VOCOM
VOCOM

THE TECHNICAL BACKGROUND

- Introduction
  - The nature of sound is vibratory. For us to hear something, the air must move. It is possible to move other media in the same way, which results in similar but non-acoustic phenomena. These equally vibratory but different forms of "sound" are familiar to all of us as (1) electrical signals in a wire, (2) physical movements of a stylus in a groove, (3) changing magnetic fields from a moving tape, (4) audio modulated high frequency radiation (amplitude or frequency modulated radio). In all these forms the vibratory movements are analogous to the original acoustic form. If they are not, the result when we eventually reconvert to acoustic vibrations (by means of a loudspeaker) is described as "distorted". But our aim, at any rate in high fidelity systems, is to preserve a true analog of the complicated and continuously changing vibratory function of the original music or speech. All of the above methods are analog methods.
  - Vocom is concerned with a different way of looking at information. Vocom processes sound in digital form, which means numbers. What happens is that the electrical analog of a sound is examined by taking samples at regular intorvals and noting its amplitude at each sampling point. If we assume that the vibrations or oscillations move around a zero line, some of the numbers noted will be positive, others negative. The result is a list of numbers all of which have a known time position. For example a pure (sine) nund might have a frequency (speed of vibration) of 100Hz (Hertz cycles per second). This is about G at the bottom of the bass stave, and one cycle takes 0.01 seconds. For half the time (0.005s) it is positive and for the other half negative, and the zero line is crossed twice. If we set a clock to sample this waveform every 0.0001s (100 microseconds-a frequency of 10KHz), every cycle will produce 100 numbers, and according to the length of the number (in the case of a computer the number of binary digits or bits) we would have greater or less discrimination in the accuracy of the numbers given. In decimal terms, the number series 1.003, 2.095, 3.416 gives much more accurate information than 1, 2, 3 because we allowed four digits instead of one. So the more bits in the number, the better the approximation to the true amplitude sampled.
  - A moment's thought will also show that the higher the frequency being sampled, the higher the sampling rate must be. In short, the two important parameters are the rate of sampling and the number of bits in the sample taken. The product of these two gives us a bit-rate per second, and in general the higher the bit-rate the better the fidelity when the numbers are later decoded into analog form.

VOCOM

I.V.M. (England)
51 Deodar Road, London, SW15 2NU
Telephone: 01-874 0042 Telex 928372

I.V.M. (U.S.A.)
408 East 78th Street, N.Y. 10021
Telephone: 212 734 7344 Telex 424033

- Why Use Digital Techniques?
  - Why bother to convert waveforms into numbers? What's wrong with my analog phonograph, tape recorder? Nothing at all in the context for which they were intended, but a lot when it comes to sending a vast amount of data over a great distance - and this is after all what Vocom is, about. The compelling reasons for using digital techniques lie in two areas - Transmission and Storage.
  - Transmission: In order to send analog signals over long distances with high fidelity, extremely expensive lines or extra high quality radio links must be provided. Long ordinary telephone lines cause drastic attenuation, particularly in the high frequencies. To overcome this, the necessary coaxial or other special cables are prohibitively costly except where quality must predominate over price (e.g. programme links between studio and transmitter). VHF/FM transmitters give adequate quality but have only short range. The quality of long distance AM radio is still appalling after over 60 years of shortwave - subject to fading, noise, cosmic and meteorological interference. And anyway on AM the normal station separation cannot give adequate bandwidth (see below) for high fidelity. But consider the transmission of binary numbers. Even if the quality is poor, the receiving end need only be able to detect whether a 0 or a 1 was sent, not the elaborate waveforms of speech or music (though there is a limit to the speed of data trans'mission, similar to the high frequency analog limit). Much poorer signals can be acceptably decoded, and since no noise or interference is in the data sent, the output will also be free from them - they are analog phenomena, outside the competence of a digital converter and therefore ignored.
  - The advantages of digital methods for speech transmission were recognised some time ago, and Pulse Code Modulation, using a form of the methods I have described, carries most of the long distance calls you make today. Without digital transmission, pictures from the moon would be a complete impossibility Naturally, in a telephone system we are not concerned with the very high data rates needed for good TV pictures. or even with hi-fi in the music sense. If we can clearly hear a noise-free voice and recognise the vocal idiosyncrasies of our caller, we are quite happy to accept certain limitations. Indeed, we have always tolerated the very low quality microphone used in telephones, which limits the possible fidelity right at the beginning of the chain. Nevertheless, in order to produce acceptable quality even by telephone standards, the PCM system used by the British GPO has to use a rate varying between 48,000 and 64,000 bits per second.
  - Storage: Analog storage on tape, film or disc is fine - but only in the applications best suited to it. Even in these applications there are plenty of poor tapes and scratchy discs, and they are bulky and cumbersome to load, and slow to start, find cues in, take off, file or destroy. For any automated system digital storage offers compelling advantages, but for Vocom it is absolutely necessary. Using the various digital equivalents of analog stores, very large amounts of data can be stored economically, reached in tiny fractions of a second, processed at speeds thousands of times faster than 'real time', sorted, re-routed, edited, reassembled, acted on by purely mathematical means, and when not needed, destoyed without fuss. Naturally some kinds of digital storage are cheaper than others -

