acta acustica 4 (1994) 215-222 June 1994, page 215

The cumulative spectral probability diagram: theory and experiments

Robert Ruiz and Claude Legros

Laboratoire d'Acoustique de l'Université Toulouse-Le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex, France

(Received 15 September 1993; revised 23 December 1993; accepted 13 May 1994)

Abstract. — This paper presents a new acoustical parameter of the speech signal. It only concerns the steady-state vowels of the french language and arises from their analysis by the linear prediction technique (L.P.C). It shows the frequency distribution of the sound level. Its theoretical method of calculation and its properties are presented and discussed. Finally, we describe its experimental application to detect the pilot's stress in a real flying situation.

Pacs numbers: 43.70Gr.

1. Introduction

The production of voiced vowels involves three stages: the power source, the vibration of the vocal folds and the time-varying filtering effect of vocal tract articulation. The first two stages lead to the creation of the acoustic excitation for the vocal tract and the third defines the acoustic timbre, perceived as the vowel uttered. The same vowel may be uttered with different excitation fundamental frequencies but, the same fundamental frequency applied to different articulatory settings of the vocal tract does not necessarily result in the same vowel.

The vocal tract acts like a time-varying filter which imposes its transmission properties on the frequency spectrum of the source. Thus, the spectrum of the excitation signal is modified by the acoustic resonances of the cavities of the tract during the sound propagation from the glottis to the lips. To separate this contribution from the periodic excitation, the homomorphic processing of the signal or the method of linear predictive analysis can be used.

In this paper, we study the transfer function of the vocal tract obtained from the linear prediction model. The vocal tract transfer function is then dissociated from the mean sound level of the signal to give the envelope spectrum zero mean. Considering these two properties, we propose the study of a new parameter of the speech signal which accounts for the frequency distribution of sound level. This gives the sound level distribution in the spectrum and allows the detection of possible spectral modifications. It can be used in the studies of intra- and inter-speaker variability of speech. Here, the experimental application concerns the detection of stress induced by flying incidents in aircraft.

After a brief overview of linear prediction analysis, the theory behind the Cumulative Spectral Probability Diagram is presented. Then its experimental application in detecting emotional disturbances of pilots by the analysis of the tape of a Cockpit Voice Recorder is discussed.

2. Theoretical aspects of the cumulative spectrum probability of sound level

2.1. *Basic relations*

The cumulative spectrum probability of sound level is defined from the spectral envelope of the steady-state vowel signal. This envelope is directly obtained from the autoregressive model (AR) by the linear prediction technique (autocorrelation method).

Before analysis, the signal is sampled into N frames whose duration is compatible with the hypothesis of quasi-stationarity of the speech signal. The linear prediction model is then calculated for each frame, leading to $(p+1)$ prediction coefficients $a(i)$, for $i = 0$ to p.

The transmittance $T(z)$ of the model is $T(z) = (\sigma / A(z))$ with the inverse filter:

$A(z) = \sum_{i=0}^{p} a(i) z^{-i}$ and σ the gain of the model (Atal & Hanauer, 1971; Makhoul, 1973, 1975a, 1975b; Markel & Gray, 1982).

The spectrum of the AR model, noted $P(\omega)$, is defined as:

$$P(\omega) = 10 \log \left| \frac{\sigma}{A\left(\mathrm{e}^{j\omega T_{\mathrm{e}}}\right)} \right|^2 \tag{1}$$

where T_{e} is the sampling period ($f_{\mathrm{e}} = 1 / T_{\mathrm{e}}$ is the sampling frequency).

Let $D(\omega)$ denote the following spectrum:

$$D(\omega) = 10 \log \left| \frac{1}{A\left(\mathrm{e}^{j\omega T_{\mathrm{e}}}\right)} \right|^2 \tag{2}$$

with $\omega = 0$ to (πf_{e}).

The linear predictor model is calculated by using the autocorrelation method: its mean is zero (all zeros lie inside the unit circle) and it is independent of the mean sound level ($20 \log_{10} \sigma$) (Markel & Gray, 1982).

