Some thoughts on Jochem Sotschek's report: *On the possibilities of the identification of phonemes*

ANABELA CARDOSO

It is a real pleasure to be able to offer to the readers of the ITC Journal the important paper by Jochem Sotschek, the renowned acoustics expert in telecommunications at the German Telekom, Research and Technology Centre in Berlin, Germany that we publish in this issue.

Several experiments with Friedrich Jürgenson (FJ) devised and supervised by Professor Hans Bender (see ITC Journal Nº 40) yielded unexplainable utterances finally considered anomalous by Bender. To check the acoustic characteristics of the latter and help determine their anomality or otherwise, Professor Bender called upon one of the most proficient Telecommunications technicians of that time in Germany, i.e. Jochem Sotschek.

The present paper documents the work done with the acoustic samples containing the utterances recorded with Jürgenson, therein called 'recordings'.

The study of Instrumental Transcommunication results is a serious enterprise. The evidence provided by ITC communications can, for the first time in the history of the research into survival of physical death, demonstrate that consciousness appears to survive physical death. The objective nature and permanence in physical form (audio and visual recordings) of the evidence enables its scrutiny by the use of technical and scientific tools.

When Professor Bender worked with FJ the most updated methods available for speech recognition at that time were used by J. Sotschek to examine the 'recordings' (i.e., the anomalous electronic voices).

Nowadays, other methods, in the main digital, are also available for the same purpose. However, they are based exactly upon the same principles described by Sotschek in the work that we publish today.

Having said this, things may seem easy to interested ITC parties. However, I have a different opinion on this highly important tool for ITC researchers and operators. I believe that proper evaluation of allegedly anomalous recordings through such sophisticated analysis methods is not at the reach of a layperson. They are very useful and in most cases better than human listening panels as the paper reveals, but they demand expertise and state of the art software to be reliable. I don't think that the basic audio editing software, generally used by ITC operators, is of help for proper identification of the meaning of anomalous recordings. In the hands of technically minded people it can serve as guidance but, even so, it should not be taken as definitive proof. Dedicated software for speech recognition is expensive and demands skilled users. It also demands sophisticated equipment such as professional, high quality soundboards, computers, amplifiers and loudspeakers. These are the means used by the scientific police which cannot be adequately used by a normal person. They require years of study and practice and real expertise to yield valid results and serve as proof as happens with the police. We should keep in mind that speech recognition performed by the police can send to jail a presumed criminal. It is only natural that such work cannot be done by a lay person with Adobe Audition, just to give an example. We, ITC operators, should not forget this basic principle if we want to be respected something that, unfortunately, decreases daily because of our superficial behaviour, most likely well-motivated but unintentionally misleading.

The fact that the guidelines above are generally not taken into consideration is at the bottom of the highly objectionable posting in Internet of thousands of 'voices' and their contents, which, in the great majority of the cases, do not correspond to what the author, certainly convinced of its authenticity, publishes. This attitude gives an extremely bad reputation to the discipline and jeopardizes the

admirable work of our outstanding pioneers – Jürgenson, Raudive, Franz Seidl, among others, not to mention the most modern ones and certainly, also of our transcendental communicators.

Publication of interpretations based upon listening tests can also be very dangerous for the following main reasons:

1. Listening is a subjective process to some degree (see Müller, 2011).

2. Correct listening requires listeners with good hearing skills, as FJ already pointed out, and first-class playback equipment.

3. Anomalous electronic voices are normally mixed with noise. Even white noise disturbances can be dangerous, as Sotschek points out, but the so-called EVP-maker and other human-speech based acoustic supports are disastrous. As a distinguished Professor of Acoustics recently remarked about this 'method': "This is the most certain way to create pareidolia in the listener; it should not be used". I could not agree more.

4. Listeners should be native-speakers of, or highly proficient in, the language the anomalous voices are supposed to be in.

Reading the full translation of Sotschek's paper for the first time today, May 1, I was happy to verify that he recommends several steps that I had decided upon, based on my personal experience, years ago, namely that the listening should, in many cases, be done sound by sound and that the experimenter can pronounce and record the same words that he/she thinks comprise the anomalous recording and compare the graphic thus obtained with the supposed anomalous recording. For this rough inspection common sound editing software such as Soundforge, Adobe Audition, Wavepad and so on are adequate because the listener only needs to compare the graphics of his/her own voice and of the anomalous recording on the computer screen. In general, the latter will be corrupted with noise but it will still be a good way of roughly and easily checking the shapes of both sound waves on the computer screen.

I had this purpose in mind when I decided to repeat the anomalous sentences I published in the CD *Electronic Voices* that complements my book with the same title. I wanted to allow the

listeners to digitally compare both graphics, a rather simple procedure. Up until now, I did not receive any negative remark about my interpretation of the content of the voices. On the other hand, the Portuguese speaking listeners said that they could virtually understand every anomalous sentence and the majority of the native speakers of other languages corroborated this evaluation although, as it should be expected, they also stated that they could not understand the meaning.

My purpose with these lines is, once more, to call the attention of ITC operators and researchers to the need to be rigorous, precise and objective in the evaluation of ITC results, either our own or other published ones, and to determinedly discard what does not fulfil these requisites. Our marvellous communicators, who spare no efforts to speak with us, will be thankful for this attitude and we will also render a great service to the study of this discipline that contains in it the power to change the prevailing human paradigm (see Laszlo 2008). But this most important goal can only be achieved if we are backed by the scientific establishment, something which has not happened.