it largely depends whether you need really rapid access to your data - but whereas you cannot significantly reduce the amount of data you must send to an analog store without degrading your signal beyond repair, with Vocom you can make drastic reductions in digital storage needs, so that the 'cost per bit' comes down to far below even the cheapest type of analog storage. Vocom reduces the PCM requirement of at least 48,000 bits/second to a mere 200 b/s. Our current research is at this moment bringing it down even further to 110 b/s, which is the standard Telex bit rate. Imagine a voice output on a Telex line! The bit-rate reduction compared with PCM is over 400:1.

- In analog terms, the greater the frequency range the wider the 'bandwidth' - the spectrum or 'band' to which the whole system must respond if fidelity is to be maintained. In general the wider the bandwidth the greater the expense, and in telecommunications where the highest quality may not be necessary but every penny counts, designers have been experimenting for years with different ways of retaining adequate quality while restricting bandwidth and hence cutting operating costs. Another way of looking at bandwidth in the telephone context is the number of separate signals which can be carried in one wire - but in either case there is a certain level of data flow which we wish either to reduce or to distribute more efficiently. Vocom does both, in ways more effective than any conceived before, as well as adding facilities not dreamed of in previous experimental proposals.

- To understand the methods so far attempted, and give the background for the extraordinary breakthrough achieved by Vocom, here is a brief summary of the main alternatives: Bandwidth ConservationS Compression Systems
  - Bandwith conservation systems, in their area of most immediate application, would enable more efficient use to be made of telephone lines, especially long distance cables and satellites. As these represent multibiJlion dollar investments, any electronic 'Black Box' which fractionally increased the return on them would seem worthwhile; particularly so if that fraction approached its theoretical value of 12,0000%. Four divergent lines of research have been followed for many years, and we here briefly describe the essence of each method: 1. Pause stuffing. 2. Coarse amplitude quantisation. 3. Time/frequency compression. 4. Voice analog vocoders.
    - 1. Pause Stuffing, properly called time assignment speech interpolation, is already widely used. It uses small pauses between words and the larger pauses, while the other man is talking, to squeeze in syllables from someone else's conversation. The system works quite well, achieving economies on a certain 36 channel system of between 200 and 300 percent. The apparatus is quite complicated, and it seems clear that further developments of it can only yield marginal improvements.
    - 2. Coarse Amplitude Quantisatlon. This system uses a gross distortion of the speech waveform to make it easier to transmit, and in a sense to make it louder. The distortion is in fact very similar to the rock and roll guitarist's 'fuzz box', and the speech output also has that nasty grainy sound. Researchers have repeatedly proved that while the quality is low, the intelligibility remains high. A great number of developments and variations of the system have been tried, and there is scope for more work. The greatest advantage of it is its cheapness. It is much used by public service mobile radio networks. The most sophisticated development of it has demonstrated potential economies of just over 200%.
    - 3. Time/Frequency Compression. This is a signal treatment process which essentially chops arbitrary sections out of a speech waveform and discards them. It stretches the remaining portions and joins them as best it can. In the stretching process all the frequencies are reduced proportionately, so that the channel requirement becomes less. Until recently the system could only be realised with a cumbersome and expensive revolving head tape machine. Attempts are now being made to make a solid state version, using the new controverslal charge transfer delay line. These devices are not quite the same as the L.S.I. (large scale integration) chips used so successfully in pocket calculators and computers, as they are analog rather than digital which means they are much harder to produce and test. The company which has announced licensing agreements for a low cost system is concentrating on an entirely new market, that of variable speed tape replay, and has made no mention in its ample pLiblicity of the bandwidth reduction potential of the system. This is very difficult to estimate theoretically as it depends almost entirely on the loss of quality and intelligibility caused by the discarding process. Experiments done by David and McDonald in 1956 and Fairbanks, Everitt, Jaeger in 1954 using tape methods to perform an essentially similar function, did not indicate that great bandwidth savings could be effected before intelligibility was reduced.