Therefore, $D(\omega)$ is like a normalized spectrum of the steady-state vowel signal.

The process of estimation of $a(i)$ and construction of $D(\omega)$ is applied to each frame. In order to reduce the slow time-varying spectral information of the steady-state vowels, we have defined a mean spectral envelope.

2.2. *Mean spectral envelope*

We define the mean spectral envelope by the mean prediction coefficients (Ruiz, 1991). Each $\overline{a(i)}$ is the arithmetic mean of the coefficient of the same order $a(i)$ for all the frames of the signal.

The mean spectral envelope (Figure 1) is usually not very different from the envelope of each frame. In the case of steady-state vowels, the shape of the vocal tract typically changes very little during their utterance.

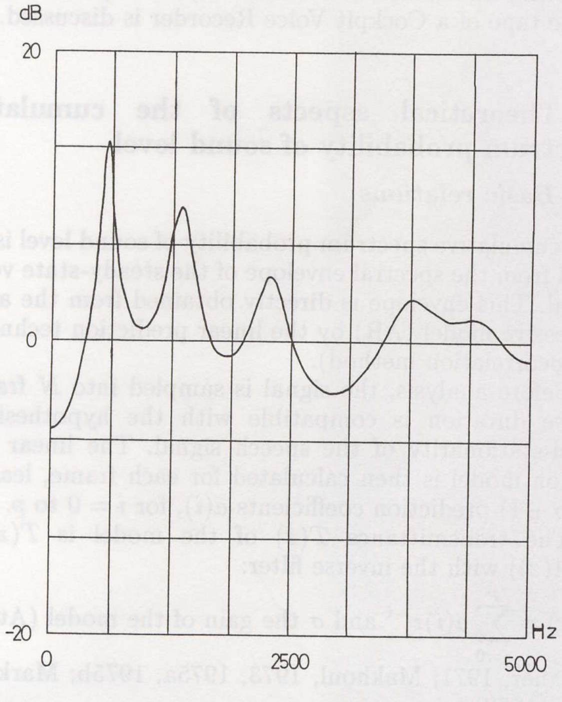


Figure 1. Envelope of a mean spectrum (steady-state vowel [a], male speaker).

2.3. *The histogram of the sound level*

The construction of the mean spectral envelope leads to N values of sound level issued from the N frequencies from 0 to $(f_e/2)$ Hz. The histogram, noted $P(L_i)$, $i = 1$ to C, is defined by the following values: $P(L_i) = (N_i/N)$, where N_i is the number of frequencies for which the sound level is in the class L_i (C is the number of classes and ΔL is the class width) (Figure 2).

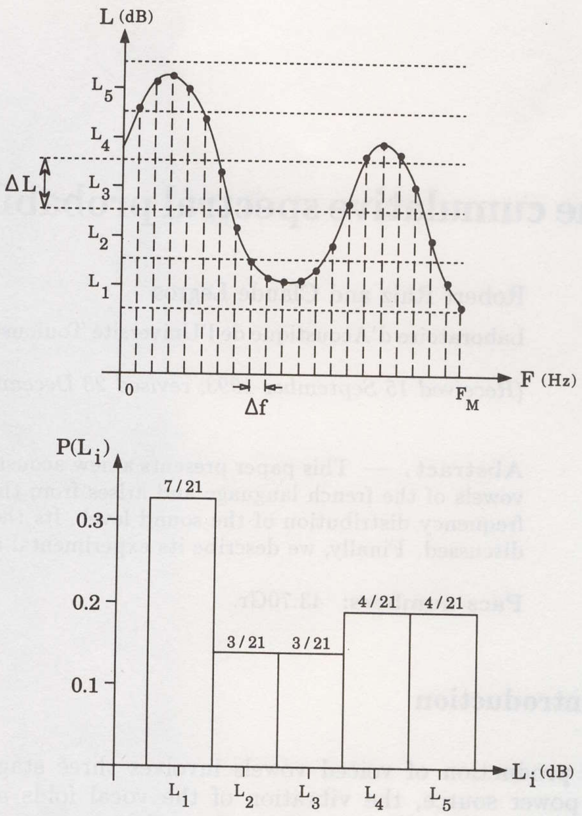


Figure 2. Spectral envelope (21 discrete frequencies) and corresponding sound level histogram.