However, it did happen in Jürgenson's and Dr. Raudive's time and the Austrian scientist, Franz Seidl, remarked in 1970: "It is rare indeed for a special area of parapsychological study – as the phenomenon of the voices is – to awaken the interest of serious men of science and researchers to the extent that the source of these voices, which emanate from beings outside the Earth, is studied with the most modern of investigatory methods in universities and research laboratories, with the result that the reality of their existence is established." (Appendix to *Breakthrough* 1971)

I think we should rightly ask what has happened since then to "the interest of serious men of science and researchers" because it is no longer visible. Many factors contributed to the present situation but one of the most important ones, if not the most important one, is the lack of rigour, critical sense and self-criticism of many ITC operators. Our communicators and our outstanding pioneers do not deserve their work to be ignored the way it presently is. Let us endeavour to change the current scientific approach to ITC. The

issue at stake is too serious and vital for humanity and for planet Earth to be treated this way. We absolutely need to understand this and the decision is in our hands.

References

Bender, Hans. (2011). *On the Analysis of Exceptional Voice Phenomena on Tapes.* Pilot studies on the "recordings" of Friedrich Jürgenson. ITC Journal Nº.40, pp. 61-78.

Cardoso, A. (2010). *Electronic Voices: Contact with Another Dimension?* Ropley, Hants, UK: O Books - John Hunt Publishing, Ltd.

Cardoso, A. (2010). CD: *Electronic Voices* (available from www.itcjournal.org).

Laszlo, E. (2008). *Quantum Shift in the Global Brain.* Vermont, USA: Inner Traditions.

Laszlo, E. (2008). An Unexplored Domain of Nonlocality: Toward a Scientific Explanation of Instrumental Transcommunication. *Explore, September/October, Vol. 4, Nº 5,* 321-327.

Müller, E. (2011). The Subjectivity of the Listening Process. *ITC Journal 41*, pp. 33-45.

Raudive, K. (1971). *Breakthrough: An Amazing Experiment in Electronic Communication with the Dead.* Gerrards Cross, England: Colin Smythe.

ITC - Historical Perspective

Editorial Note

Readers will recall that we published in issue 40 (pp 61-78) Professor Hans Bender's account of experiments with Friedrich Jürgenson and the methods used to assess the paranormality of recordings where likely anomalous utterances were detected. Those were labelled 'recordings' («Einspielungen» in German) and therefore, the same word used in the present report by Professor Bender's invited telecommunications expert, Jochem Sotschek, should be interpreted in that sense i.e., the recording of anomalous electronic voices.

On the possibilities of the identification of phonemes**

JOCHEM SOTSCHEK

The utilisation of visible-speech-techniques and other methods for the analysis of tape-recording 'playbacks'

A summary is given on the methods of determining speech components in the technique of communication. By means of these methods, analogous conclusions can be drawn for the analysis of 'play-back-phenomena'. Techniques and possibilities of visible-speech-techniques are explained by various examples.

Auditory experiments in audiometry and telecommunications

The usual and easiest way to recognise speech information is based on auditory experiments. The speech information to be tested is presented for identification to one or several subjects, for instance

over headphones. This method is generally applicable. It is used, for instance, in audiometry for measuring and evaluating the human listening behaviour or in telecommunications to evaluate the transmission quality during voice transmissions over unknown channels.

In audiometry, groups of uniformly shaped words (multisyllabic numbers or monosyllabic substantives) chosen from a greater repertoire are played to the subject to recognise the word samples (Freiburg word test, DIN 45 621).

For measuring the speech quality in telecommunications one can present language elements (sentences, words, syllables) to a testing group where the audibility is a measure of the transmission quality. Mostly in these auditory experiments sense-empty syllables, so called logotoms, are used, which auditory impressions have to be noted by the subjects. They are performed from a fixed repertoire of initial sounds, vowels and ablauts by randomly composing a great quantity of different syllables. The syllable components are sound elements of at least one certain language; sound lists consisting of sounds of all spoken languages are also used. However, in every case the number of different sounds is limited: a list published by the International Association of Telecommunications Administrations contains 45 different initial sounds, 5 different vowels and 49 different ablauts. The percentage of correctly understood syllables in relation to the total number of transmitted syllables is called the syllable intelligibility of the transmission channel concerned.

Auditory experiments for recognising 'recordings'

Conditions similar to those in the previously described method for recognising sense-empty syllables are present during recognition attempts when listening to 'recordings'. But in contrast to the aforementioned examples here the possible repertoire of voice elements is not taken only from one or several languages, which are usually known, but rather are completely unlimited. This applies not only to the number of different sounds but also to the rules after which the individual sounds are later put together into syllables and words in the described experiments.

In the Freiburg word test the listener only had to compare the sound heard with a rather limited vocabulary, and this just taken

from his mother tongue. He soon realised that the test words were composed of only two – or multisyllabic – numbers or represented a collection of monosyllabic substantives. Although the members of testing groups in intelligibility measurements do not know all these different logotoms individually as syllables they can split the sound heard into sound parts and classify it into a repertoire of initial sounds, vowels and ablauts they have learned over time. Also in this case there is a strictly limited pool of possible language elements.

In the case of trying to recognise 'recordings' these facilities of a limited repertoire do not exist for the listener. He would make it too easy for himself if he would assign to that which he heard only words or parts of words (i.e., the sounds) from his own language or vocabulary. But the possibility of extending this to other languages that the subject also knows does not provide any major improvement. Proof of this is shown when a number of 'recordings' were differently evaluated by German subjects having good English language skills and American subjects with excellent German language skills.

During listening attempts of 'recordings' the subjects should care less about the 'supposed' semantic content of the speech samples but rather document the phonetic content.

Because the two criteria cannot always be cleanly separated – a plausible sense is too easily interpreted into the heard sound – the sound content should be analysed at least during the processing of the listening results sound by sound. Only after a statistical analysis of logged answers of an enormous number of subjects one can conclude from the accumulation of sound groups a possible word context.

Analysing in accordance with these considerations the protocols of only a small number of subjects – for instance the results of 10 or 20 subjects – the sound analysis nevertheless will give only random results, which are statistically not significant. Naturally, the question immediately arises of how many subjects are necessary to achieve statistically secured results for recognition of the sound content of 'recordings'. For this purpose experimental results from telecommunications can provide valuable guidelines. To provide an estimation of the required number of independent hearing tests of

the same material, well-known connections of intelligibility measurement may be used.