    - 4. Voice Analog Vocoders. There is a large family of speech coding systems called vocoders. Unlike the ones described above, they separate the speech into its components at different frequencies, and then start cutting it down. Voice analog vocoders attempt to deduce from the signal what was making the sound, such as at what frequency the vocal cords of the speaker were vibrating, and where in his mouth his tongue was. At the receiving end circuits are controlled which imitate the physiology of a speaker and thus reproduce the sound. A recent review of vocoder activity listed no fewer than 35 independent research projects but gave no mention of any viable installations. Many of the studies demonstrated that bandwidth savings of the order of 1,000 per cent are realisable, but all the systems described used very large quantities of analog signal processing circuitry, and often a large general purpose digital computer as well. The machines were not easily reproduceable. There is. however a fifth method. This is the
    - 5. Frequency Domain Analyser/Synthesiser. We call it Vocom. Instead of trying to reproduce a speaker, like voice analog vocoders, Vocom instead recreates and reproduces a sound. By making an elaborate analysis of a vocal input (and this does require high speeds - up to 1,000,000 bits/second - but it is all inside the machine), it is p05sible to identify and then reproduce it using coded data of much reduced density. The output is not a reproduction of the original, in the sense of being some sort of recording or copy, but a synthesised recreation of it using extremely sophisticated sound generators. Research work has been speeded by the extensive use of L.S.I. technology (mentioned above - 3.), which enables very high speed, very compact circuitry to perform the analysis and subsequent synthesis of the input. Each parameter of the sound is quantised according to its perceived importance - the result is that the essence of the sound is retained while the pith is discarded - like crushing oranges' and concentrating the juice.

- How Does Vocom Work?
  - Well. obviously at this stage we are not ready to reveal all, but briefly it goes like this: A computer is able to programme 64 special devices which are both filters and oscillators, and are of extraordinary versatility. The 64 devices are not unlike the fibres in the basilar membrane of the inner ear - each fibre is attuned to a particular frequency, and gives a maximum response at that frequency, falling off gradually to zero at pitches above and below this resonance point. Our ears identify pitch by noting the varying amplitudes sent to the brain by these fibres. Vocom's filter/oscillators not only detect but also generate sound, and unlike the fibres of the cochlea they can be tuned to any frequency from 0-16KHz, so that particularly detailed study can be made of a chosen part of the spectrum. Each device (when oscillating) can have its level controlled to 8-bit accuracy (28=256 different amplitudea). The rate of frequency or amplitude change (defining e.g. the sharpness of a consonant) is separately determined for each filter/os~illator by 8 bits. There is a choice of three waveforms, and three digitally controlled output amplifiers supervise the overall dynamic levels (6 bits).
  - Apart from anything else, this machine can generate any sound, the problem being only how to define the sound for the machine. It is the most advanced device ever made for electronic music synthesis and acoustic research. But we are here concerned with Vocom as a speech processor. How does it do this?
  - When the filter/oscillator bank (in filter mode) receives an incoming signal, some filters do not respond at all, the rest in varying degrees according to how the energy in the signal is distributed. The computer's clock sends regular interrupts' to sample the entire bank of devices, and the resulting numbers are stored. After a few samples the computer begins to analyse the sound in the filter bank by examining the numbers. It has already been given a great deal of information about speech patterns, and as the amount of stored data increases with each sample the computer processes' the information, rejecting that which is unnecessary for the clear transmission of speech. For example if a vowel sound lasts for 70 milliseconds, it is to all intents and]urposes the same sound throughout its length (for our perceptions, 7Oms is not very long). It would be normal to renew instructions continually, but in this case it is not necessary. It is the difference between saying "Do it now. Have you done it? Yes. Do it again now. Have you done it? Yes. Do it again . . . etc", which is very tedious for a human being but quite usual for a computer though extravagant in bit usage, and the much more compact command: &Do it for 7Oms and stop.& Similarly certain phonemes, some consonants for instance, can be synthesised from prepared generators, and need not be elaborately recreated. Once the input is detected and identified, the message is '~Do Sound No. 143." But perhaps the most important breakthrough lies in the meaning of 'Frequency Domain.' Most digital conversion is done by sampling amplitudes in the way I described in the introduction. By repeated amplitude measurements and zero crossing detection the frequency is adduced. With Vocom, on the other hand, the frequency identifies itself because the relevant filters

