*This article examines the problems of analysing EVP ('electronic voice phenomenon') and other anomalous or paranormal sound recordings (particularly those containing apparent voices). Various analysis methods are described including listening (!), waveform and frequency analysis and phonetic software. The article also considers the vital importance of frequently neglected aspects of assessing paranormal sound recording such as audio context, automatic gain circuits, 'phoneme restoration' , 'verbal transformation' and software artifacts. It considers where the apparent voices recorded on vigils may come from. Is it really ghosts?*
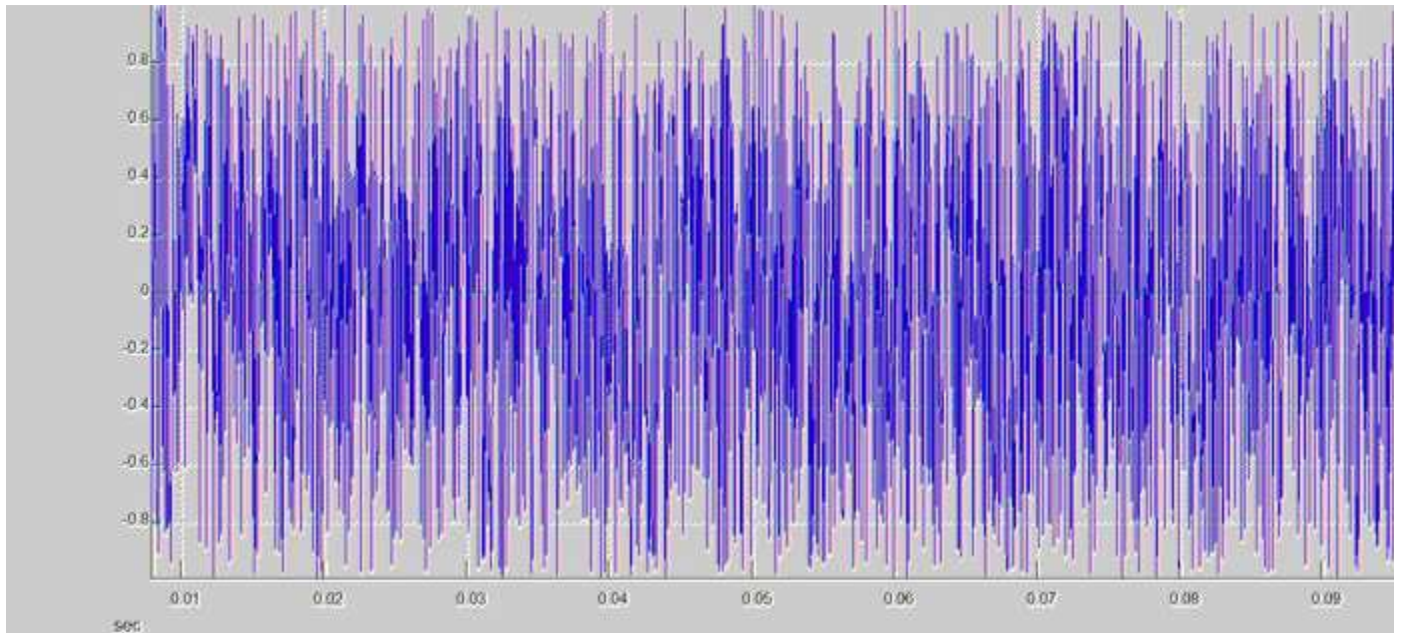
White noise

Note: there is an EVP gallery to go with this page where you can experiment with actual sounds.

Introduction

Please note that is this is **not** intended to be an EVP guide. Rather it is advice to those researchers who take audio (or video) recorders on investigations in the hope of capturing anomalous sounds, usually paranormal voices, who wish to avoid some of the pitfalls. This material may also be of interest to EVP researchers using 'open microphone' methods.

Tape recorders have been taken on vigils for decades and have frequently recorded odd sounds. Sometimes the intention was simply to replace note books by dictating records of observations. On occasion, there were specific known audio phenomena to be recorded. Most hauntings involve some sort of unusual sounds, sometimes including whispering or voices.

More recently, there has often been a deliberate intention, while on vigils, to catch ghostly voices. These voices are different from general sounds heard during hauntings because they are not heard at the time of the recording but only appear later during play back. It is, therefore, assumed by many people that the sounds must be paranormal in nature. Indeed, they are

generally thought to be Electronic Voice Phenomena (EVP), which have been researched (usually away from haunted locations) for decades. Some people think that EVP voices are caused by 'spirits'. Since many people assume that 'spirits' are the same as ghosts (even though investigations of hauntings generally fail to produce evidence supporting any such link), it probably explains why they seek EVP in haunted houses.

The recent trend towards seeking EVP-type recordings on vigils is different from traditional EVP research (also known as ITC or Instrumental Transcommunication). EVP researchers generally maintain that the voices can be recorded anywhere at any time. In addition, researchers sometimes deliberately include high levels of simulated background noise (like the 'white noise' hiss of a radio not tuned in to any particular station) in their recordings which they believe assists with the production of paranormal voices. On vigils, by contrast, people usually seek out quiet locations for recordings (though they may address questions out loud to any ghosts who might be present). Another key difference is that in traditional EVP research, strenuous efforts are usually made to exclude the possibility of any natural voices. In a vigil situation it is often difficult or impossible to avoid the voices of other investigators sometimes getting onto recordings.

Sound

A major problem with reviewing audio recordings for paranormal material (whether voices or anything else) is that, unless you are using a video camera, it is often difficult to recall all the possible natural sources of sounds present. Even if you are reviewing a video recording, any odd sounds that appear might come from somewhere out of the shot.

At least with a video camera present, it might be possible to eliminate some natural causes of strange sounds. There is no point locking off a room with just an audio recorder in it as any sounds recorded will inevitably remain 'unexplained' but by no means necessarily 'paranormal'.

We humans are used to seeing and hearing at the same time. If we are deprived of the sense of sight, it is very easy to misinterpret noises and not realise their source. In particular, if we are primed to hear something special (such as a ghostly voice) it is all too easy for expectation to be fulfilled even when there are, in fact, mundane causes (such as another investigator's voice) for it.

The Cocktail Party Effect

Imagine you are at a noisy party with loud music and everyone talking. You're not paying any attention to the hubbub which is too difficult to listen to anyway. Then you hear your name mentioned. You look up and see it is someone you know on the far side of the room talking about you. The fact that you can suddenly hear your name among the din shows that hearing is something that happens not in your ear but in your brain. Your brain filtered out your name from the noise because it was relevant to you. This is the Cocktail Party Effect. It is significant here because people quickly get used to background sounds (even in noisy factories) so that they don't hear them any more - because the sound is not relevant.

The same can happen on vigils. Just because we don't remember hearing a weird sound during a vigil, we shouldn't be surprised that an audio recorder picked it up. A sound recorder is a neutral instrument that records everything within its capabilities without discrimination. It never gets used to background noises, like squeaky chairs and creaking floorboards. In addition, the recorder may actually be more sensitive to quiet sounds than the investigator. Many microphones are strongly directional so that they pick up sounds from one particular direction preferentially (more strongly than humans) while ignoring others. So a sound recorder may pick stuff we didn't hear and could also miss things we heard. In neither case does it mean the sound was necessarily paranormal, it is just that recorders work differently to human hearing.

Another interesting point arises from the Cocktail Party Effect. Different people will get used to background noises at different rates. That means that some people may hear a sound whereas others don't (because only some are already not hearing such noises)! Needless to say, if this happens many people will claim that the sound must be paranormal as only some people heard it! This sort of thing is frequently reported on vigils!

Recordings aren't like the original sound

When we listen to a sound in a room, our brains use various different auditory cues to make sense of it. There is, for instance, the frequency of the sound, the time difference between its arrival at each ear and the sound's phase. In addition, there are the reverberations of the same sound (see 'Sound sources: Real voices' below) whose characteristics depend on the size and shape of the room. Our brains use these cues to decide which sounds, in a mixture, are most 'important', so that they occupy our attention.

When we listen to a recording, much of that original information is not present, even with stereo. Thus, our brains may place different priorities on which sounds we pay most attention to. The result is that ambient noise we didn't notice at the time of the recording may now seem much louder. This may account for apparent 'voices', not heard at the time of recording but seemingly obvious at the playback of the recording. Similarly, what appeared loud ('important') sounds at the time of the recording may now seem insignificant or even absent on the playback. When stressed, humans hearing sensitivity is increased, amplifying previously unheard natural background noise.

So, in summary it is perfectly possible for:

- Some people in an area, and not others, to hear a sound
- Sounds to be heard by investigators but not recorded
- Sounds to be recorded but not heard by investigators

Any of these situations can arise from completely mundane causes.

Don't edit your recordings!

Once you have a sound recording containing apparently anomalous sounds, make copies of it and keep it safe. It could be valuable evidence. Very importantly, don't 'enhance', edit, improve

or manipulate your recording in any way. Some people 'tidy up' or 'enhance' their recordings but this has several serious side effects: (a) it removes vital evidence of context, (b) it can make later analysis impossible, (c) it can 'create' false sounds (see *Audio Editing Software* below) and (d) it can look bad to people who wish to dismiss your possibly paranormal recordings out of hand.

If you wish cut out a section of your recording to show a suspected paranormal sound to others, leave a few seconds before and after the event as well. These are important to establish to context of the suspected paranormal bit. If you just play people the sound itself they have no idea of the context. For instance, is the anomalous sound one of many previous similar sounds? Is it loud or soft (many apparently paranormal sounds are very faint) compared to the ambient sounds? Is there any 'build up' to the sound or does it just suddenly appear? All these points could provide clues to the sounds origin that cannot be picked up by listening to the sound alone.

Audio editing software

Audio editing software is easily available nowadays, some of it very cheap indeed (like shareware or freeware). It is tempting to use such software to 'enhance' apparent voices or other paranormal sounds. However, there is a serious problem with this approach. While it is true that you can remove background noise from a recording, the process is not perfect and it inevitably alters the recording permanently, including the bits you are interested in. If there is a ordinary real human voice in a noisy recording, audio editing software may well clarify it (though it is also likely to alter it). If, on the other hand, there is a random noise that sounds like a voice, but isn't one, repeated use of noise removal, filtering, etc, could make the noise sound more like a voice. You could end up listening to an artifact of the overuse of audio enhancement rather than a real voice!

For example, suppose you hear something faint in a noisy background that could possibly be a voice. If you apply noise filtering and it sounds more like a voice, there is inevitably a temptation to do more filtering to see if it gets better. If this process does indeed make the sound more voice-like, you may carry on applying more filtering. However, if you continue to apply 'enhancements', the process may merely exaggerate the features that are voice-like so that you might be convinced that it IS a voice. Indeed, 'enhancement' can even introduce 'new' features to the sound that are just artifacts of the process rather than real recorded sounds. Such audio enhancement processes are not 'intelligent', they don't know when to stop but just apply the same rules repeatedly. On each iteration you get further and further away from the original sound until, ultimately it may be completely lost.

If you really must use audio editing software on your paranormal recordings, never apply more than one or two 'enhancements' per sample and apply exactly the **same** process to ALL your recordings. If you apply different types of 'enhancement', to different degrees, to each recording, then you will not be able to objectively compare samples. In some noise reduction processes you have to nominate an area of the recording as noise so that it can then be removed from the sample you are interested in. Obviously, this selection is subjective so, again, the results will vary from sample to sample.

Most audio editing software is designed to edit and enhance sound clips with a reasonable signal to noise ratio. It is not usually designed to recover faint signals from significant ambient noise, which is a typical scenario in paranormal sound recording. For this reason, the tools provided in non-specialist audio editing software, such as noise removal, may be too aggressive for use with paranormal sound recordings. If such software is used to remove an annoying hiss from an otherwise clear music recording, it should leave the music relatively untouched. However, when used to try to enhance faint paranormal sounds, it may change them so significantly that the results are of no use as evidence.

EVP can sometimes sound like speech speeded up or slowed down. For this reason, some people occasionally edit recordings to slow them down or speed them up. This is something else to be avoided as it destroys evidence.

This is why, ideally, you should aim to avoid all such audio enhancement, otherwise you could be left with software artifacts rather than paranormal evidence. You should, at the very least, always keep an unedited copy of your original recording for comparison. Try using phonetic analysis software (see below) instead.