There are results of syllable recognition measurements on a wide range of transmission systems. In this connection the results captured on transmission systems evoking a similar speech impression as in the 'recordings' are of special interest. The most important parameters are the transmitted *frequency band* and the *signal to noise ratio* which results from the ratio of the signal value (speech – i.e. 'recording') to the value of all disturbances (i.e. noise and hidden speech). Due regard must be given to the spectral distribution ('colouration') of the disturbance signal. All these characteristics can be measured quite accurately. From the Visible Speech Spectrogram of the 12 'recordings' of F. Jürgenson, in Mölnbo, Sweden, it can be seen that the frequency range of these 'recordings' extends between 200 Hz and 6...7 kHz, the range of perceiving 'normal' speech. The disturbance signals simultaneously occurring with the 'recordings' were in the same range. These spectrograms perform a three-dimensional visual presentation of a speech or sound event. Information about the signal noise ratio of the sound event can be obtained by a level scriber. Here, the signal will be recorded over time without any frequency evaluation.

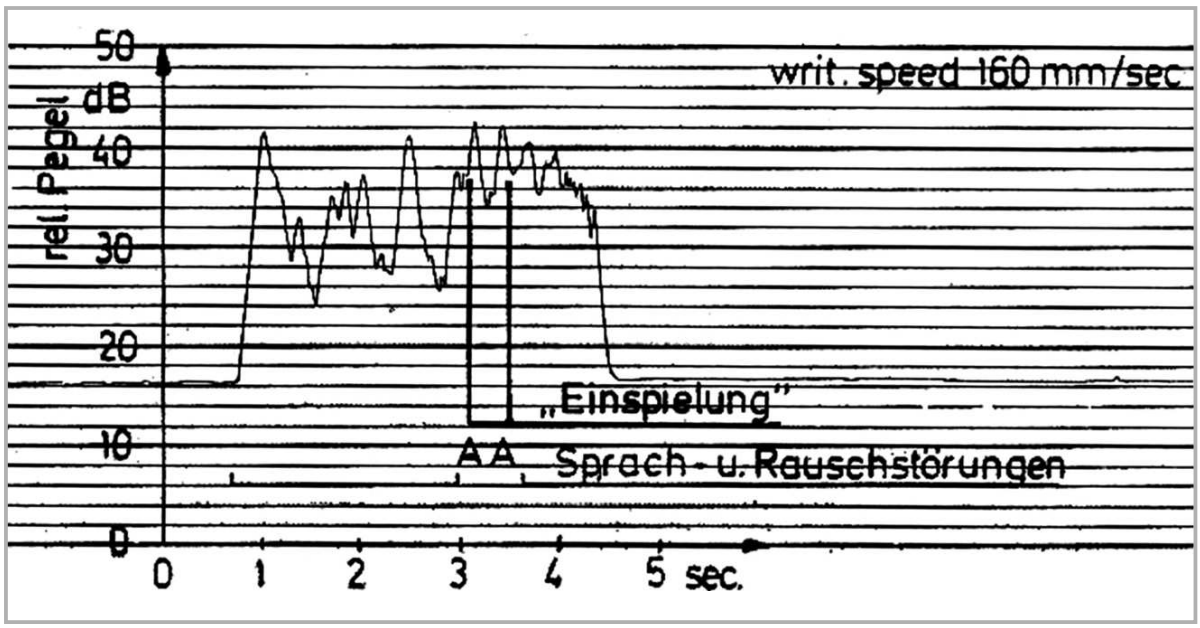

Figure 1

Based on the twelve aforementioned 'recordings' the signal to noise ratio of the voltage levels between 'recording' and disturbance signal was determined to be in an area between +10 and -10 dB. (Because of the large range between the lowest and the highest possible value, amplitude specifications of sound events in telecommunications are made in the measurement decibel [dB], the twenty times logarithm of a voltage or sound pressure ratio). Figure 1 shows such a level scriber writing. It belongs to the outlined Visible Speech Analysis of the 'recording': 'TAN(NER) TAN(NER)' in figure 3. In the level scriber recording of these speech samples only the most powerful sound 'A' is shown. In the above example the signal to noise ratio compared with the most powerful vowels is around 0 dB. From intelligibility measurements, i.e. by Fletcher and Galt[1], the context between syllable understanding and signal to noise ratio is well known. These results apply to sense-empty syllables as speech material and white noise as disturbance signal (a noise signal is called 'white noise' if it has constant power in equal frequency intervals). Desired and interfering signals have been in bandwidths limited to 125 ...5700 Hz, which approximately corresponded in the frequency range to the conditions during the 'recordings'. However, the nature of these 'artificial' disturbance signals only approximately corresponds to the actual combined speech and noise interferences. For this still quite favourable signal to noise ratio of +10 dB, which occurred in only two cases during the 'recordings', Fletcher and Galt reported a syllable intelligibility of 55%; in the case of a signal to noise ratio of -10 dB a syllable intelligibility of 5%. If one assumes that the 'recordings' could be reasonable speech, i.e. single words, one can specify the result according to other authors[2] in word intelligibility. In this case the word intelligibility is about 90% (for +10 dB signal to noise ratio) and about 10% (for -10 dB signal to noise ratio).