It can be seen that the histogram will be the closest to the spectral distribution of levels when the class width is minimum and therefore the number of classes (C) is a maximum.

2.4. *The cumulative spectral probablility of sound level*

2.4.1. *Definition*

The Cumulative Spectral Probability (CSP) of sound level, for the class center L_i, is called the statistical spectral index, noted $\hat{I}_i$, and defined by the following relation ($i = 1$ to C):

$$\hat{I}_i = \left(1 - \sum_{j=1}^{i} P(L_j)\right) \qquad (3)$$

where L_i is the class center for which we want to know the sum of the previous $P(L_j)$ (i.e. those corresponding to class centers less or equal to L_i.

When we plot $\hat{I}_i$ *versus* L_i, we obtain a cumulative histogram. The corresponding smoothed graph (polynomial regression) is called Cumulative Spectral Probability Diagram (Figure 3).

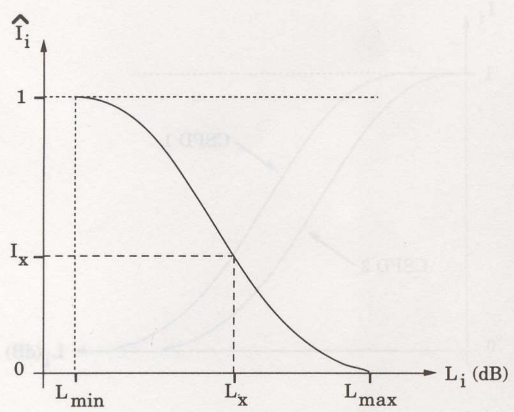


Figure 3. Diagram of Cumulative Spectral Probability (CSPD).

2.4.2. *Properties*

- The histogram of the sound level is not a good characteristic of a mean utterance from a group of different utterances or from frames of the same utterance: the minimal and maximal values of the sound level in each spectrum can be very different from each other. On the other hand, the CSPD is bounded for all dynamics between 0 and 1. So, an utterance will be defined by a mean CSPD rather than by a mean histogram.

We will calculate the CSPD of the vowel by way of the envelope of its mean spectrum.

- $\hat{I}_x$ is called statistic spectral index of the sound level L_x (Figure 3) ($0 \leq \hat{I}_x \leq 1$). Indeed, L_x represents the level reached or overrun by $(\hat{I}_x \cdot 100)$% of the frequency domain. For example, if $L_x =$ 20 dB corresponds to $\hat{I}_x =$ 0.4, it means that the level 20 dB has been reached or overrun by 40% of the frequency band 0-5000 Hz. One may also say that for a "bandwith" of 2000 Hz, the level is at least 20 dB.

So, $\hat{I}_x = 0$ indicates the maximum level on the envelope spectrum and $\hat{I}_x = 1$ the minimal. The slope of the CSPD indicates the dispersion of the level values.

Corresponding to the index $\hat{I}_x$, is the value L_x of the sound level. This value is no longer than that of a level class because of the continuity of the drawing. The result is that the level L_x is affected with an uncertainty of $\Delta L/2$.

- The bandwith of the formants can be obtained from the CSPD (Ruiz & Legros, 1994). To demonstrate this property, let us consider a simulated spectrum envelope with its corresponding CSPD (Figure 4).

For example, this envelope may be a resonance of a steady-state vowel spectrum. In our case, it is only a simple parabola.

The quantity $(100\hat{I}_x)$ is the percentage of discrete frequencies for which the sound level is greater or equal to L_x.