Even if you don't edit your recordings, you could be effectively filtering them without realising it. This happens if you use a low specification recorder or use a low quality mode. See here for hints on selecting a recorder.

Manipulation

You may come across other people's recordings that have been manipulated by audio editing software. This may have been done with the best of intentions, to point out the 'important' bit but how are you, or anyone else, to know? Unfortunately, it is not always easy to tell if manipulation has occurred. There is less information in sound recordings than in photographs, making them easier to manipulate without people noticing. With computer software it is possible to make more or less seamless edits to sound recordings and make them difficult to detect.

One common manipulation is to remove 'background noise'. This is often noticeable when listening to the recording. The paranormal fragment will suddenly start, apparently preceded by total silence, and end just as abruptly without fading into any background electronic hiss or ambient noise. Such absolute silence is very rare on vigils.

If you have access to audio or signal analysis software you could also examine the waveform. It should never abruptly jump from one level to another in a vertical line (at least not at the highest time resolutions). If it does, it could indicate that something has been inserted or removed. Similarly, if the waveform goes completely flat (eg. at the start or end of the recording), or very nearly so, this could be caused by noise filtering.

If you can analyse the waveform frequency data (through FFTs, for instance) note if there are any frequencies apparently entirely missing. This could indicate that the sound has been filtered to remove particular frequency bands.
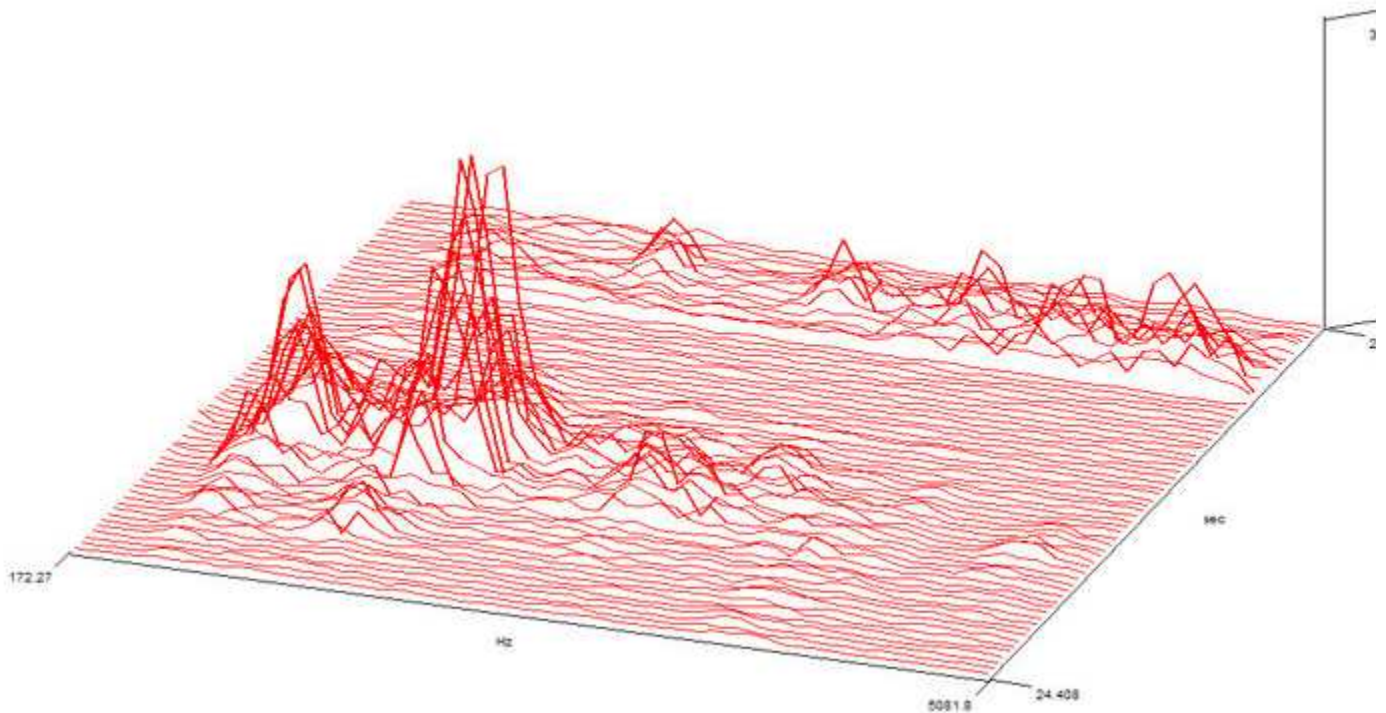
This picture shows the frequency (bottom axis) distribution for the word 'LET' (ordinary human voice, not paranormal) versus time (right axis). Compare it with the same recording below (in 'Analysis: Frequency' section). Note how all the lower frequencies, below 500 Hz, have been eliminated by a software filter in this version. It gives a distorted view of the real frequency distribution.

Consider the context

Background noise is important! It provides an audio context and reassures the listener that the recording has not been manipulated. It can also provide clues to any possible natural explanations for apparently paranormal sounds. For instance, is the paranormal sound louder than the ambient background noise, about the same level or fainter?

If the sound is louder than the background noise then there is the possibility that it is a real sound that was not noted at the time or was forgotten. It is also possible that the microphone was directional and happened to be aimed at a sound source that no one noticed at the time. It could also be radio or electrical interference.

If the apparently paranormal sound (APS) sound is at the same level as, or fainter than, the background ambient noise, then it could be a chance effect. Background noise is, typically, random and unpredictable. If two elements of it (such as a squeaky chair and a creaking floorboard or an electric fan and a noise from outside) happen to occur at the same time, they may combine to sound like something quite different and weird. By listening to the background over a long period, you may be able to deduce that this is precisely what has happened.

You might think it unlikely that you could pick up a sound fainter than the background noise. However, if it contains apparently meaningful information, like a voice, it could be the Cocktail Party Effect.

Sometimes you may record sounds from outside the building (assuming you're inside) where you are recording. These may appear to be coming from inside the building as there is no easy way of telling, from a simple recording, where a particular sound originated. You might even pick up the voice of someone passing the building outside.

You should record such details as where, when and how the recording was made and the equipment used. Recorders vary a lot in specification. This helps to establish the content.

Auto gain circuit

While listening to the recording, you should note if the background noise appears to fade drastically when a real, loud sound is picked up. This means that your recorder probably has an auto-gain circuit (AGC) whose job is to keep sound levels roughly constant. Most voice recorders have AGC and it usually can't be switched off.

The problem is that the AGC turns to high gain during quiet periods(amplifying background noise) and low during noisy ones (when the background apparently fades away). This can make it difficult to judge the relative loudness of different sound sources from the recording. It also tends to exaggerate background noise which can produce false APSs. In addition, the AGC affects the very part of the recorder (the microphone input and lead which can act as an antenna) that is most susceptible to radio interference.

In addition, the AGC can give normal recordings a 'weird' feel. The unnatural way in which the background sound can vanish and re-appear around normal sounds can feel quite spooky. Sadly, it's just normal operation for the AGC which is designed primarily for 'average' applications where there are not usually long silent periods.

Perhaps the only good thing about the AGC is that, if it reacts to a sound, it means the sound is real and fairly loud which is useful information to have.

Some people deliberately ask questions out loud to sound recorders to elicit EVP. They leave a silent gap for the answer before posing the next question. If the recorder has AGC, it will tend to amplify the bits between the questions so that noise will be more prominent during the 'answer' section.

Automatic gain circuits are common both in voice recorders and video cameras. If you can get hold of a recorder with manual controls, so that you can override AGC, it would certainly make life easier. Such equipment is, however, likely to be rare and expensive.

Compressed digital audio files

Many people use digital audio equipment to record paranormal sounds and EVP. Even if they use analogue audio equipment, they may well load sounds onto a computer to analyse and edit. This potentially introduces a problem. Many popular audio formats (like MP3, WMA) are compressed. In particular, such formats are 'lossy' which means that the sound stored is not identical to the one recorded.

There are various different methods for compressing audio data. Some are based on the way humans perceive sound so that, for instance, anything that we would not normally notice is simply eliminated before it is stored. In recordings of normal voice or music (for which such compressed formats are fine) you would probably never notice the difference. However, other sounds, such as background noises, often contain a wider spread of frequencies than in voice or music. Such sounds may be noticeably altered by compression techniques.

You should try to use uncompressed digital formats, like WAV, if you can, for recording and storing paranormal sounds. If you cannot, please bear in mind that what you have stored in your audio file is probably different from what was recorded. It could, potentially, make background noises sound more like voices.

Outdoors

There are special considerations when looking at recordings made out of doors. For one thing, the wind can affect the microphone, producing loud sounds that drown out others. You can buy microphone wind screens to stop these problems.

You also need to be aware that ambient background sound levels out of doors are likely to be higher than indoors. Also, the wind can affect sound so that it can carry further than usual as well as producing other odd effects. In particular, real voices (or fragments of conversations) might appear on recordings made outside.

Analysis : Listening

The most obvious way to analyse a sound recording is to listen to it, though it is also the most subjective. Indeed, scientists have discovered that more of the process of hearing happens in the brain than in the ears. It is important to listen to large sections of the recording at any one time so that you become accustomed to the context (see above), particularly the ambient background sound level and other natural sounds. Listening to small sections can leave you with a false impression of the occasion and any anomalies found. Were there, for instance, lots of sounds similar to the APS or was it unique?

It is difficult to judge what caused a particular noise (like the sound of a moving table) unless you go back to the site of the vigil and examine the scene carefully. You may be able to reproduce certain sounds by moving likely objects around or tapping them gently. This does not necessarily imply that the objects were moved paranormally. You really need to use a video recorder to obtain good evidence of phenomena like that. One of the most common things people listen for is apparent paranormal voices (or EVP).

People learn to recognise voices instinctively from childhood (it is done unconsciously like walking or swimming). Unlike most voice recognition software, humans can pick up words despite differing pronunciations and accents. Human language recognition accepts the possibility of 'false positives' (unlike software) as a price worth paying for not missing potentially important information. It does mean, however, that sometimes we can become convinced of a word or message even if we're wrong. If words are missed, we often fill in an 'appropriate' substitute based on the context. So, if someone said to you '*that ball shred*', you would almost certainly hear '*that ball is red*' because it makes sense from the context (unlike the real words). Your brain would substitute 'is' for 'sh' without you even being aware of it and you would be convinced that you heard it correctly. APSs are frequently not even as clear as '*that ball shred*'. Indeed, they can sound like gibberish to people not familiar with EVP.

You will often find that people disagree about what words are apparently said. Perhaps the best way to sort this out is to find a group of friends, preferably people not interested in EVP, and ask them what their opinion is and then take a vote. You could include a few 'control' voices in your sample too. These 'controls' would be ordinary, real voices saying known things. To make them sound more ambiguous (more like APSs), you could muffle the voice, maybe by putting a cloth over the microphone of your recorder. Make sure the clips you play include a few seconds of context (see above) on either side of the APS. Very importantly, do not tell your judges what words to expect as expectations strongly affect results in speech interpretation.

There are other tools for analysing apparent speech fragments.

Analysis: Waveforms

How do we know that a certain sound is human speech? It may seem obvious to most people but it is a difficult problem to scientists trying to make reliable voice recognition software. It is called the 'speech detection problem' and is still being actively researched. Once you have detected speech, there is another hard problem, to understand precisely what is being said (everyone's speech sounds slightly different). This is the 'speech recognition problem'. Both problems are central to the idea of apparently paranormal voices. Are the sounds recorded really voices and, if they are, what exactly are they saying?

In detecting human speech, an important concept is the idea of 'voiced' sounds. These are sounds (like vowel sounds) that involve the use of the larynx - they are called 'voiced' sounds. If you put your hand over your larynx (at the base of your throat) you can feel it vibrate when you make 'voiced' sounds. Some sounds, like 's' (as in 'less') or 't' (as in 'let') are not voiced - they require only lips and tongue to produce. Also, if you whisper, you do not voice sounds. This 'voicing' sound is called the fundamental frequency.