What does that mean for the practice of auditory experiments concerning 'recordings'? Considering first the 'recordings' as sense-empty phoneme compositions, for the specified signal to voice ratio there are only 5...55% correct answers as an average of a large number of individual decisions. In logotom measurements the evaluation of 1000...2000 logotoms per transmission system is sufficient in this case. If, however, a limited stock of real words of a language is assigned to the 'recordings', the hit rates in auditory

experiments under the same circumstances are between 10 ... 90%. Because it cannot always be presumed that the language is known to the subject, this hit rate is certainly too high. So, transferring the conditions and measuring methods of intelligibility tests to recognition attempts in the above 'recordings', one would have to perform at least 1000 independent auditory experiments. In the case of an adverse signal to voice ratio only about 5 ... 10% of the results will be correct. The evaluation method should allow recognizing this small number of hits. Due to an assumed limited word repertoire of the subject the evaluations would not be sufficiently detailed. As already explained, a better result can be expected only by an analysis performed sound by sound. However, in the case of the currently known 'recordings', which unfortunately are strongly disturbed, this method seems less promising.

Speech analysis from Visible Speech Representations

A completely different method for recognizing speech sounds is offered by the analysis of *Visible Speech Spectrograms*. The Visible Speech Spectrograph, a modern device in linguistics, is an analysis device that allows the visible representation of a sound event and its electric voltage in three coordinates: time, frequency and amplitude.

The spectrograph manufactured by Voiceprint Laboratories and used for these studies has a 300 Hz wide fixed filter which glides slowly as a 'frequency window' over the whole frequency band to be examined. It provides within 80 sec an analysis of a 2,3 sec lasting sound event stored on a tape.

Analysing, for example, a voice signal, the spectrogram gives a spectral representation of the energy distribution over time (abscissa axis) of the speech sample. Thereby the frequency of the event is plotted on the ordinate axis (frequency marks at every full kHz) and the different size of every portion is presented by different strong blackening. The device enables the representation of a dynamic range between 0 dB (no blackening) and about +50 dB (strongest blackening).

Now we know that the different voiced speech sounds have energy concentrations in certain frequency ranges, the so-called *formants*. Every voiced speech sound has several characteristic formant areas with characteristic frequency positions. Thus, with the

recognition of formants in certain frequency positions in a spectrogram, which are not identifiable during listening to the speech samples due to disturbances, it is possible to reconstruct under certain circumstances the original sound from the physical analysis of the speech sample.

Further support is offered by another property of voiced speech sounds generating particular speech-specific characteristics in a spectrogram. This special property results from the way in which the human voice formation system produces voiced sounds. Here, initially a constant air stream is pressed from the lungs as an energy source through the epiglottis of the larynx. The vocal folds forming the epiglottis are stimulated to perform periodical pulse-shaped vibrations. Their fundamental frequency determines the pitch of the intonated speech sound. The vocal fold frequency may be modified arbitrarily within certain limits. It varies according to the way of speaking within a sentence. However, during normal speech it is usually located within a given range. For women's and children's voices this lies between 220 Hz and 330 Hz, for male voices between 120 Hz and 160 Hz. Then the periodical pulse-shaped air stream, which is very rich in harmonics, is passed through the three variable cavities of the vocal tract (nasal, oral and pharyngeal cavities). According to the adjusted volume of the three cavities the desired vowel is formed by their different resonance frequencies.

Unvoiced sounds are produced by airflow or shutter sounds inside the vocal tract, on the teeth or lips without the involvement of the vocal folds. The same mechanism of voice generation applies to the ‘unvoiced’ soft speech where also no vocal fold oscillations occur.

A third group consists of voiced consonants, in which in addition to non-periodic sound components occurs a periodic sound portion due to vocal fold oscillations.

These vocal fold oscillations, described above as a characteristic of specific speech sounds, can also be seen during the analysis with the Visible Speech Spectrograph. If they occur in particular frequency ranges and can be assigned to particular formants, they perform an important criterion for differentiation between useful signals and interfering signals, occurring in addition, during voice recordings. These interfering signals, which mostly have a non-

periodic course (noise), are mapped accordingly, randomly, in the Visible Speech presentation. However, if the disturbance also consists of speech, a differentiation of the two speech samples becomes very difficult. It would be conceivable at best only during a simultaneous occurrence of different vocal fold frequencies (i.e. different pitches of the two voiced speech samples). In this case unvoiced sounds from different speakers are not distinguishable at all.

To illustrate this, the Visible Speech Spectrogram, in figure 2, of a clearly undisturbed voice sample is shown. It is the sentence 'EINSPIELUNG NUMMER EINS' pronounced by the male speaker B.

