plate the capacity of the conductor 1 is changed, correspondingly altering the intensity of the waves or impulses generated by the conductor.'

Manuscript received by the Institution on 6th February 1970. (Paper No.1331/Com. 31.)