Voiced sounds are so common that exist in almost all words. So any fragment of speech of more than a syllable or two should include at least one voiced section. This is useful to know because voiced sounds have recognisable characteristics. The sound the larynx makes is low frequency and usually at a higher volume than unvoiced sounds. Look at the following examples of natural human speech. They come from signal analysis software. Shareware audio editing software often includes this sort of display.
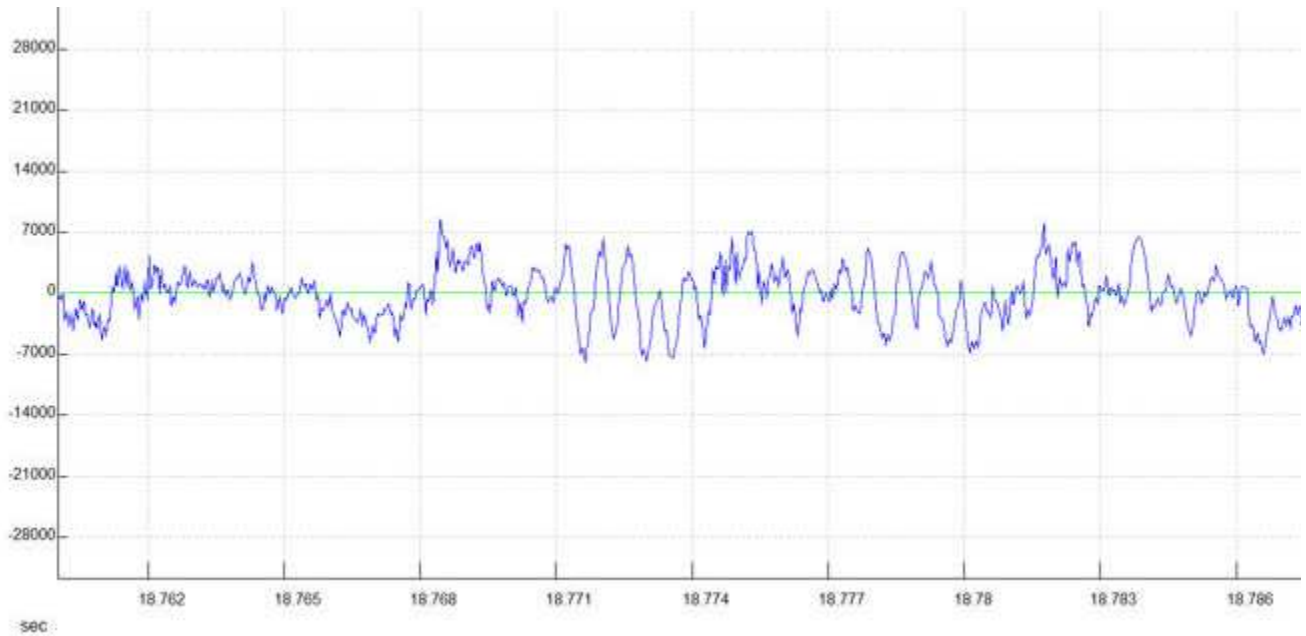
Above you can see the voiced vowel sound EH, as in the 'e' in the word 'let'. Below is the unvoiced T sound, also as in the word 'let'.
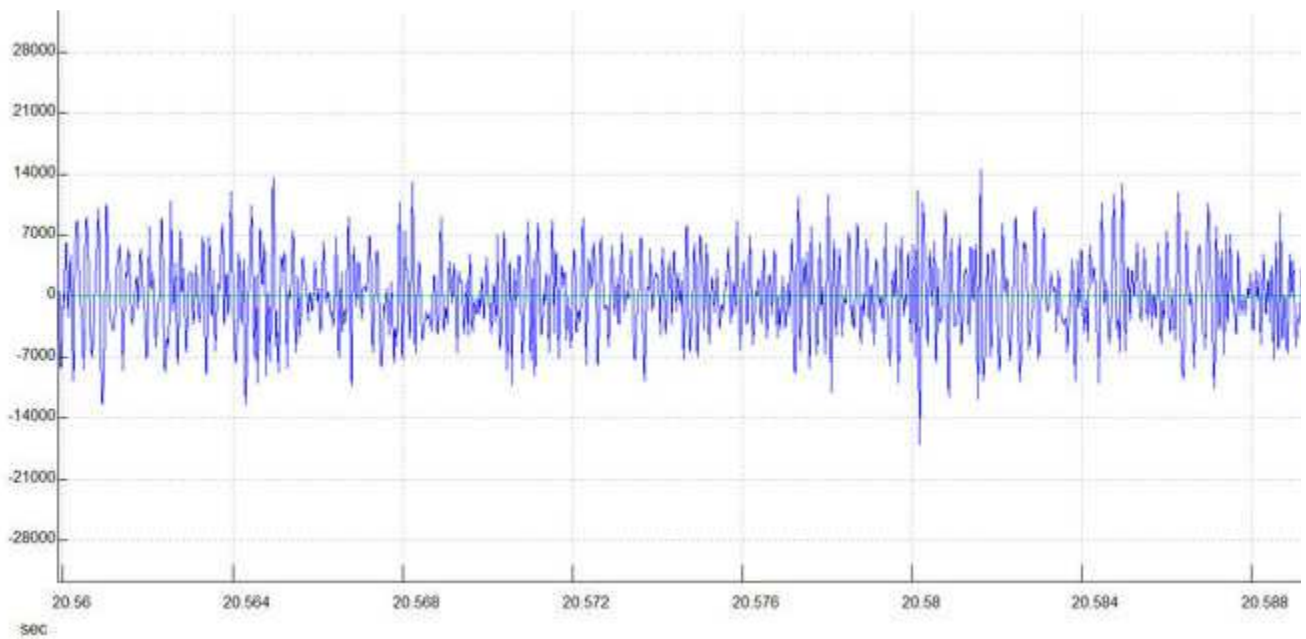
They are from the same recording of one voice. Notice that in the E (voiced) sound (top), low frequency dominates (the wave crests are further apart). This is the low frequency sound of the larynx vibrating.
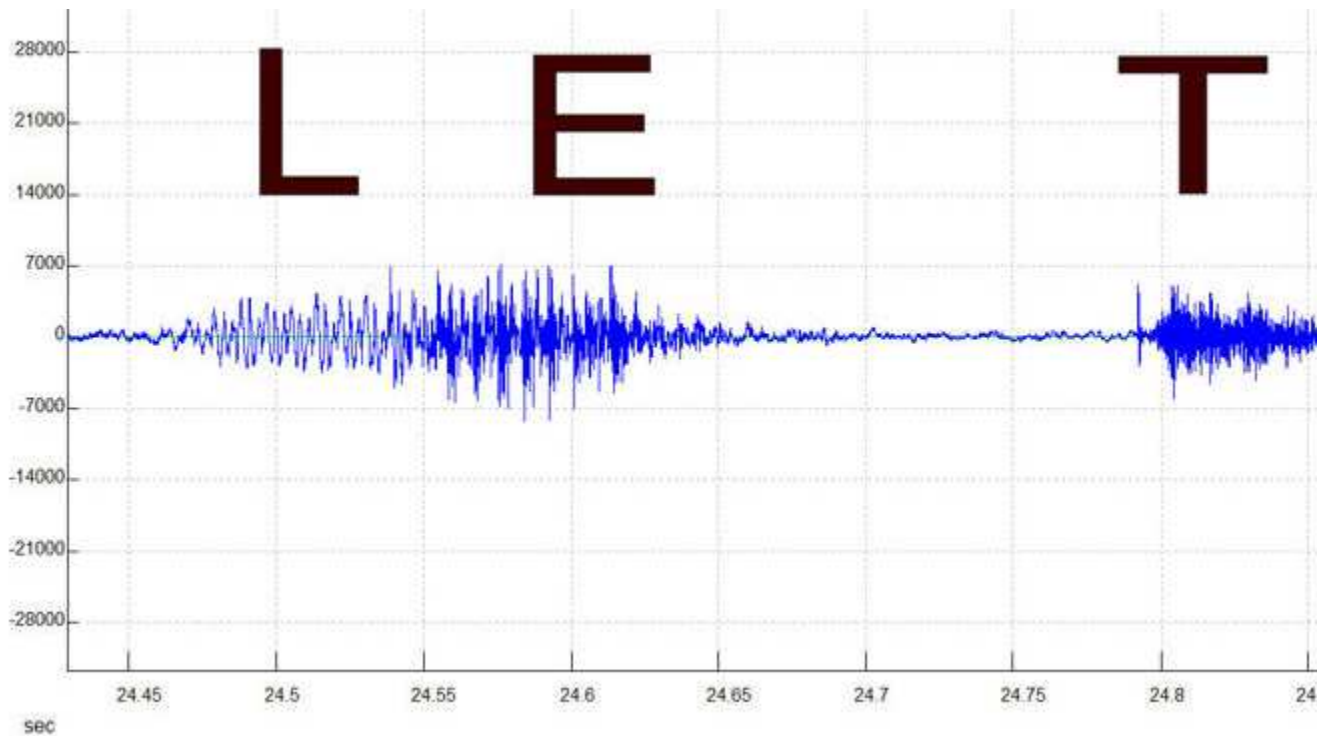
Above is the word 'let' (ordinary human speech). Notice the frequency increases from the L to the E (which are both voiced) but both are much lower frequency than the unvoiced T sound. There is a gap of over one tenth of a second between the E and the T, while the tongue and lips are re-arranged to sound the T! Note how the voiced section is generally louder (wave crests are higher) than the final unvoiced part.

These individual distinct sounds are called 'phonemes'. You can see them easily in the LET example above where the phonemes happen to correspond to individual letters, though they don't have to be. Some phonemes are composed of multiple letters (like 'ph' being pronounced as 'f'). Phonemes are generally reasonably distinct from each other when shown as waveforms. If you cannot see obvious transitions between phonemes, in what is supposed to be a word, then it probably isn't normal human speech. If, on the other hand, you can see typical sections of low frequency, high volume interspersed occasionally by high frequency, low volume bits, it probably IS normal human speech. You can try recording your own voice to see what various different phonemes look like.

If the phonemes look right for normal human speech, and it sounds reasonably intelligible, then it probably is a natural voice. It could just be a voice not noticed at the time of the recording. Even if there are no obvious phonemes then it could still be a normal voice but distorted for some reason. At least you can positively identify normal human speech, if it is present and undistorted, though.