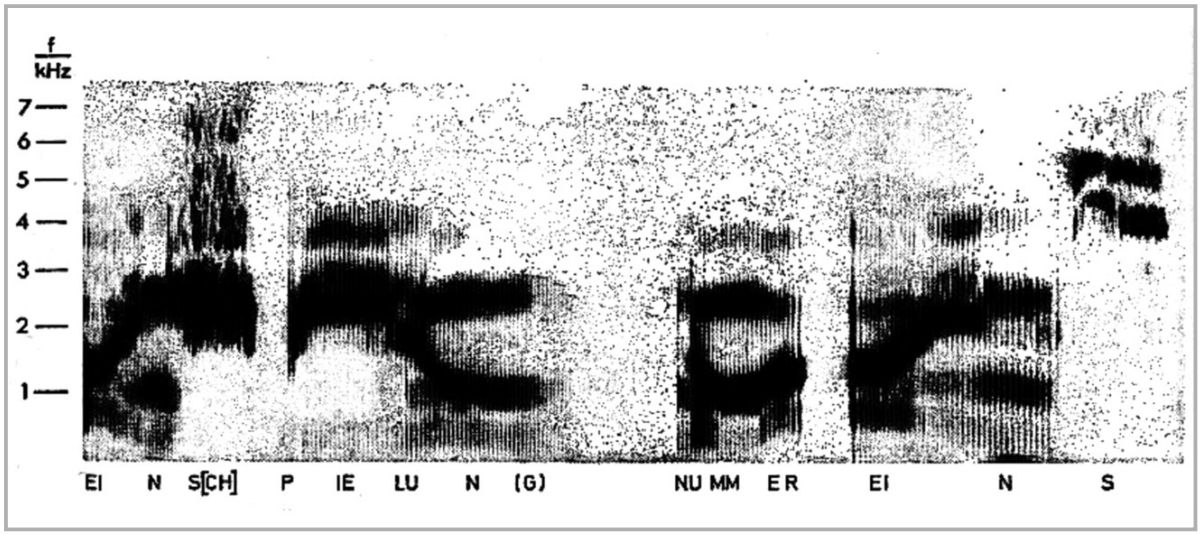

Figure 2

In the figure the abscissa describes the course of time of the speech sample from 0 to 2.3 sec and the ordinate shows the frequencies from 1 to 7 kHz. Clearly the individual restricted formant regions of the voiced sounds can be recognised, in the darkest black areas. The broadband character of the sound 'SCH' in 'EINSPIELUNG' ranges from 2 kHz to 7 kHz and beyond. The sound 'S' in 'EINS' also extends over a wide frequency range. The representation of the voice lip vibrations of vertical scanning is clearly visible with all tones together with the voiceless tones [the white areas in] SCH, P and S. The distance between the stripes indicates the period duration of the vocal fold oscillations. It can be seen that this period duration varies within a word or a whole sentence: The pitch of the speech changes with the sentence melody.

The syllable 'EIN' is intoned higher than the ending -'IELUNG', the sound 'N' in 'EINS' is spoken very deep (long period duration of the vocal fold oscillations).

With the described analysis method the opportunity is created to finally detect speech sounds, which are not clearly recognizable by the ear because of disturbances, or by optical decision criteria based on the position of the individual formant regions, considering the presence of vocal fold frequencies. The prerequisite is of course that the noise and the speech sound are distinguishable in the visual representation. With regard to the frequency selection as well as to the reception of volume differences, the analysing skills of the ear are very sensitive in a wide dynamic range and seldom are outdone by technical means. But for the exact technical measurement of sound events, often the ears are only accessories.

If useful and interfering signals are exactly in the same frequency range and the disturbance has a higher energy than the useful signal it is very questionable whether a separation of the two signals could be made naturally. However, there are also cases where a sound event is made inaudible to the ear by another sound event which has not the same frequency as the first but a neighbouring one. In the case of this ear-physiological phenomenon, the so-called occlusion, the determining of this hidden and inaudible signal by metrological analysis, i.e. as described above, is quite possible.

A sound detection of 'recordings' using determinations of voiced/unvoiced sounds is of course only viable if the analysis of the 'recordings' performed in the described manner provides representations of vocal fold frequencies. However, this can be assumed if the speech character of a 'recording' is clear by simple listening and most of normal spoken speech sounds have verifiable vocal fold oscillations which are also necessary for their generation. As shown in the next section, these premises are obviously applicable to 'recordings'.

Application of Visible Speech Analysis to 'recordings'

Figure 3 shows the Visible Speech Analysis of a 'recording'.

Next to the 'recording'-signal it displays interferences in the form of noise and speech. However, these interferences [made by the experimenters] are easily distinguished by listening to the recorded

material because, although both appear approximately equally loud, the ‘recording’ has a higher pitch than the language interferences.

The ‘recording’ has the characteristics of a female voice.

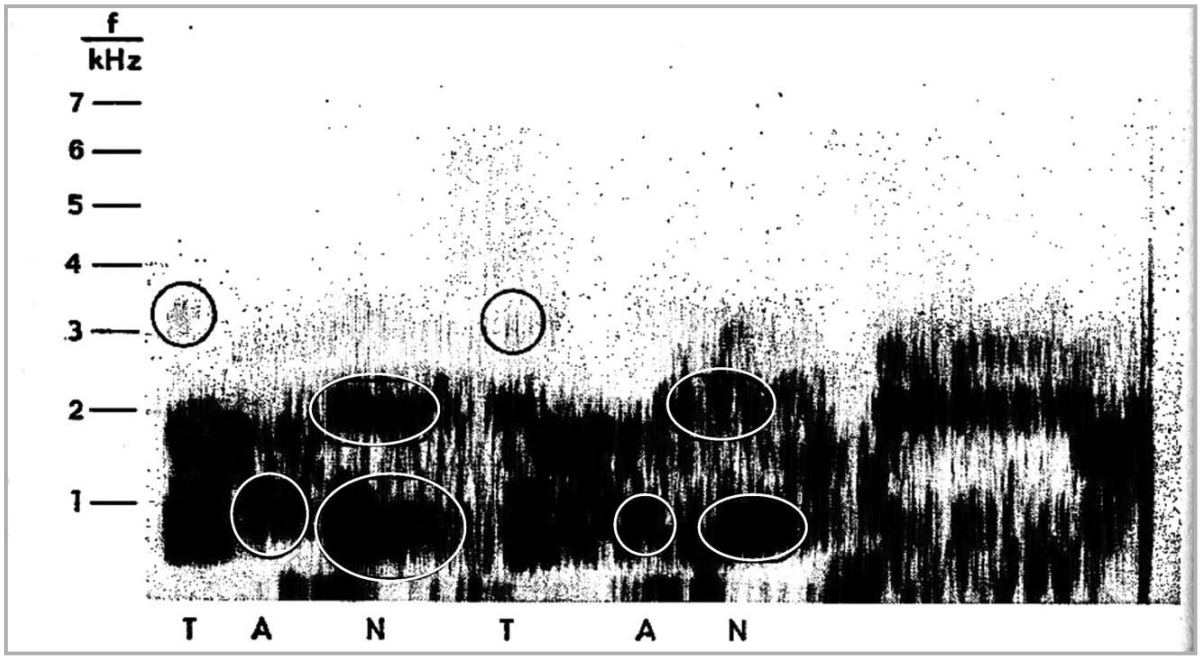

Figure 3 (3)

In the analysis the formant areas of the ‘recording’ are highlighted by framings. The very quiet sounding ‘T’ is only visible in a small range of 3...4 kHz although it contains additional energy portions particularly in higher frequencies. The formant of the sound ‘A’ is visible. Mainly in the first syllable the fine structure caused by the speech fundamental (vocal fold oscillations) is signified. However this is clearly visible in the first formant (about 1 kHz) of the first ‘N’-sound. In comparison with the fundamental frequency representation in figure 2, one recognises a quite high speech fundamental frequency. This fits with the female voice pitch, subjectively perceived, of the ‘recording’. The other formant structures belong to the speech interferences [by the experimenters].