The total bandwith of the spectrum $(F_{max} - F_{min})$ represents 100% of the frequency domain considered. So, the estimated bandwith corresponding to the quantity $(100\hat{I}_x)$ is:

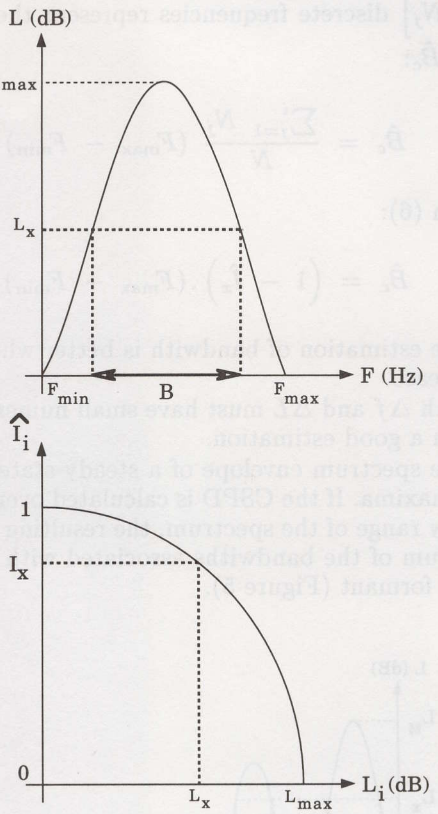


Figure 4. Calculation of the bandwith B at $-(L_{max} - L_X)$ dB from the spectrum envelope and from the corresponding Cumulative Spectral Probability Diagram: $\hat{B} = \hat{I}_X . (F_{max} - F_{min})$

$$\hat{B} = \hat{I}_x\,(F_{max} - F_{min}) \qquad \text{(Hz)} \qquad (4)$$

For $\hat{I}_x = 0$, i.e. for the sound level L_{max}, we have $\hat{B} = 0$ Hz, and for $L_x = L_{min}$, we have $\hat{I}_x = 1$ and $\hat{B} = (F_{max} - F_{min})$.

The bandwith $\hat{B}_c$ defined by the following relation:

$$\hat{B}_c = \left(1 - \hat{I}_x\right)(F_{max} - F_{min}) \qquad \text{(Hz)} \qquad (5)$$

corresponds to the frequency domain $\left[(F_{max} - F_{min}) - \hat{B}\right]$. It is the complementary one to $\hat{B}$ in the spectrum: $\left(B + \hat{B}_c\right) = (F_{max} - F_{min})$.

For $\hat{I}_x = 0$, $\hat{B}_c = (F_{max} - F_{min})$ and for $\hat{I}_x = 1$, $\hat{B}_c = 0$ Hz.

Indeed, the index $\hat{I}_x$ corresponding to the level L_x is (from (3)):

$$\hat{I}_x = \left(1 - \sum_{j=1}^{i} \frac{N_j}{N}\right) \qquad (6)$$

where N is the total number of discrete frequencies between F_{min} and F_{max}. Then, in proportion, the

$\left[\sum_{j=1}^{i} N_j\right]$ discrete frequencies represent the frequency domain $\hat{B}_c$:

$$\hat{B}_c = \frac{\sum_{j=1}^{i} N_j}{N} \left(F_{\max} - F_{\min}\right) \tag{7}$$

and from (6):

$$\hat{B}_c = \left(1 - \hat{I}_x\right) . \left(F_{\max} - F_{\min}\right) \tag{8}$$

- The estimation of bandwith is better when Δf and ΔL decrease.
- Both Δf and ΔL must have small numerical values to obtain a good estimation.
- The spectrum envelope of a steady-state vowel has several maxima. If the CSPD is calculated over the whole frequency range of the spectrum, the resulting bandwiths are the sum of the bandwiths associated with each maximum or formant (Figure 5).