Analysis: Frequency

In the next stage of voice analysis you will be able to resolve that vexed question - what exactly is being said? In simple speech recognition software, frequency analysis is used to determine what each of the phonemes actually is. Some audio editing software allows you to use fast fourier transforms (FFTs) to analyse the frequency spectrums of sounds. Below you can see the same word LET (normal human voice, not paranormal) as a frequency spectrum.
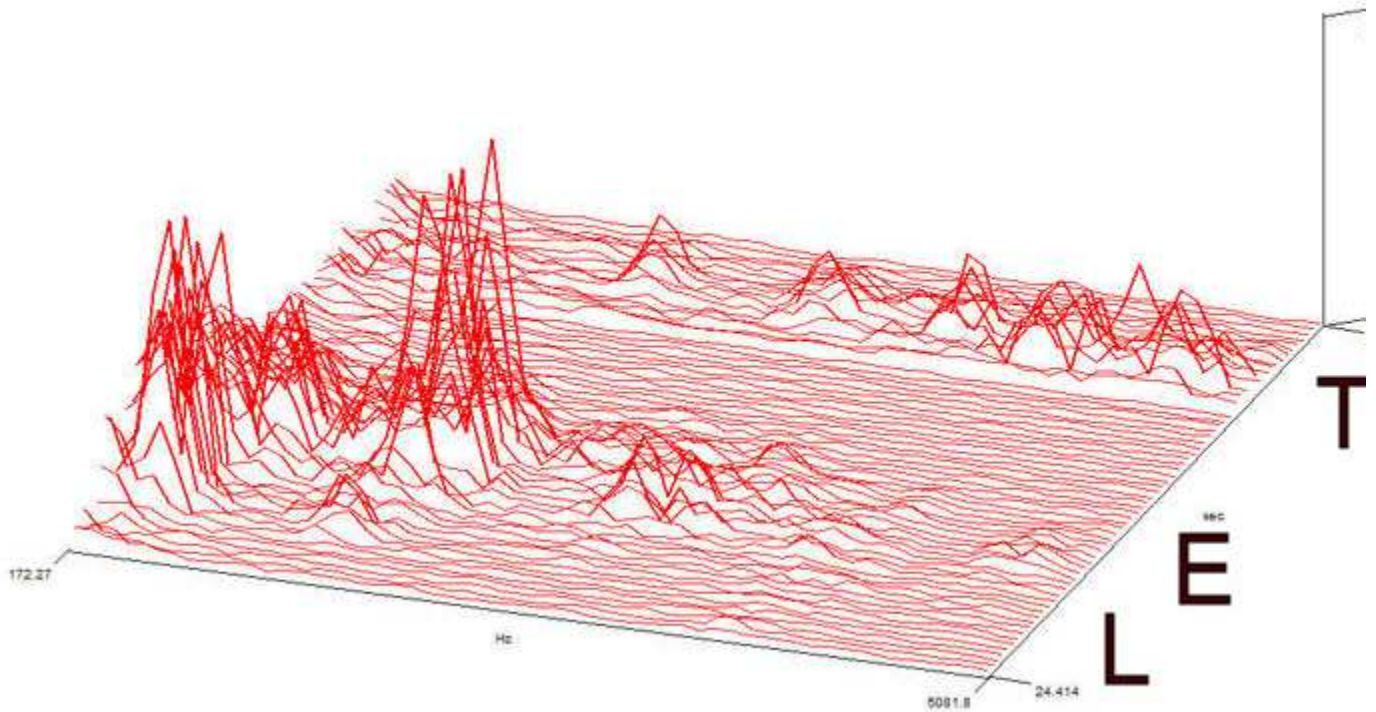
The figure above shows the frequency (bottom axis) versus time (right axis) with volume as a vertical axis in a three dimensional representation.
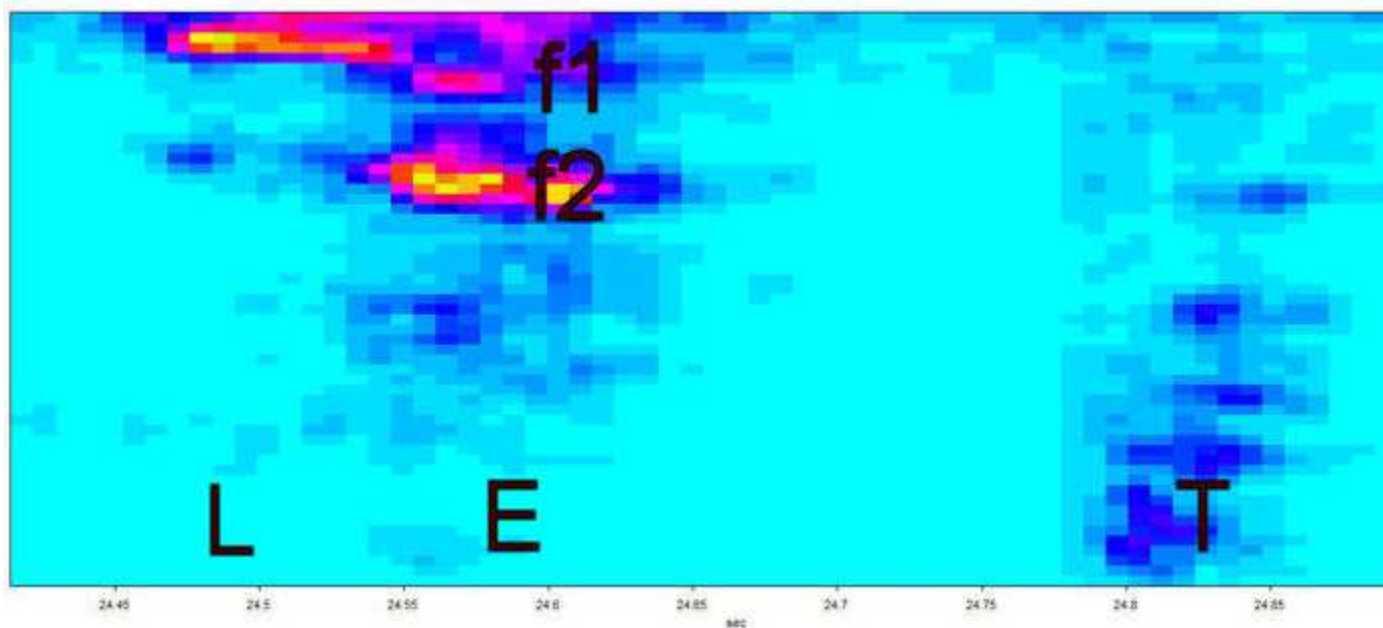
The next figure (above) shows the same information in two dimensions (a spectrogram) with time along the bottom axis, frequency along the sides and colours representing the volume (blue lower, through red to yellow higher) of the sound.

The frequency peaks you see in the figure above are called formants. The exact frequencies of such formants are determined by the precise shape of the vocal tract of the person talking. Essentially, there are certain shapes in the vocal tract where sounds resonate (this means that certain frequencies are effectively amplified). As a result formants generally occur at around 1000 Hz intervals. In the case of voiced vowels, the frequencies are harmonics (whole number multiples) of the fundamental frequency. However, not all harmonics are usually present. Only those that are resonated emerge. Also, the lower frequency formants (below around 2500 Hz) may be suppressed in consonants (see the T in figure above, for instance) by a phenomenon known as antiresonance due to oral constrictions. Overall, formant frequencies generally fall between 250 and 3800 Hz.

Phonemes are characterised by unique combinations of such formants. These formants are used by humans to recognise phonemes. (This is, of course, a huge simplification but sine wave speech, discussed below, demonstrates that formant frequencies alone are enough to understand speech). The two main formant frequencies in the E sound in the figure above are around 700 (f1 in figure) and 1800 Hz (f2 in figure). These frequencies vary significantly from person to person and time to time. Phonemes may contain up to 6 formants, though the main 2 are sometimes enough to uniquely identify them but 3 is more typical.

If you can see these formants in an APS, it should be possible to deduce the phonemes from them. Once you've done that you can assemble the word phonetically. If you say it out loud, you should be able to work out what it is. In the above example the answer should be 'L-EH-T' which is, of course, pronounced 'let'.

You should be aware that phonemes can be modified by sounds before and after (called coarticulation). For instance, where a 't' follows an 's', the two may run together as 's', as in 'get some' which may sound like 'guess-some' .

Analysis: Phonetic software

There is software available to help analyse speech phonetically. A good example is Praat, written by Paul Boersma and David Weenink. It is available to [download](). The software can measure many speech parameters used by phoneticists as well as displaying them graphically. You cannot use the software to 'validate' paranormal speech as real or otherwise, as it was never designed to do that, but it can highlight differences between ordinary real speech and random noise. You can then see where your own recordings of paranormal speech lie along that spectrum.
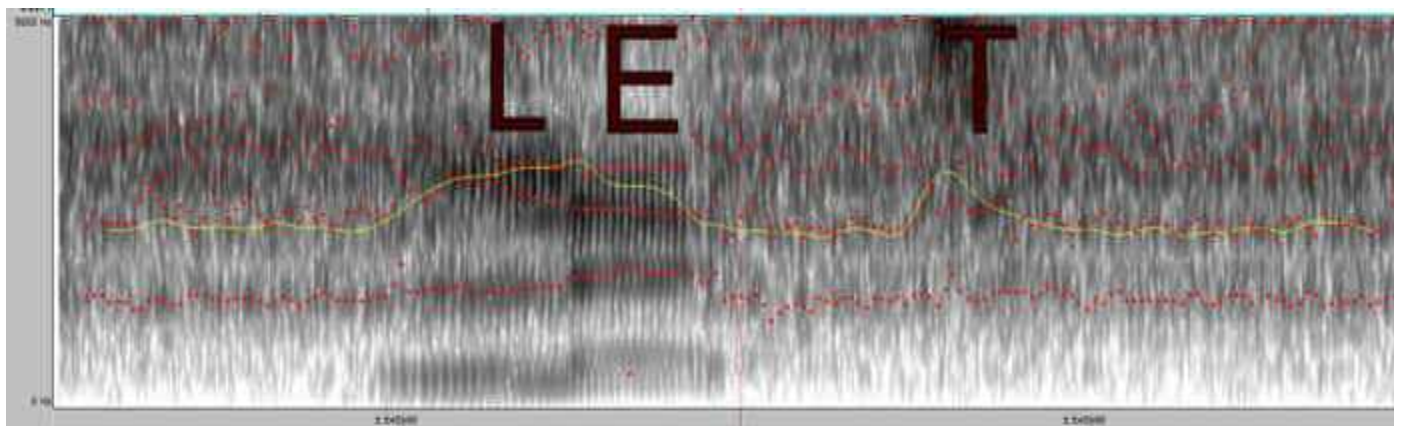
In the Praat diagram above, the word LET (ordinary human speech) is shown in a spectrogram (the letters L E T were inserted afterwards for illustration purposes and do not appear in the software) with frequency increasing upwards and time going right. The red dots are formant frequencies and the yellow line denotes sound intensity. The background sound is radio 'white noise' (see below). Note how the lines of red dots tend to wander up and down at random, outside the voice sounds, but are more closely spaced, and moving in a definite direction, within them. This is typical of ordinary human speech.

Analysis: Are they really words?

The techniques mentioned above should allow you to identify real speech. If your recording does indeed look like real speech, when analysed in software, it may be paranormal, if you can show there is no possible natural origin (such as a real voice that was not remembered or noticed at the time of the recording). But what about sounds that don't look like speech when analysed in software but still sound like words? How can you tell if there are real words present and not just an auditory illusion? To do this, you need to know more about how people recognise words.

Speech is understood by humans, completely unconsciously, using phonemes. People use these basic building blocks to understand speech. We can identify phonemes at a rate 20 per second but can only follow similar non-speech sounds at a rate of around 1 per second. Clearly the phonemes are built into our memory for quick recognition (as are whole words). Interestingly, it

is not simply a question of identifying each phoneme and constructing words. Phoneme sounds can be modified by the phoneme before and after them. They are also frequently slurred together. So the system of recognising speech (which also has to work for different accents) has to be very flexible. The cost of this flexibility is that it sometimes makes errors.

Phonemes are recognised by humans using the formants (see above). This has been demonstrated using the experimental tool of 'sine wave speech'. In sine wave speech, the formants in ordinary speech are identified by software and then synthesised into a new sound using pure sine waves. Thus, all the usual sound of speech has been removed apart from the basic formants. The result sounds, at first, like strange whistles but on further listening it becomes, surprisingly, perfectly intelligible (try samples here and read more here).

Considering that speech is such a complex sound, it is surprising that so little information is actually vital to understanding words. However, there is a second set of information present in sine wave speech that is not often mentioned. Sine wave speech also retains the original timing and intensity of the formants as well as their frequencies. This, too, is clearly important in speech perception. So, any random noise that contains frequencies typical of formants and varies in intensity with timings similar to real speech will inevitably sound voice-like.

Another vital point to realise is that word recognition takes place in tandem with phoneme recognition, though we are never directly aware of either process. By the time we are conscious of hearing a particular word, our brain will have already decided what it is and passed it on to our memory fully formed and sounding perfectly normal. The word will sound the same to us whether it has been correctly recognised by our brain or not! So even when our brains make a mistake, we remember erroneous words no differently from a correct one, convinced that we heard correctly.

Research into word recognition has shown that people:

- recognise frequently-heard words more quickly
- recognise words more quickly than non-words (eg. 'sing' rather than 'sning')
- context (ie. the words around the one being recognised) speeds up word recognition

Modern theories of speech recognition say that words are recognised by processing each phoneme in turn and eliminating possibilities as you go. Thus, the first phoneme may fit several hundred words. The list will be reduced a lot by the second phoneme, as fewer words start with that combination. This carries on until there are sufficient phonemes heard to distinguish the word unambiguously. In certain circumstances, previous phonemes may be re-examined to help speed the recognition. Context may also be used to speed recognition (ie. what word makes sense in that position in a phrase).