The analysis of a further ‘recording’ is shown in figure 4. Here, too, the suggestion of a fundamental speech frequency is visible particularly at the first ‘O’.

In this Visible Speech Diagram the limitations of this method of sound analysis become evident. The sound ‘F’ (F corresponding to

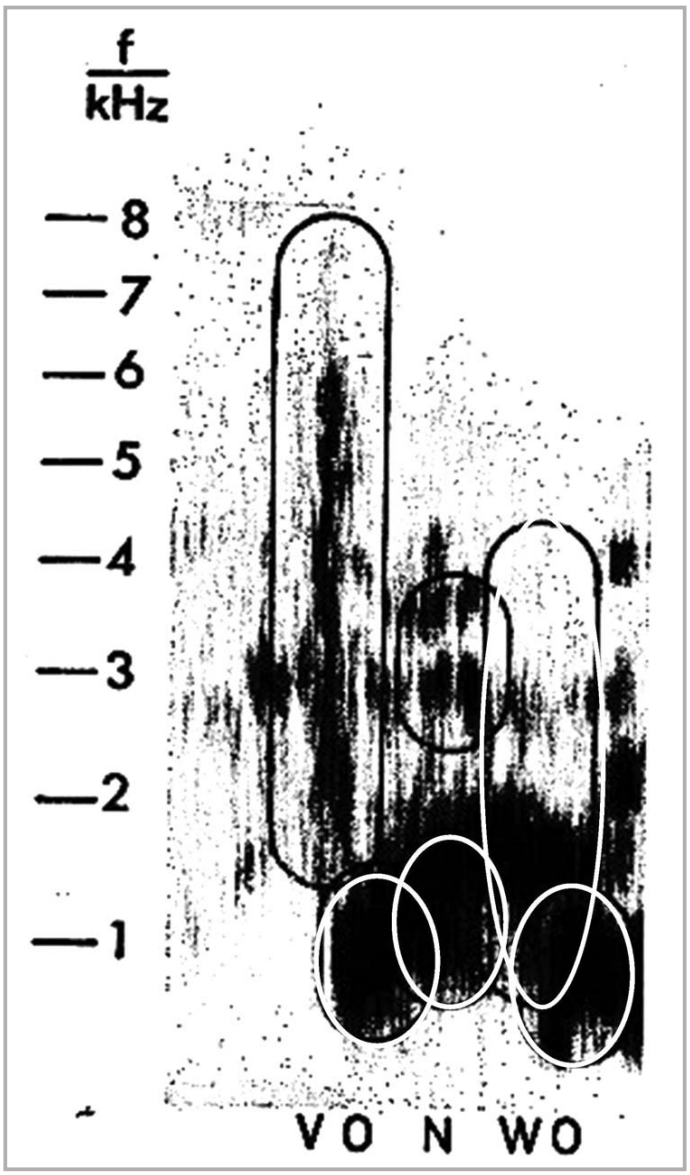

Figure 4

the V of the word VON) is clearly visible in the diagram. In accordance with the Visible Speech Analysis the vowels “O” [of VON WO] sound more like "A" than “O” and the sound "W" is only indistinctly visible. Thus, for "reading" the Visible Speech Diagram the phonetic listening impression is necessary as an aid.

An example for the case in which the listening impression provides significantly more information than a Visible Speech Analysis is shown in figure 5. This ‘recording’ was done at the end of a conversation in which, from the beginning, it had been agreed that it should be finished with a pause of about one minute.