are responding-and their frequency is known. So a complicated spectrum becomes - so much of Filter 1, so much of 2, and so on up to 64. This is enormously simpler than full scale mathematical assessment by amplitude sampling. And as mentioned above, almost any degree of fidelity is possible by concentrating the filters in the relevant spectral areas.

- After data compression, the filters (now oscillators) recreate the sound they originally 'heard'. The computer can also store a complete vocabulary of 'pre-packed' words', selected by program and sent to the synthesiser on demand - in fact it can do all the normal manipulations on digital material which one would expect from any processor. In its digital form the speech data has now been compressed to a rate of no more than about 200 bits/second. This amazing order of compression means that it can be transmitted over lines at whatever the bit-rate capability of that line may be. Sent by telephone line capable of 60,000 b/s, for example, a Vocom transfer could be made at a speed 300 times faster than real time. In terms of line rental savings, this is spectacularly good business.
- Vocom is extremely adaptable. In the above description I have covered only a fraction of the possibilities, and as I have said we do not at this stage wish to make all our secrets public, with an intensive research program in hand. But we hope these notes may have served to show that Vocom is really new, is really effective, and promises an extraordinary advance in the science of communications - and advances in this field mean more fruitful dialogue between people. In all areas where the spoken word is preferred to cold type, and there are many, Vocom will change the world. As an appendix to this brief exposition, you will find enclosed a few examples of the programs and print-outs used in our current recearch program.

VOCOM

- offers, for the first time ever, a practical method of storing human speech in a highly compressed digital form, and at a very low cost per stored bit. The coded voice data can use literally any existing communications system telephone, Telex, computer terminal service, wire, radio, satellite - well, we admit you can't send Vocom in a letter! But whereas long distance high quality broadcasting by line needs coaxial cables, waveguides, microwave links and other costly installations, Vocom is just as happy with ordinary telephone lines.
- Potential users of Vocom range from large businesses with their own computer terminals to any member of the public with a telephone. The Vocom terminal is simply a small box, not unlike a compact calculator in size and general appearance. This stands beside the telephone or Telex, and the renter would be supplied with a directory of code numbers. Some of these would be public services but there would also be confidential services for which special codes would be issued. To operate Vocom it is only necessary to key numbers on the terminal, and the reply is in ordinary speech, heard either through the telephone earpiece or a small loudspeaker. Though generated by Vocom, this voice is clearly recognised as that of the original speaker. The terminal box will certainly cost no more than $30, although we anticipate that it would usually be rented rather than purchased, as part of a Vocom subscriber service.

- VOCOM
  - is versatile. Not only can it store spoken messages of any length, but it can actually assemble sentences of acceptable grammatical structure from a stored bank of individual words. The modest computer (typically one of the DEC PDP8 series) which is all that is necessary to service Vocom can also perform other calculations and instruct Vocom to speak the result. For example, an inventory control system using Vocom might receive input data from the storekeeper as he despatched items, and would calculate what remained, continuously updating Vocom. The salesman on the road has only to phone in from a Vocom terminal to check the latest stock situation - but the message he actually hears may never have been spoken by anyone in that form. By arrangement with the telephone companies, there is in fact no reason why Vocom should not be dialled like an ordinary telephone number, particularly where security risks are not high.

- VOCOM
  - is aimed at people and businesses at every level, not just top professIonal users and the wealthy individual. Millions of the public who have previously considered any kind of data access quite beyond their reach will now, for the small rental of a Vocom terminal, have a mass of information instantly available, delivered in normal everyday language and adaptable to personal requirements.