The exact way that the human brain recognises formants is still being researched. However, the fact that formants can be recognised, despite varying from person to person and from time to time, points to the idea that we use frequency ratios. The ratio of formant frequencies (F1, F2 etc) making up a particular phoneme should always be roughly the same in ordinary speech, even though the absolute values vary. For voiced vowels, for instance, the formant frequencies are

harmonics of the fundamental (voicing) frequency. When our brains recognise sound frequency peaks at known ratios, and in timings and intensities typical of speech, they will usually start to hear words, whether they are real or not.

Formant noise

So what if someone was played a random sound that contained various formant-like sounds which was **not**, however, real speech? If the person was convinced the sound was speech, their brain would try to identify formants in it by looking for familiar frequency ratios (particularly harmonics). All of this would be completely unconscious, of course.

Obviously, such 'formant noise', though voice-like, would sound like nonsense, at first. However, with repeated listening,certain chance formant combinations and sound intensity changes might suggest words (though different people might disagree on what the 'words' were). Indeed, there might be dozens of possible words suggested by one chance sequence of two apparent phonemes. The effect would be more convincing if there were noticeable changes in sound intensity, such as bursts of noise. These might, depending on their timing, suggest individual phonemes and words. Such changes in sound intensity could arise naturally, depending on the sound source, or from auto-gain circuits or audio editing, for instance.

So what words, out of the many possible, would the listeners, albeit unconsciously, choose? Frequently-heard words would certainly be one strong influence. Another would be expectation. When presented with such ambiguous and incomplete stimuli, people tend to be heavily influenced by expectation or suggestion. Thus, it would be relatively easy to turn random formant sounds (or chance sequences of them) into whole words and for those 'words' to make sense in terms of the circumstances of the recording. Once one 'word' was identified, the strong influence of context would come into play, so turning the rest of the sound fragment into an entire sensible-sounding 'phrase'. Thus, if an investigator asks a question out loud to the recorder, they would, almost inevitably, receive a relevant 'reply' (after a few listenings). This could all be based around a few phoneme-like sounds and chance sequences thereof. There are specific reasons why this can happen and they are related to the way people process speech. Surprisingly, rather than gobbledegook, we would expect reasonable words or phrases probably reflecting the context of the situation.

Thus, if an investigator on a vigil asks a question out loud to the sound recorder, they would, almost inevitably, receive a relevant 'reply' (after a few listenings). This could all be based around a few phoneme-like sounds and chance sequences thereof. There are specific reasons why this can happen and they are related to the way people process speech.

How we understand speech

Research has illuminated the, perhaps surprising, way we humans understand speech. Consider, for example, the phoneme restoration effect. In this effect, a phoneme in a word (or phrase) is replaced by noise but is still 'heard' by the listener (ie. the brain 'restores' it). The missing phoneme sounds as though it is there but it is an auditory hallucination. Indeed, the precise phoneme replaced can be altered by the context of the phrase. The phoneme restoration effect

works straight away in a single listening session. Interestingly, the effect only works if the phoneme is replaced by random noise - if there is a simple gap, no replacement phoneme is heard. This demonstrates how random (white) noise can substitute for real phonemes when the context is right.

On hearing the phoneme restoration effect: When you hear the phoneme restoration effect for the first time it is extraordinary (try the link above for samples). Even though you KNOW a phoneme has been replaced by white noise, you can still 'hear' the missing sound (usually a letter in the word) plainly! It graphically illustrates that, though it is your ear that hears sounds, it is your brain that turns them into meaningful words.

Now, imagine listening to someone talking in a language you are unfamiliar with (maybe Chinese for native English speakers). You can hear a person obviously speaking but all you hear is unfamiliar sounds rather than words you can understand. That is what your ear actually hears when you are listening to speech, even when it is in your native tongue. It is your brain that 'decodes' those vocalisations into meaningful words. The phoneme restoration effect shows what happens when you try to listen to speech in a noisy environment. Your brain just 'fills in the gaps' even when your ear hears only noise.

Another important thing to consider is the verbal transformation effect. If you hear a word (or phrase) repeatedly, it will eventually transform until it appears to be a different, though similar sounding, word. The transformed words have a similar phoneme structure (so 'truce' may transform to "truth" when heard repeatedly). Research has shown that if you hear a nonsense word (ie. a series of phonemes that are not a real word) repeatedly it is more likely to be transformed to something else than if you start with a real word. Indeed, the most frequently transformed nonsense words are those that comprise phoneme sequences that are never normally heard in a natural language. The word transformations can end up in either real words or nonsense words. Obviously, if you are expecting real words then it is more likely they will end up as real words. So, if you listen to 'formant noise' repeatedly, it is likely to be transformed into real words with the 'missing phonemes' being supplied by phoneme restoration.

Making formant noise

Based on the above, a perfect formula for 'formant noise' would be a sound of speech-like frequencies containing at least two frequency peaks that happen to be in a frequency ratio typical of speech (particularly a harmonic relationship). In addition, some time gaps between intensity peaks (to produce 'words') and a noisy background (to allow the phoneme restoration effect to kick in) would be very useful.

Something else that can enhance the voice-like quality of formant noise is to restrict its overall
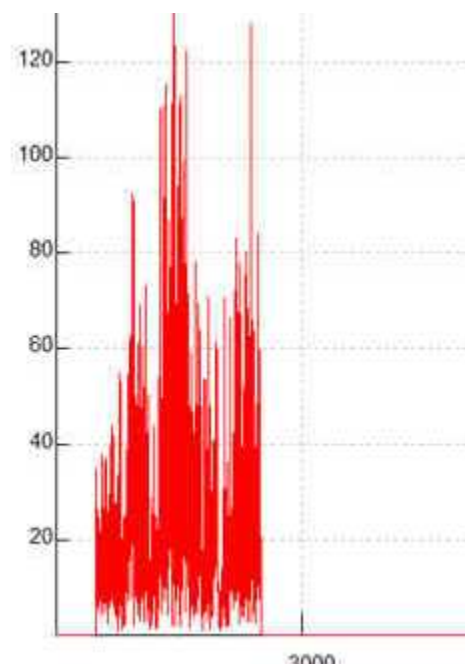
frequency range, producing a tight 'spectrum envelope'. A spectrum envelope is the overall range of frequencies and sound intensities – see figure right (the envelope is the shape enclosing the waves). Reducing the frequency range helps to enhance the illusion that the 'voice' is being produced by a single individual talking. Without this neat spectrum envelope, the noise may sound like lots of different voices contributing individual phrases or phonemes, which is less likely to fool the brain into thinking it is hearing speech.

You can reduce the frequency range of a recording by frequency (or FFT) filtering using audio editing software. Noise filtering (using the same software) may also accentuate any existing frequency peaks that happen to exist. This may increase the possibility that such peaks will be recognised by your brain as formants. This is why it is a bad idea to edit recordings to make them sound more 'voice-like'. You could be turning formant noise into a spurious voice.

Real speech not necessary

So, in summary, there is a possibility that sounds other than real speech could be interpreted as voices. This formant noise would contain:

- frequency peaks in ratios typical of formants (especially harmonics)
- changing sound intensity resembling the timing of phonemes and words

Some people will, on repeated listening, start to hear 'words' whose identity is likely to depend on context and expectation and will be rich in common words. Such sounds could arise from:

- natural non-voice sound sources - 'background' sounds
- real natural voices
- any sounds altered by recording compression and / or audio editing

Certainly, such formant noise has to be considered as a possibility when analysing EVP samples. So what are the likely sources of formant noise and other apparently paranormal sounds?

Sound sources: Background and forgotten noises

Probably the most important source of apparently paranormal sounds (recorded though not heard or noticed at the time) has to be ordinary background noise. This has already been discussed above in the 'Cocktail Party Effect'. As an experiment, try sitting in an ordinary quiet room with a sound recorder running and note down all the sounds you hear. Then play back the recording and see how many you missed and see if you can identify them!

Sound sources: Radio, electrical and magnetic interference

It has often been said that EVP might be caused by unintentional radio reception. It is hypothesised that people might pick up words from fragments of radio broadcasts picked up inadvertently by sound recorders. While this is possible and may happen from time to time, it seems unlikely to be a major source of apparently paranormal voices. If you try sampling many radio stations, by moving the tuning dial on your radio, it is easy to see why. Many stations

broadcast mostly music, rather than speech, these days. And if you did pick up a speech radio transmission on a recorder by mistake, the chances are very low that the words you would hear would be very relevant to your situation when you are recording. Most EVP researchers report that the words are frequently relevant to their situation and even sometimes appear to answer their questions! Wireless microphones are prone to picking up unwanted radio transmissions and should be avoided in paranormal research.

Electromagnetic interference from electrical equipment can be a source of anomalous sounds. Most interference is likely to be heard as a vague hiss, hum or whistle and a variety of other odd noises only heard on playback. If this happens to be of a similar frequency to formants, it could produce anomalous voices as described above. Electrical interference affects the electronic circuits of the recorder directly, producing apparent noise on the recording, despite there being no sound heard at the time. Modern electrical equipment is designed to minimise such interference but it could still occur.

Some designs of microphones (including popular ones) can pick up magnetic field disturbances. Thus, a magnetic field varying at similar frequencies to sound could produce a recording via a microphone, despite no sound being heard. Such magnetic field disturbances need to be quite strong, however, for this effect to occur. In a lot of cases it is likely that the source of such a magnetic field disturbance (eg. a motorised device) would probably produce real sounds as well, giving a clue to the source of the anomaly.

As an experiment, try using a cell phone near your sound recorder (within a metre or so). When you play back the recording you will probably hear the musical tones as the phone contacts the network. Then, when you make a call, you may well hear a buzzing sound from the call itself. You won't be able to hear the words being spoken (the signal is digital and encrypted and cannot be decoded by your recorder) but you will hear what at least one kind of electrical interference sounds like.

Sound sources: Real voices!