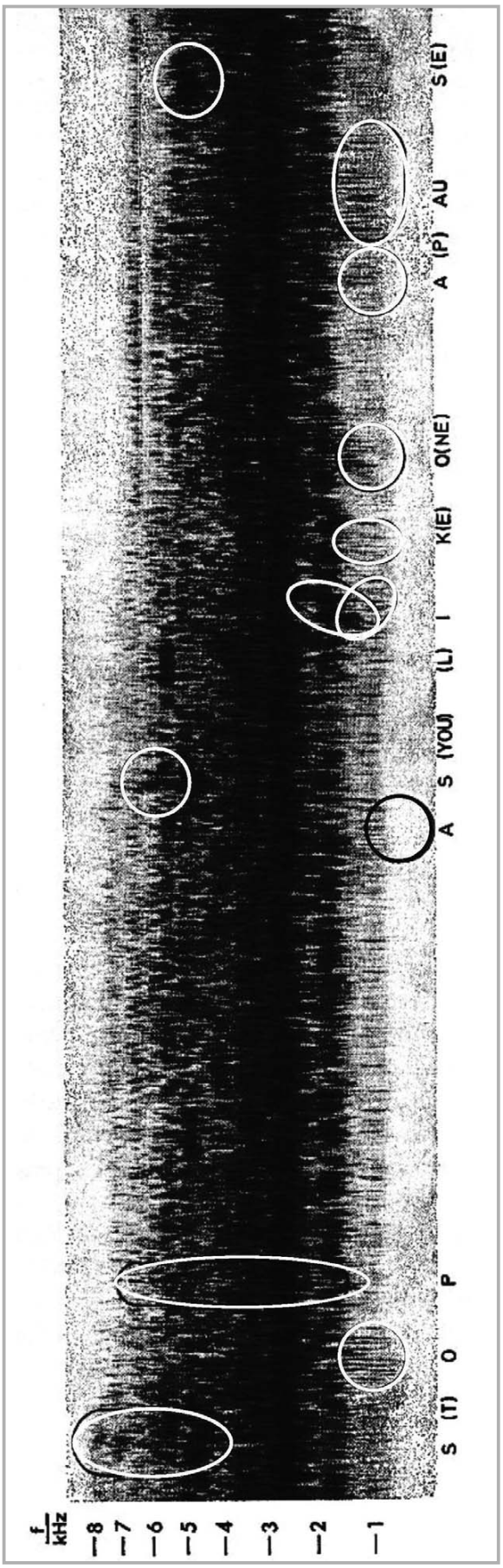

Figure 5

That is why the analysis shows on a first superficial observation only the (very highly amplified) image of an almost random noise interference with a predominance of certain frequency ranges, particularly at about 2.8 kHz. However, if we examine synchronously a listening attempt with the Visible Speech Diagram, we can find some temporal and frequency bounded energy concentrations. They are in accordance with sounds in the heard text 'STOP – AS YOU LIKE ONE (?) A PAUSE'. The analysis also confirms the American pronunciation of the 'O' as an 'A' in the first word, already noticeable during a first listening attempt. If here in figure 5, in contrast to the view expressed in the first part of this paper, the heard sounds are given a meaning, this is done solely for a better interpretation of the Visible Speech Diagram. The correctness of this interpretation should not be assumed at all. This also implies the sounds and sound sequences appearing in brackets, which can't be verified in the diagram itself.

As a final example of an analysis of a 'recording' we will examine a whispered 'recording' (figure 6). This 'recording' occurs within a sentence spoken by B., during a break which is only disturbed by noise signals. Of course this is very suitable for sound detection. On the other hand, the whispered expression is very low, so that only the first syllable in the Visible Speech Diagram can be verified. In addition, the characteristic structure of the fundamental frequency of the speech in the diagram cannot be used here to distinguish between speech and noise because soft speech arises only due to

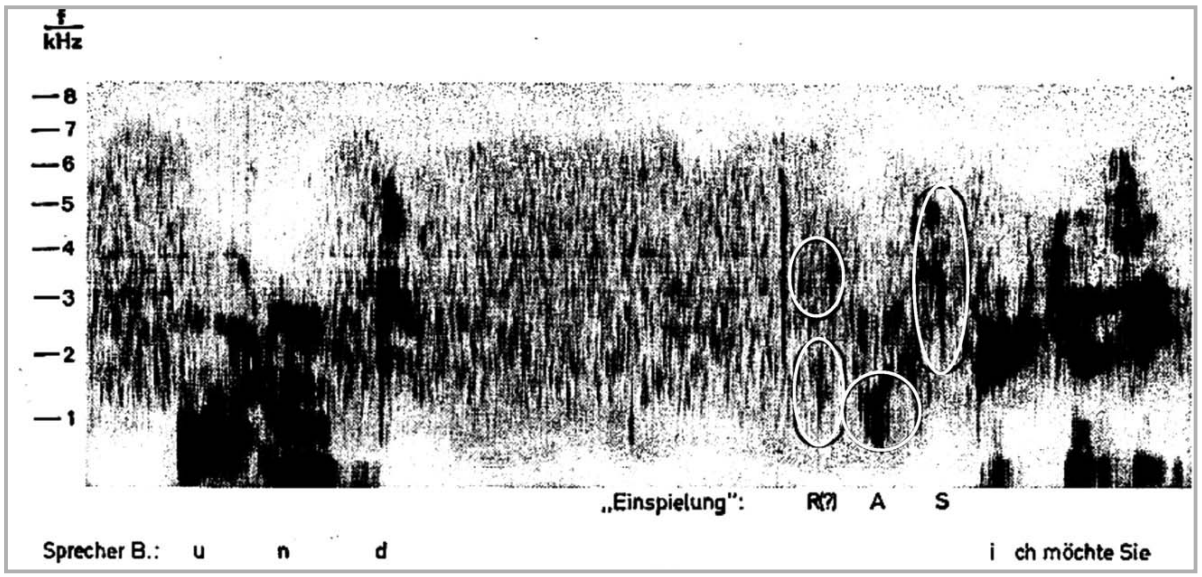

Figure 6

noises in the voice formation channel without involvement of the vocal folds. In figure 6 the actual 'recording' (denoted by 'Einspielung') and the utterances of speaker B. ('Sprecher B.') are clearly distinguishable. For a better clarification of the sound content of the 'recording' in figure 6 there was another recording produced in which the speaker B. repeated the alleged meaning of this 'recording' and whispering occurred during his utterances. Figure 7 shows this repetition. The relationship between the whispered voice sample and the first syllable 'R (?) AS' (the only part shown) in the 'recording' can be clearly recognised. In figure 6 the formant area of the 'A' and the area of the 'S' are relatively clearly identifiable.