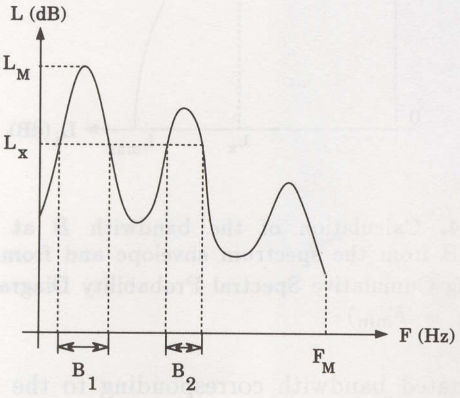


Figure 5. $B = (B_1 + B_2)$ is the bandwith (in Hz) at $-(L_M - L_X)$dB.

If we consider only the frequency range of one formant, it is possible, from (4), to calculate bandwiths. In this case, $F_{\max}$ and $F_{\min}$ are the frequency limits of the band only including the formant under consideration.

The CSPD can be obtained by frequency bands on the spectrum considered. When two complementary bands are defined, the comparison of the two corresponding CSPD leads to a spectral balanced criterion.

2.4.3. *The spectral balanced criterion*

Let us separate the whole frequency band $(0, f_e/2)$ into two complementary ones $(0, f_1)$ and $(f_1, f_e/2)$ and let us calculate the two CSPD (see Figure 6).

The area between CSPD 1 and CSPD 2 defines the spectral balance for the statistical indexes between 0 and 1. If it is great the spectra are said to be "unbalanced". If not, they are said to be "balanced".

A spectrum will be balanced if the sound level distributions in the two frequency bands are identical. When they differ the spectrum is unbalanced.

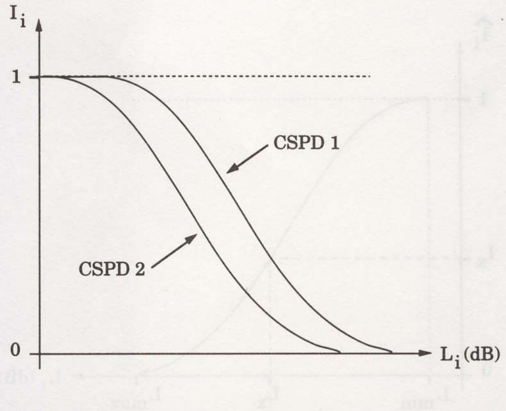


Figure 6. CSPD_1: diagram of cumulative spectral probability for the band $(0, f_1)$. CSPD_2: diagram of cumulative spectral probability for the band $(f_1, f_e/2)$.

Spectra of steady-state vowels decrease with frequency. So, these are naturally unbalanced, having greater energy at low frequencies.

The CSPD corresponding to the low frequency band $(0; f_1)$ is "far" from the one corresponding to the high frequency band $(f_1; f_e/2)$.

Now, these general considerations are applied to a real signal of monophtong in order to detect the acoustical correlate of stress.

3. Application of CSPD to stress detection

The CSPD has been used to investigate a tape from a Cokpit Voice Recorder (CVR) in order to investigate the quantification of emotional disturbance.

The variability factors of the voice of a speaker are numerous. They are mainly of a physiological, psychological, emotional and linguistic nature. Therefore, the utterance within the same language varies with sex, dimension of the phonatory organs, but also with the psychological and emotional state of the speaker. The same individual will utter a sentence differently according to the emotion he wishes to express. The vocal changes that can arise may occur when the speaker lies (Brenner et al., 1979; Horvath, 1978; Streeter et al., 1977), when he witnesses a tragic event which he has to describe (Williams & Stevens, 1972, 1981), or when he performs mental or psychomotor exercises (Armstrong, 1980; Armstrong & Poock, 1981; Brenner et al., 1979; Griffin & Williams, 1987; Mosko et al., 1983).

It has been previously demonstrated that for such situations, and also other areas (Legros & Ruiz, 1988, 1990; Ruiz, 1991), the fundamental frequency rises significantly from its mean value measured without disturbing factors.

The fluctuations of the other acoustical parameters of the voice signal (such as sound level, or formant frequencies) have only been estimated for some isolated cases (Griffin & Williams, 1987; Hecker et al., 1968; Mosko et al., 1983; Williams & Stevens, 1972).