If you record real human voices in the distance, it may be difficult to work out what they are saying (even when heavily amplified). Naturally, such voices will contain formants typical of voices. Unable to understand what is actually being said, listeners' brains may re-interpret the formants to make new 'words'. Such words would tend to follow the pattern of interpretation mentioned above, ie. predominantly common words with the overall meaning biased strongly by expectancy. Effectively, distant voices could form a 'formant noise' capable of re-interpretation. In the figures below we see what happens when a voice becomes increasingly distant from a recorder.
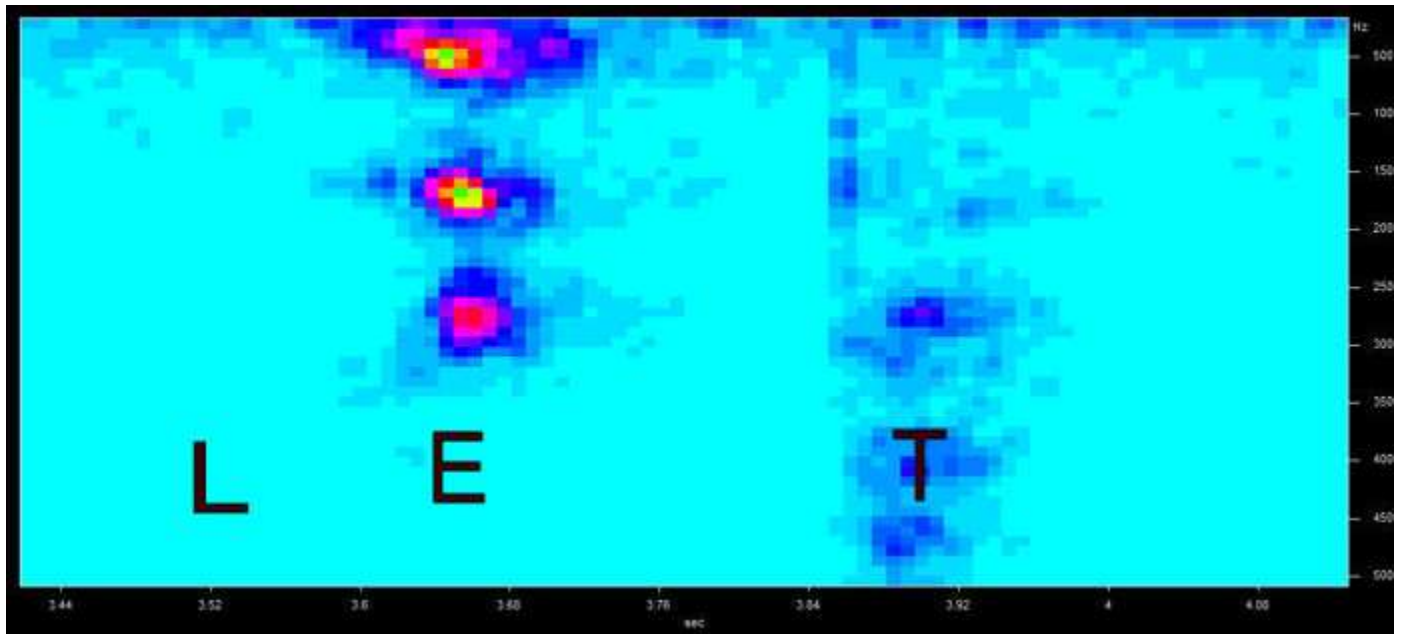
In the figure above, the word LET is spoken again (ordinary human speech). However, this time it is more distant (a couple of metres from the recorder). Compare this with the earlier spectrogram. Though the E sound is still distinct, the T sound has faded somewhat and the L has vanished! Notice how the low frequencies (towards the top of the spectrogram) from L and E (produced by the larynx) have gone.
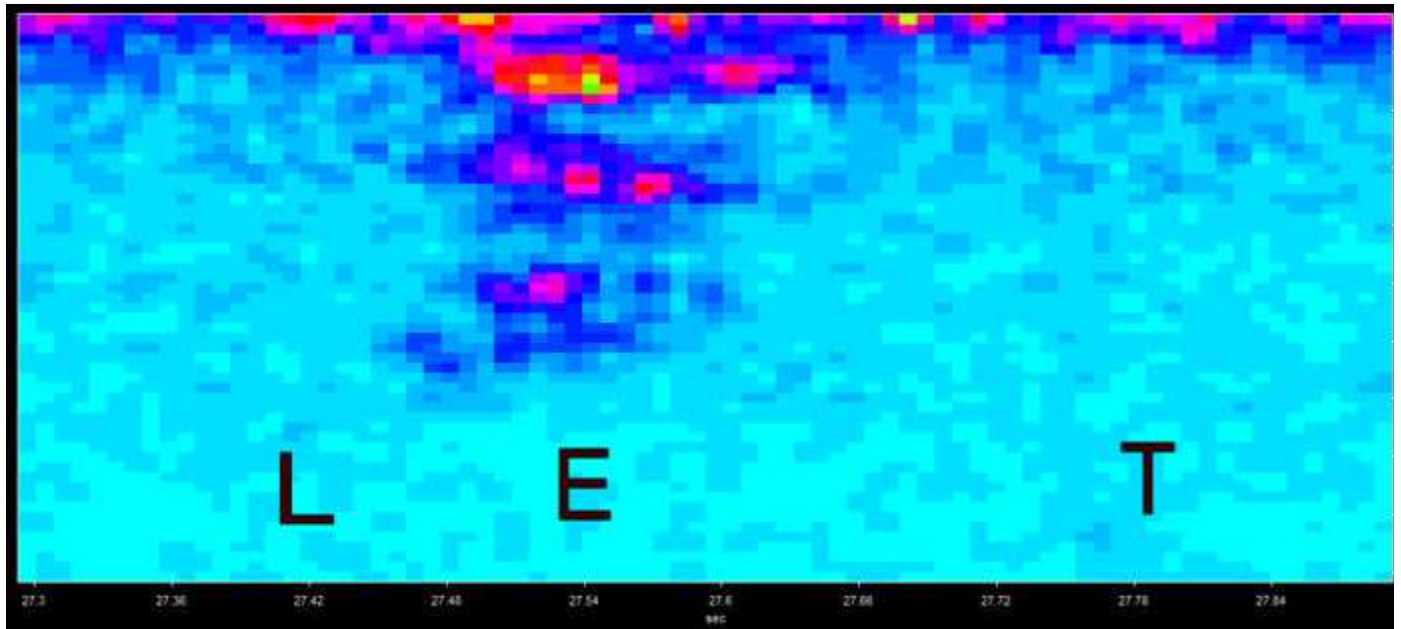

In the figure above the word LET was repeated at a distance of several metres and out of direct sight. Things have deteriorated further. The T has now vanished and only the E is still recognisable. Note also, the greater level of noise at all frequencies, particularly the low ones.

This is mainly because the LET sound is much fainter, bringing it closer to the general background noise level.

Research has shown that (as you would expect) speech intelligibility drops with distance from the sound source. In particular, as can be seen above, consonants suffer more than vowels. This loss of consonants has a profound effect on speech intelligibility because they are more important than vowels in understanding words accurately. Indeed, the loss of consonants is used in the ALCONS (Percentage Articulation Loss of Consonants) formula as a way of measuring speech intelligibility. Intelligibility is also affected by the level of background noise and, when indoors, the reverberation time of the room.

Reverberation time is the time taken for echoes of the original sound source to fade away. It is typically high in large rooms. A long reverberating time can affect speech intelligibility because it is strong enough, for long enough, to interfere with the original sound. Many vigils, of course, take place in old houses and buildings with large rooms. Speech affected by excessive reverberation could be difficult to understand and may be interpreted as paranormal.

As we have seen, any voice recorded from a distance would tend to be faint, low in consonants and may be affected by background noise. It would sound like a voice but it would be very difficult to understand the words being spoken. Looking at the last figure above as an example, there is a lot of general noise and only one formant (for the E sound in LET) which is not very distinct because of the intrusive noise. Anyone listening to this sound would be hard pressed to recognise it as E, never mind the word LET, and could easily re-interpret it as a different word. Such distant voices could easily be seen as a good source of 'formant noise'.
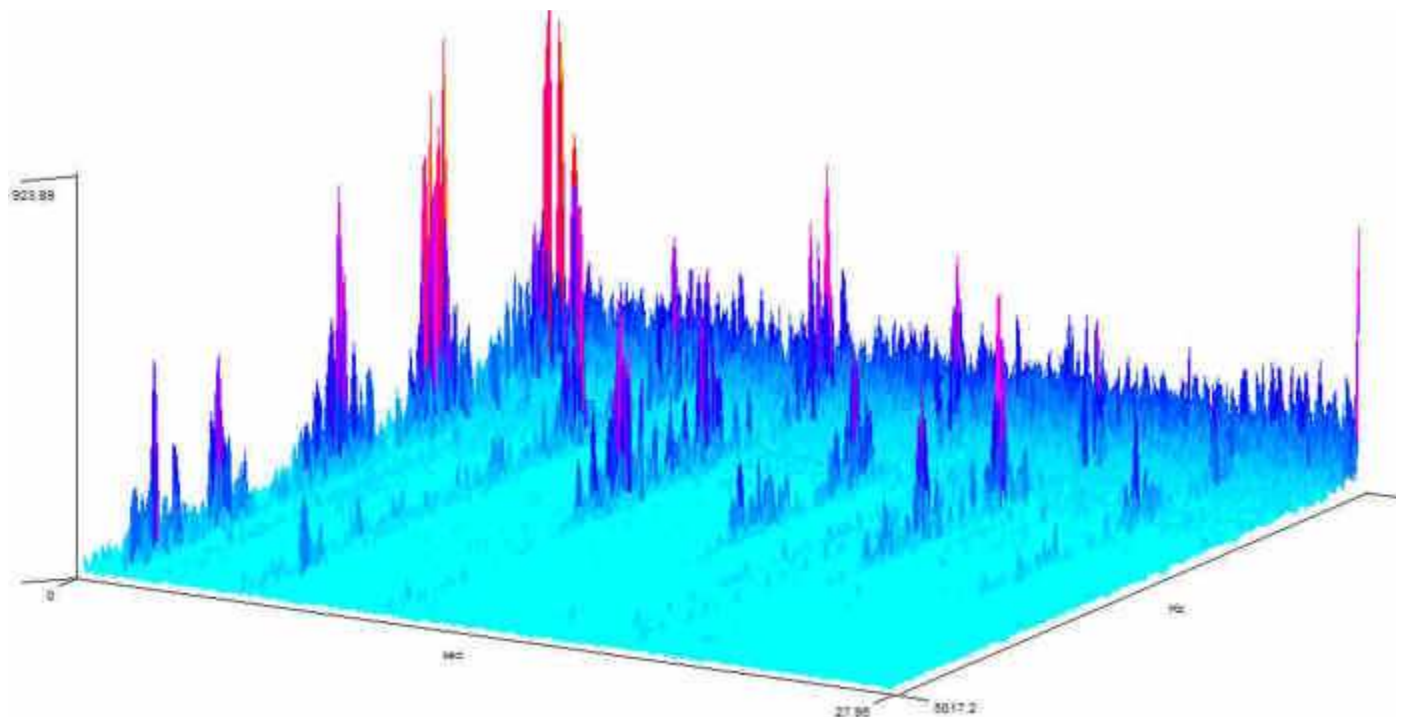
The figure above shows the frequency spectrum (a 3-D spectrogram) as the word LET is said out loud (ordinary human speech) at increasing distances from the recorder. Time is the bottom axis, increasing rightwards (0 to 28s). The right axis is frequency, increasing towards the front (from 45 to 5000 Hz). The vertical axis is loudness. You can see that the word LET is said six times in all (the parallel lines of peaks coming towards the front like chains of mountains poking out of light blue clouds). When the word LET is first said (far left), there are obvious high peaks - the formants - up to 5000 Hz (good enough for voiceprints!). As the speaker moves further away (going right), the number of formants gradually diminishes. They also shrink in height until they are no higher than the low frequency background noise forming a 'wall' at the 'back' of the graph. The higher frequency formants (above about 3000 Hz) vanish completely. In addition, the lowest frequency formants vanish into the background noise. Speech recorded at a distance will have few formants, mostly in the region below 3000 Hz.

Sound sources: White noise

Other sounds, apart from speech, contain formant frequencies. EVP researchers use 'white noise' to stimulate apparent voices. Pure white noise would seem to be an unlikely source of 'formant noise' because it contains all frequencies equally with none standing out. However, there are many white noise-like sound sources which are often reported to produce EVP. These may include sufficient random irregularities to give rise to suitable formant frequencies from time to time. The white noise surrounding these random irregularities is useful for filling in 'gaps' in words using the phoneme restoration effect.