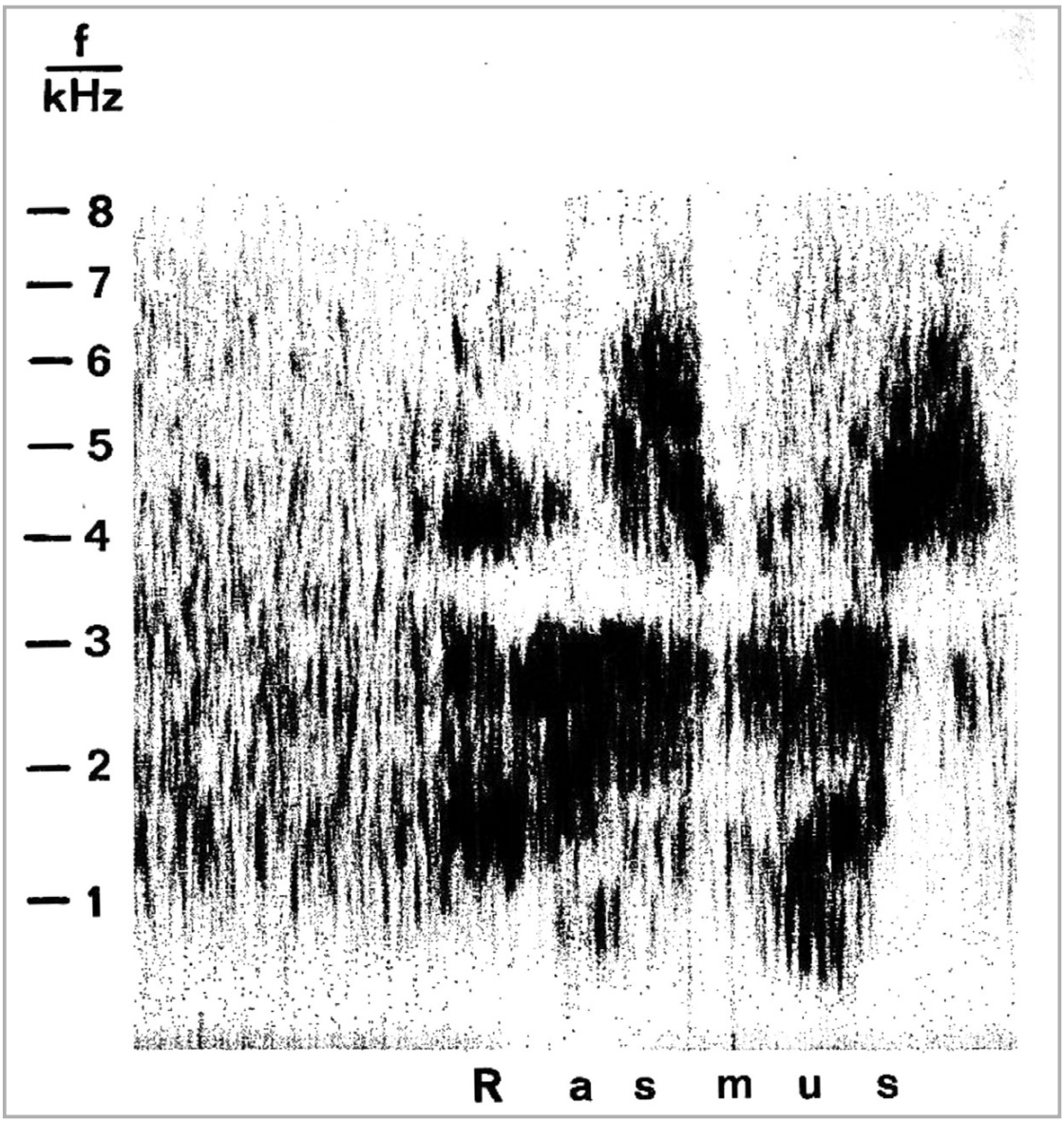

Figure 7

Because of the absence of the fundamental frequency of the speech in the figures nothing can be stated about the possible identity of the two speakers nor is the representation of the 'recording' sounds suited to assign the depicted energy concentration with exact certainty to certain speech sounds. The normally advanced argument that the 'recording' in figure 6 is produced by the speaker B. himself, can be responded to in that it seems impossible to switch from an unvoiced and very low whispered utterance to a voiced and loud spoken 'I' in the word 'ICH' that followed, in such a short period.

This period can be seen directly in figure 6. Taking into account only the visible prefixes of the 'recording' it takes about 0.05 sec. from the end of the 'S' until the beginning of the 'I' of 'ich möchte Sie'. Taking into consideration the second syllable [MUS], which is not visible in figure 6 but audible, this period is certainly still shorter.

Concluding remarks

From these examples it can be concluded that the measuring and assessment methods of telecommunication undoubtedly can provide assistance in identifying 'recordings'. This applies to both the methodology of auditory experiments and the application of the Visible Speech Analysis to the sound detection of 'recordings'.

Although there are limitations to the sound comparison between 'recordings' and *de facto* spoken speech, in specific cases the evaluation of such an analysis can give some indication of what the words/ speech might be in a recording. However, the Visible Speech Analysis does not give the opportunity to determine specific sounds with absolute certainty, but it only offers a help in combination with the listening impression of a 'recording' to indicate its sound composition. In the case of voiced speech sounds, the sound determination becomes much simpler with this analysis. An exact certainty, with the conclusion drawn on the basis of the Visible Speech Analysis, cannot be achieved.

The method of the visible presentation of voice and noise signals is most suitable for precise timing between the individual speech sounds, from which it is often possible to draw conclusions regarding the origin of speech utterances.

Notes

*Translation by Professor Uwe Hartmann of the German original article by J. Sotschek: Über Möglichkeiten der Erkennung von Sprachlauten - Zur Anwendbarkeit des Visible-Speech-Verfahrens und anderer Methoden auf die Analyse von Tonband-"Einspielungen").

*Published with the kind permission of *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie,* Jahrgang 12, Nr. 4, 1970, S. 239-254. (Journal of Parapsychology and Border Areas of Psychology). Grateful thanks to Professor Eberhard Bauer for providing the original text for the present translation.

1. Fletcher, H. and R. H. Galt: The Perception of Speech and Its Relation to Telephony. *Journ. Acoust. Soc. Amer. 22* (1950): 89-151

2. Flanagan, J. L.: Speech Analysis, Synthesis and Perception, in: *Kommunikation und Kybernetik in Einzeldarstellungen, Band 3,* Berlin/Heidelberg/new York 1965, S. 234 ff.

3. The delineation in white of areas of sound in Figures 3,4,5 and 6 has been added to make them more visible on the page than the original black outlining.