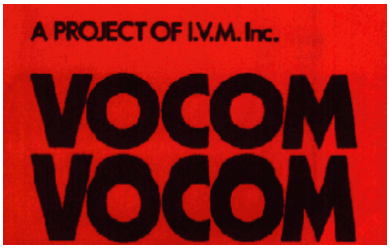
- VOCOM,
  - by selecting essential and rejecting unnecessary voice data, stores the highly complex patterns of speech in a more economical way than has ever been possible before. Far more flexible than an analogue tape recorder, a Vocom unit can accept voice inputs from any distance, process them in any way required, store them indefinitely, and finally transmit them in a fraction of the time it takes to speak them. The importance of this compression facility is incalculable. People will be able to talk for as long as they like into Vocom, and because the entire long message can be sent at many times its original speed the line usage costs will still show spectacular savings on conventional methods. A mass of non-urgent material, including whole books for publishers, could be packed into the least busy hours on the world's telephone and Telex lines. Both the command numbers and the message itself can be coded to ensure complete security, and it is possible to change codes automatically and frequently (every hour, for example) where maximum secrecy is essential. Though not normally necessary as well, a Vocom output can be scrambled like an ordinary speech signal. The addition of Vocom to an existing system does not affect its other functions in any way.

- For a private renter of Vocom, there will be numerous services at his command which at present are either not available at all or only in an inefficient form. Let's take a few examples-telephone answering, for instance. At the moment there is a choice between human answering services, which are often less than good, slow and very insecure, and telephone recording machines capable of not more than about 30 seconds per message, and also insecure. With Vocom, keying a personal code will automatically put all calls in to Vocom, which will answer callers in any way the subscriber wants, even answer questions if programmed to do so, and record a message of any length. The subscriber could access his messages from anywhere, update the answer to callers, anything.