There are various sound sources that sound superficially like white noise that you might record on vigils. These include the wind, electrical equipment (particularly anything with a fan), flowing water, radio noise, etc. The following spectrograms all show 2 seconds of various 'white noise' sound sources between a few Hz and 5000 Hz.
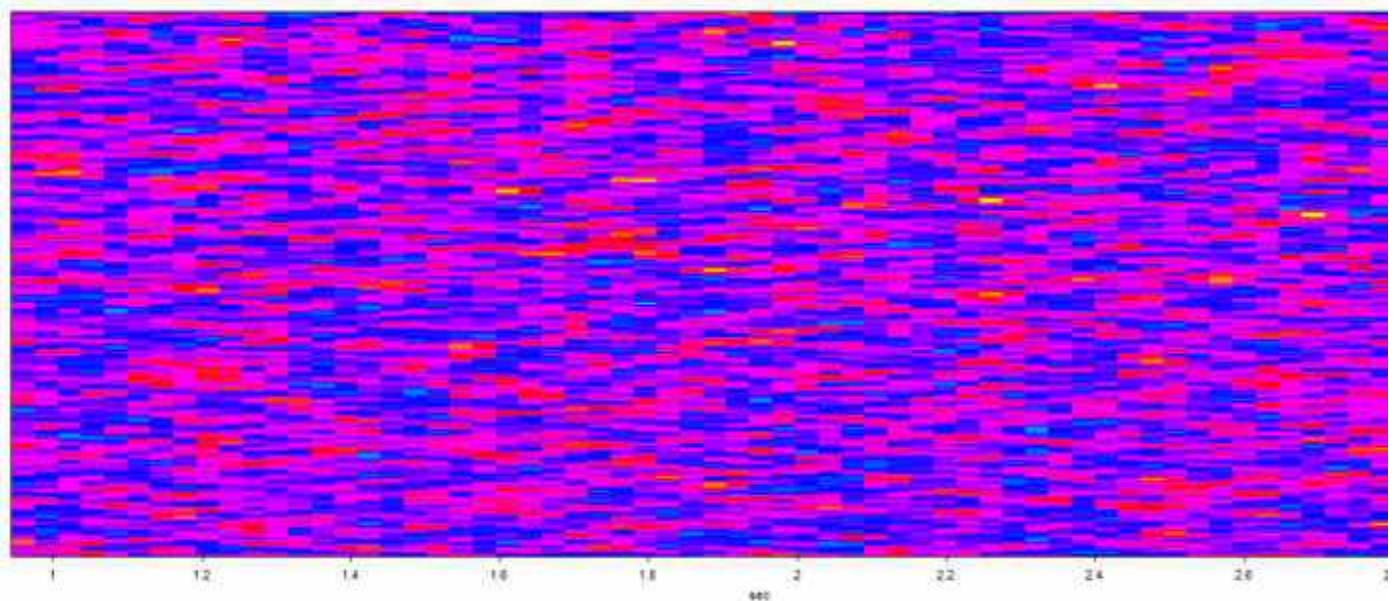
The figure above is a spectrogram of pure electronically generated white noise. All frequencies are represented equally so that there are few random irregularities likely to stand out as formant frequencies.
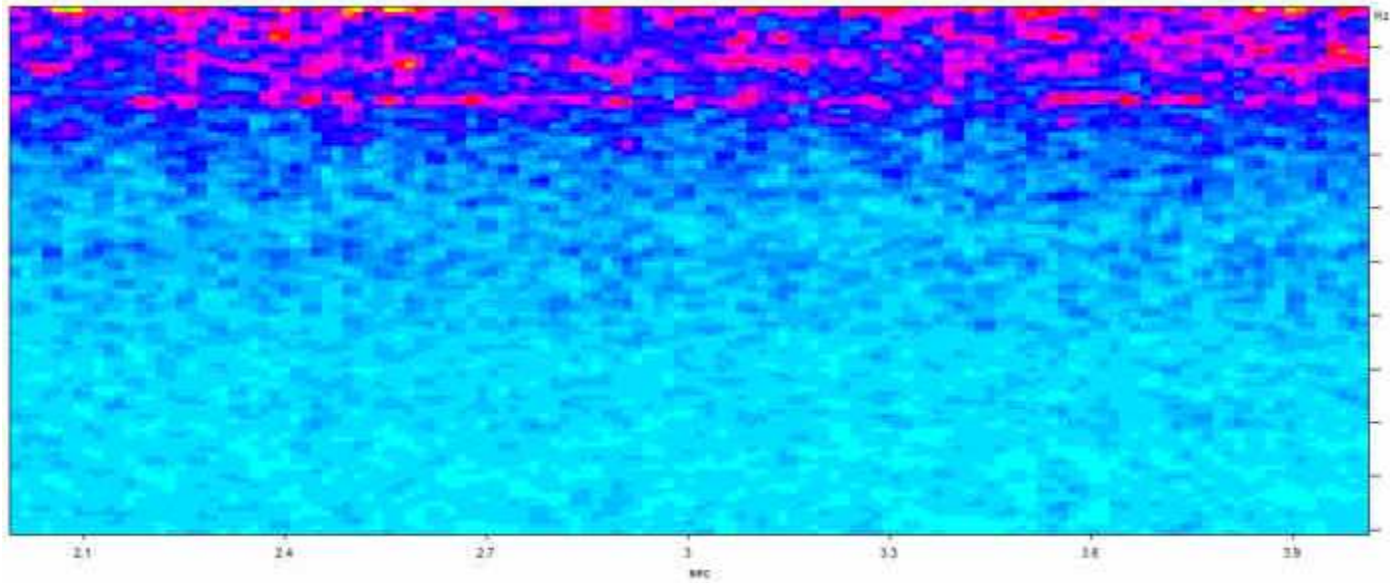

In the figure above is the noise from an electric fan. Most of the noise is concentrated in the lower frequencies, below 1500 Hz. Notice that the sound is by no means constant. There are random irregularities, some at formant frequencies. Some people report hearing voices in the sound of fans.
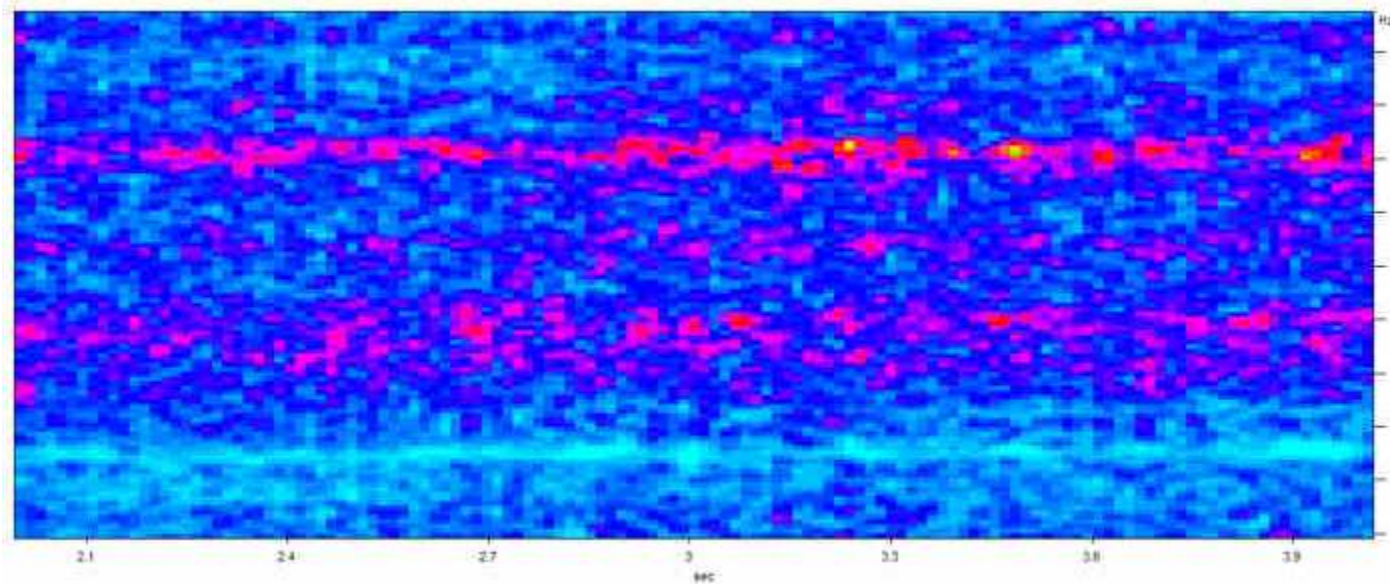
The figure above shows radio 'white noise'. It was recorded by tuning a radio between stations, a method used by some EVP researchers (no radio voices can be heard on this recording). It is certainly not white noise in the strict sense (compare it with true white noise above). There are

notable peaks at the 1500 Hz and 3000 Hz frequencies as well as a more irregular string of peaks around 2200 Hz. There are many random irregularities that could give rise to brief formant sounds.

Finally, here is an interesting recording of real, normal words against an electronically generated white noise background (over a longer time-scale of 6.5s compared to previous figures in this section).
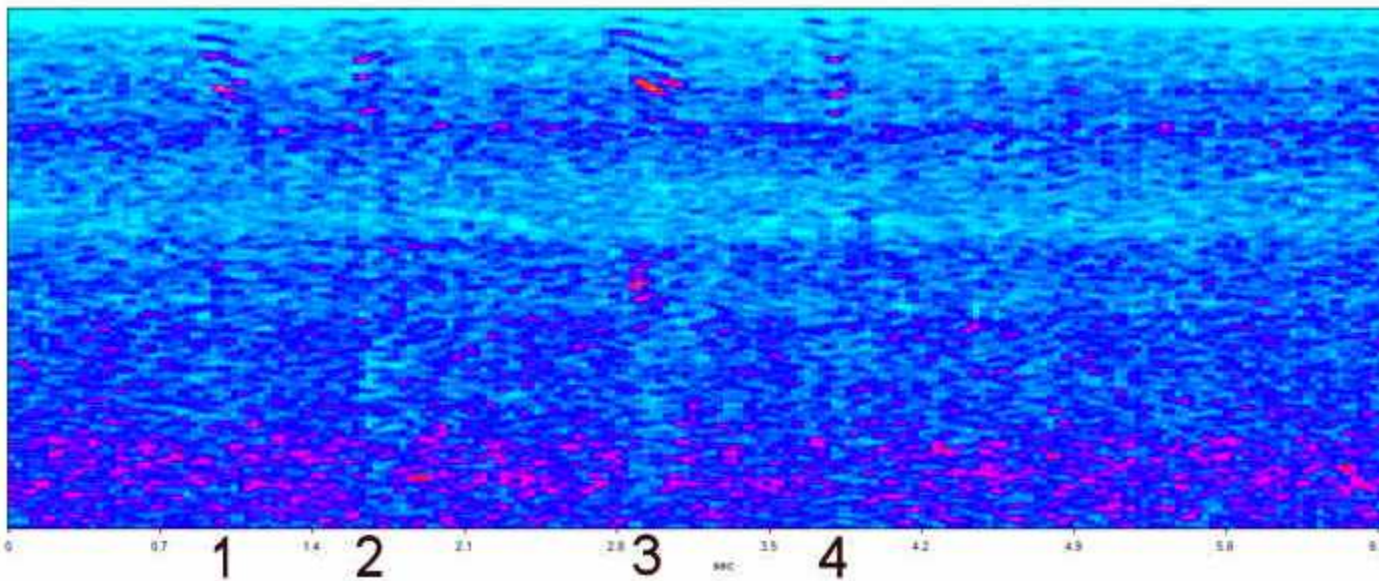
The words are indicated by the numbers. They look like vertical stripes, typical of real, normal speech. To produce this recording, pure white noise was played through loudspeakers and words were then spoken over it into a digital recorder (using a lossy compressed file format). As you can see, the background white noise sound frequencies are not evenly spread out. Note the lack of very low frequency sound and the excess above 4000 Hz. Compare it with the top figure in this section which shows pure electronically produced white noise with the same frequency range. Various factors may have introduced the uneven frequency distribution. These include the frequency response of the equipment used to reproduce the white noise and re-record it, as well as the compression used by the recorder. So even using pure white noise can be distorted by electronic equipment.

Testing voices

Bearing in mind the foregoing discussion, it is clearly difficult to decide what is a paranormal voice and what merely an auditory illusion caused by the way the brain's speech processing works. So, what can be done in practical terms?

- One thing to try is to break down phrases of apparent voices into their constituent words and play them individually. Do they still stand up as words on their own?
- Another thing you can try is to listen to the individual words (on their own) and allow yourself the 'luxury' of repeating exactly what you hear, even if it sounds like a nonsense

word. Does a nonsense word (like 'srin' instead of 'skin') actually fit better than your original choice?

- Looking at the frequency analysis, are all the expected phonemes present or are one, or more, 'replaced' by random noise (phoneme restoration)?
- Reproduce your interpretation of the words as closely as you can and say them into your recorder and then compare the frequency analysis of your real words with the sample you're examining. Are all the expected formants present and are their frequencies distinct or do they run into each other?
- Try playing your voices to someone with no interest in the paranormal without telling them your interpretation (see below).
- Try selective frequency filtering (see below)

There are many other tests you could apply. It is important keep an open mind and follow the evidence rather than personal theories you may have.

Assessing content

Assuming that you are satisfied that you have recorded a voice, there is still the question of actual word content to be considered. Does the apparent vocalisation make any sense as real words? Do the words make any sense as a message?

EVP researchers often say that their messages make sense within the context of their recording (though they can appear meaningless or cryptic in isolation). Of course, expectation and the verbal transformation effect are important in content. If you listen to the same words repeatedly (whether a normal real voice or EVP sample) you will hear the words shift around similar sounding alternatives (sometimes including nonsense words). If, as often happens in EVP, you are told what words you are going to hear in advance, it is no surprise when you hear the expected message - a clear case of suggestion.

While assessing content is always going to be a subjective process, there are things that can be done to overcome the more obvious biases.

- Try selecting third-party judges who have had nothing to do with the experiment or investigation. Ideally, get people who have no experience of, or interest in, EVP or even the paranormal.
- To avoid the effects of suggestion, give your judges no information in advance about your interpretation of the apparent voices or the circumstances of the recording.
- Prepare a single recording for them to judge, containing a number of EVP samples, each interspersed with other recordings of normal, real voices and non-voice sounds. The normal voices should be third parties unknown to the judges. Put silent gaps between each sound and introduce each one with a number to identify it.
- Ask your judges to say, in each case, what they think the sounds are and, if they are voices, what they are saying.

All the sounds should be of similar duration to the EVP samples. Some of the normal, real voice messages should be sensible words or phrases while others should be nonsense words or phrases.

The reason behind this approach is to reduce psychological bias. By separating your EVP samples and not repeating them in isolation, you reduce the verbal transformation effect. By including random sounds, you give your judges psychological 'permission' to say your samples might be random noises, if they think they are. By including sensible and nonsense voice messages, you give them 'permission' to say that, though your EVP sample sounds may resemble a voice, the message makes no sense, if that's what they honestly think.

Natural causes to look for

By way of summary, here is a brief account of the sort of natural causes for anomalous sound recordings that you should be looking out for.

- faint sounds that were either not noticed at the time or simply forgotten - don't forget that the AGC will amplify very faint sounds during quiet periods and a directional microphone may pick up noises from small areas preferentially that you may not notice - try analysing sound waveforms to look for voiced phonemes
- background sounds - it is easy to get used to repeated or continuous background sounds (such as electrical equipment, such as fans and pumps, or the wind) that can be quite loud - they may come as a surprise when you play the recording back as sound recorders don't get used to sounds (though the AGC may amplify them) - again, these could include formant frequencies
- electrical or radio interference - this would only be noticed when the recording is played back - it doesn't have to be obvious radio transmissions but could just be hums, buzzes, whistles, etc. - the AGC may tend to amplify these as well
- multiple sources of sound may come together, by chance, to form apparently novel sounds, particularly in association with 'background' sounds
- real voices so distant and faint that the words cannot be made out

If any of these sound sources produce a noise containing frequencies typical of formants ('formant noise') they are likely to sound **like** a human voice but the 'words', if any, may be difficult to understand on first hearing (this is because, if they are random noise, there are no words, as such). Repeated listening may produce reasonably robust 'words' (though different people will probably disagree about what they are). The content of these 'words' is likely to be dominated by common words and the overall 'message' (if any) will be strongly affected by expectancy (ie. it will probably appear relevant to the situation).

Of course, once you have eliminated these natural causes, there is always a chance that you have a genuinely paranormal sound recording.

Selective frequency filtering

How can you tell if a voice is genuine or just an illusion caused by 'formant noise'? In the case of formant noise, the 'formants' being picked up are just chance (rough) speech-like ratios between two or more frequency peaks. Unlike in real, natural speech, the number of formants producing each perceived phoneme is rarely likely to exceed two! This gives a possible way to test for illusory voices.

Assuming you have the software available, you need to selectively filter out bands of frequencies (sometimes called FFT filtering) from your sample. For instance, you could eliminate the following frequency bands in turn: 0-400 Hz, then 400-800 Hz, 800-1200 Hz and so on. In each case you leave all the remaining frequencies intact. You can then repeat with overlapping bands eg. 200-600, 600-1000, 1000-1400 Hz and so on. Ordinary, real speech will be altered by this treatment but the words will generally still be understandable. Illusory voices, by contrast, are likely at some stage to vanish or change radically (maybe different voice or different words). This is because you will break up the chance frequency ratio (probably formed from just two frequencies). Repeat the same experiment on a recording of yourself saying the same message and compare it with the sample you are analysing.

There is also a better, quicker way of doing the same thing. Instead of selecting pre-defined equal bands of frequencies to eliminate, you could examine the spectrogram (see examples above) to choose them. If there are two or more obvious bands of more intense sound at particular frequencies in your recording, try filtering out each in turn. For instance:
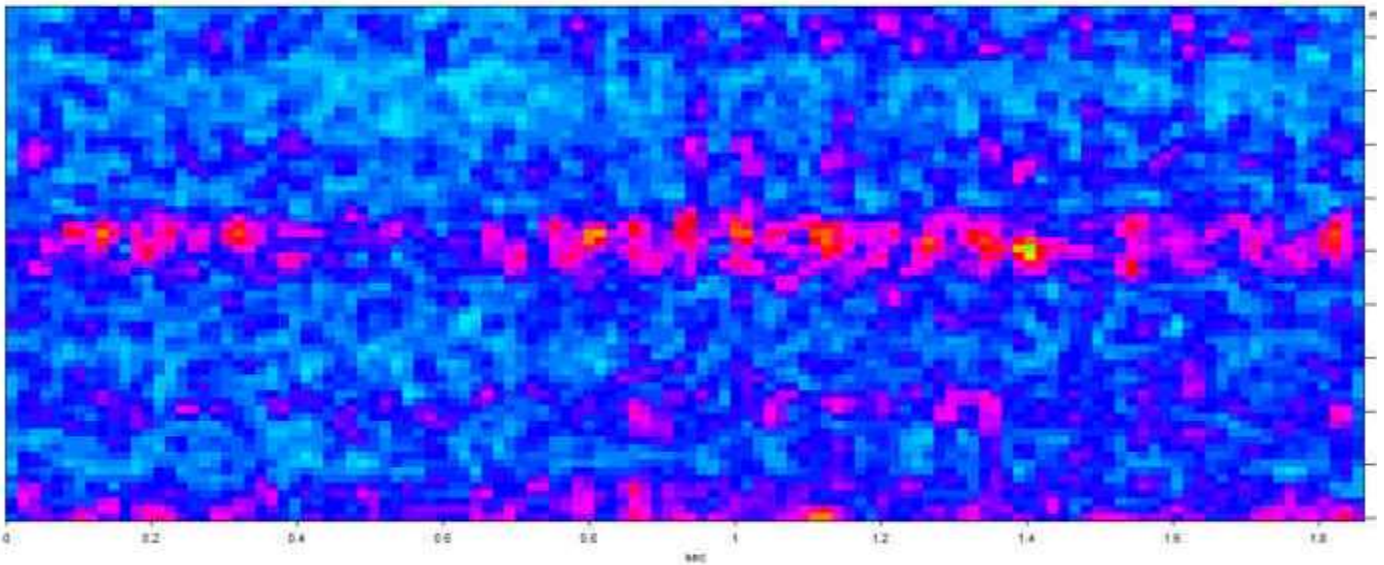
This is a spectrogram of some randomly selected radio white noise showing frequencies up to 3000 Hz. You can see there are bands of higher intensity sound around particular frequencies. In this example, these bands are around 0-500 Hz, 800-1000 Hz, 1200-1800 Hz, 2200-2600 Hz and above 2900 Hz. So, to test this sample you could filter out each of these bands in turn to see if any of them eliminated the apparent voice. A real voice, by contrast, should survive such treatment.
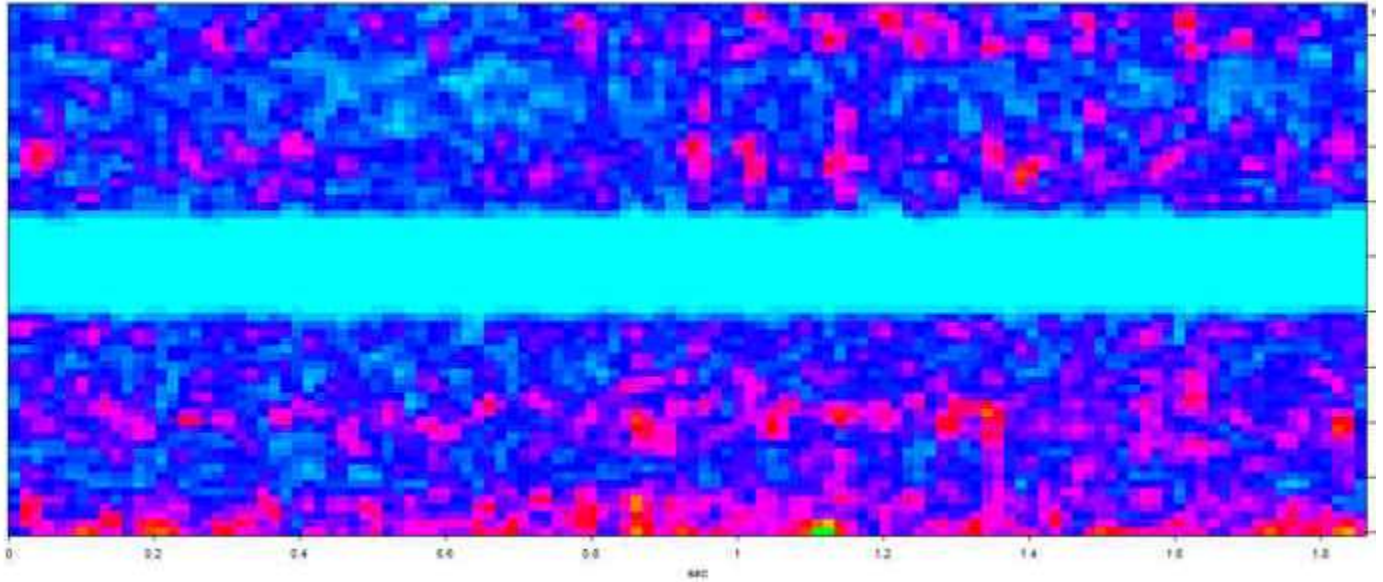
This is the same sound but with the major 1200-1800 Hz band removed by FFT filtering.

EVP samples

There are some sample EVP-type sounds available to hear. They are set up in the format of an experiment to demonstrate some of the points covered in this article. The sound clips are here.

Sound recorders for paranormal research

There are many portable digital sound recorders available now that are suitable for use in paranormal investigations. You should use digital recorders because analogue recorders may produce their own internal mechanical noises. However, before purchasing one you should be careful to check its frequency response. This should not be confused with the sample rate. Though many digital recorders often have an impressive sample rate of around 44000 Hz, they may have a frequency response far lower than this. The frequency response is important because, if it is 3000 Hz (top limit) or less, it will act like a tight spectrum envelope (see 'Making Formant Noise' above) so making formant noise sound more like voices.

A cheap dictaphone may only have a frequency response around 3000 Hz so these should be avoided. Even a higher specification recorder may have different quality modes (or long/short play modes). Check what the frequency response is for these modes and select one giving at least 5000 Hz. Such 'high quality' modes will make for bigger sound files which can be difficult to handle (some software will not accept very large files). It may also be inconvenient because large files will require a big memory in the recorder. The best advice is to split your recording session into several sections, each producing sound files small enough for your software to handle.

If you make recordings with a frequency response of 3000 Hz or less, you will inevitably hear more 'voices' from what may well be only formant noise. You may also form the conclusion that

EVP is concentrated in the lowest frequency range, which may well be wrong. Using such a restricted frequency range may even distort normal voices so that they are no longer recognisable (leading to the idea of a voice of unknown origin being present). To do scientifically useful research you need to use a frequency response of at least 5000 Hz (which covers most of the normal vocal range).

Summary

Searching for paranormal voices, or EVP, is more difficult than it might at first seem. As well as eliminating real voices, unheard or forgotten at the time of recording, there are other, more insidious, natural causes to consider. It is perfectly possible to hear 'words', that are not really there, in certain forms of noise, because of the way the brain interprets speech. Sine wave speech demonstrates that all that is necessary to understand words is the relevant formants. The phoneme restoration effect demonstrates that when certain sounds within words are missing, and their place taken by noise, the brain will 'insert' the appropriate missing sound to make sense of a word. The result will sound like real speech.

Formants are recognised by the brain using frequency sound peaks in known ratios to one another. Such frequency ratios can occur by chance in random noise ('formant noise'). As few as two formants can produce a phoneme which the brain will interpret as part of a word. Someone listening to formant noise may 'hear' phonemes that are effectively manufactured by their brain. Given the random, ambiguous nature of the sound source, the brain can turn this collection of odd apparent phonemes into complete words and phrases that sound perfectly real to the listener. This is a 'top-down' process, so the exact phrases, though constrained by the apparent phonemes, are nevertheless determined by the listener. Suggestion can have a powerful effect in determining this message content, which is why you should never tell people what to expect when playing them EVP samples. It may be possible to test for the 'formant noise' that produces illusory voices by selectively filtering out certain frequency bands to break the spurious ratios that resemble formants. Audio editing, such as noise reduction, may actually make random noise more like formant noise.