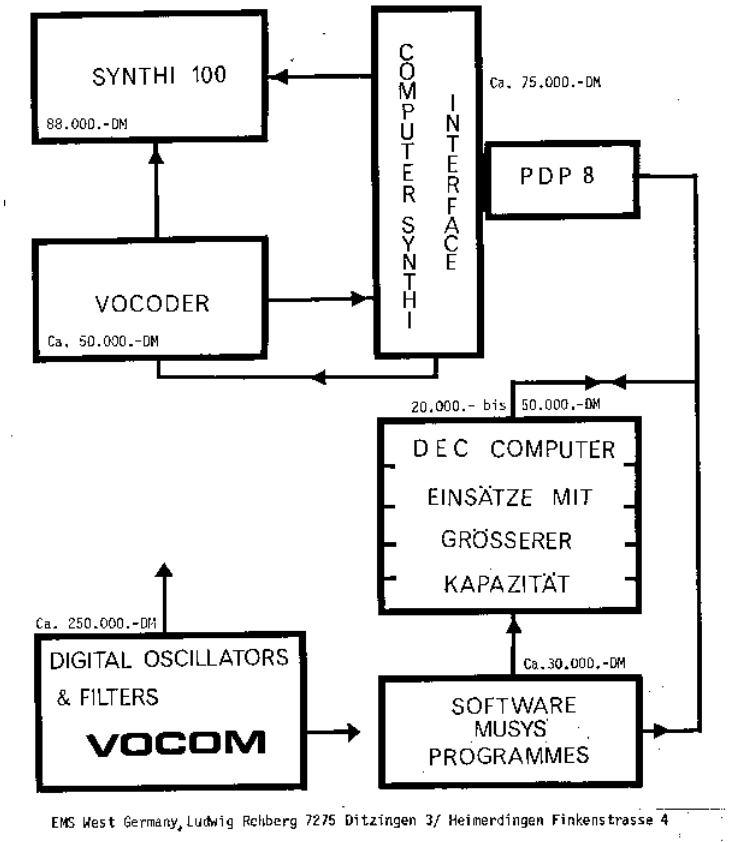
- Or take the stock market. The Vocom subscriber can have up-to-the-minute prices direct from the floor of the market. He can have his stockbroker's advice on trends and the latest changes. Suitable programming would enable the computer to accept portfolio information from the customer, line this up with current trends and compute the most profitable course.
- The possibilities of Vocom in the public field are endless. Airline, railroad or any other transportation reservation services would not suffer, as they do today, because one human being is not doing a job properly. The spoken message will actually come from a computer, and if the booking position is correctly stored the message must also be correct. Car hire, truck rental, personal credit situation instantly obtainable from the bank's computer on a confidential code. Phone directory enquiries, changes of address, hotel room booking, hospital bulletins. All this and more, continuously updated and continuously available in clear, understandable form. Since the initiation Is by key and not normally spoken (unless of course the user wishes to record something) it is possible for people not very familiar with the language stored to use Vocom where they might fail to communicate with a human being. If you have a little German, for example, you might not be able to ask for train schedules but you might know enough German to understand them. Simply key Vocom and listen. If you need it again, key it again Vocom does not get tired or impatient. In many cases, a variety of languages will in any cace be available on demand.
- For the business, civil and military user the expansion of capability is even more dramatic. Large concerns which already have computer terminals can put in a Vocom facility without any trouble at all, ei~her by adding a 'talking output' to the existing terminal or by installing the Vocom machine itself if the size of the business justifies the outlay. Vocom comes complete with its own computer or can be interfaced with an existing processor. Thus anyone with a computer can make it vocal and articulate.
- Let's again look at a few examples. A banker could verify the credit rating of an intending borrower, or the financial state of a company, confidentially and simply. International banking houses would have a low-cost world communications network of far greater scope than now. An entire meeting in a New York head office could be sent to Vocom as it happened, flashed across to Europe in low-rate time, and be heard (rather than read in hastily typed form) in the Zurich office an hour or two later.
- Again, let us suppose a large company has a sales meeting in Chicago, but one of their top executives is in San Francisco waiting to make a decision. He needs a directive from the meeting but much more than Yes or No. He can hear back the whole meeting from a San Francisco Vocom terminal, key another number to retrieve a six-month-old file stored in New York, go ahead and make his decision and then input Chicago's Vocom to dictate a long report giving his opinions and reasons. There have been no mistakes due to under-briefing or attempted economies on telephone charges. Everything is in Vocom as long as required. When the heat is off a typist can take it all down for the records, or a digital tape can be filed holding the entire operation in voice form exactly as it happened.
- For some users the word bank facility will be particularly useful. Imagine, for example, a permanent store of more than 10,000 words-a larger vocabulary than possessed (regrettably) by many of us. Apply this capability to a police problem in New York. As well as a large selection of ordinary words, all Manhattan street references would be stored in the bank, using clear, unhurried speech. In an emergency situation a person under stress can speak indistinctly or make mistakes. A message like "Blue Buick taken from Lexington at 45th, 11.52, believed heading Queens" could be keyed into Vocom using a rapid shorthand code. Vocom would then assemble the message from its word bank. Patrol cars would be in constant touch with Vocom by radio and receive the clearly spoken message, which would itself change minute by minute as it is updated. The original input keying would have simultaneously commanded another Vocom bank to assemble the same message in Spanish, and print it out in both languages for the permanent file. Criminal records, court cases, traffic control. There is no police operation which would not benefit from Vocom, and its installation will initiate a complete overhaul of police methods.
- As a last example (but there could be many, many more), consider the advantage to a military organisation in having the day's orders to distant units spoken in plain language, in as much detail as is needed and with no ambiguities due to telescoped or mal-transmitted messages. In high security situations like this, elaborate scrambling, codes and ciphers would of course be used but all decoding would be automatic and the destined unit commander would receive his orders directly, if need be from the Commander-in-Chief himself. Because of low line usage, massive amounts of secret material could be safely sent to foreign locations.

- VOCOM
  - is without any doubt the biggest thing in communications since the invention of the telephone. It can be added to existing systems with practically no disturbance and at very modest cost both to installer and user. Its impact on the community will be vast and wholly beneficial, and its potential for enhancing the quality of inter-national dialogue, leading to more efficient promotion of world trade and improved human relations, cannot be exaggerated.

- The Vocom
  - research programme is already well advanced, and the present need is for additional capitalisation to bring the benefits of Vocom into the world's communication networks with the minimum of delay. A stake in Vocom is truly a stake in a better world for tomorrow.



EMS - "ZUKUNFTSMUSIK"



Back To Top this Page

Back *Back to* Richard Davies NLnet Home Page