Therefore, the fundamental frequency appears to be a vocal parameter able to quantify emotional disturbance.

Other studies, mainly Soviet and American, have been conducted using specific acoustic parameters sensitive to the emotional state of the speaker: a mathematical function of the energy in third-octave bands (Soviet ones), and frequency modulation of the fundamental frequency (American ones).

The results of earlier studies have led to the detection and characterization of the emotion as "positive" (joy, pleasure...) or "negative" (anxiety, fear...) in 95% of the cases. But, the experimental conditions and the results are not clearly detailed (Simonov & Frolov, 1973, 1977).

In the USA, the Psychological Stress Evaluator (PSE) detects and displays a slow frequency modulation of the fundamental frequency in the case of a lack of emotion. This modulation disappears when the disturbance grows. The PSE has been mainly tested as a lie detector (Brenner et al., 1979; Horvath, 1978) then by the N.A.S.A. during an Appollo mission to estimate the stress of astronauts (Older & Jenney, 1975). The voice analysis technique used by the P.S.E. is not predictable enough to guarantee its use in the evaluation of the level of psychological stress (Older & Jenney, 1975).

The review of previous studies shows clearly that modifications to the speech signal are apparent. The purpose of our work is to evaluate quantitatively the spectral fluctuations of the vowel signal as a result of stress induced by flight incidents aboard a plane. We studied the transfer function of the vocal tract by way of an auto-regressive model (Ruiz, 1991).

The magnetic speech recording from the CVR of a crashed plane is the data used in our study. Its quality is very good: no damping by mechanical or chemical constraints have been noted, the magnetic tape was new, the signal to noise ratio is of 30 dB and the pass band of the CVR is 150-5000 Hz.

3.1. *The experiments*

We studied the CVR speech record of the two pilots. The timing of the events during the last thirty minutes shows the occurence of two successive failures in flight commands. These are followed by discussions between the pilots about the reality and causes of the events. Finally, after a third occurence, the plane crashed.

It is obvious that this record contains a stressful phase (at the end of the flight) and several other phases with lower stress (during discussions after the occurence of technical problems).

3.1.1. *The analysis of vowels*

The sentences are numerous. We kept those which occur just after incidents. We also kept those which can show, from a syntactic point of view, an emotional disturbance. On the other hand, we selected some other sentences during the periods where there was no obvious trouble.

From these sentences, we kept about 300 steady-state vowels for each pilot.

Data analysis was conducted with the Interactive Laboratory System (ILS) software. The monophtongs are analysed during their stable period (i.e. frames for which the frequency of the first three formants is not modified). This signal, low-pass filtered at 8 kHz and sampled at 16 kHz, was divided into analysis frames of 256 samples, each with an overlapping of 192 samples. A hamming window was used and the signal was pre-emphasized (pre-emphasis factor $\mu = 0.98$). The linear prediction model of each frame was calculated by ILS (order 20) which uses the autocorrelation method for calculating the model parameters of every frame.

With specific routines, these coefficients were averaged according to the process described in 2.2. The mean spectral envelope is defined by 512 discrete frequencies between 0 and 8000 Hz (FFT order = 10). We define the two frequency bands: one below the second anti-resonance (low frequency band) and the other from this anti-resonance to 5000 Hz (high frequency band) (cf 4, discussion). Then we calculate a cumulative histogram for these two bands and for the 0-5000 Hz band and we plot the Cumulative Spectral Probability Diagrams.

3.2. *The experimental results*

They are performed with the steady-state vowel [a] which is the most common in the selected sentences (67 for the left pilot and 66 for the right pilot).

3.2.1. *The spectral balance criterion*

The balance criterion is calculated for the two pilots from the mean spectrum of the vowels uttered during non-emotional periods. The first three mean formant frequencies are 750 Hz, 1500 Hz and 2500 Hz. The lower band contains the first two formants up to 1850 Hz and the upper band goes from 1850 Hz up to 5 kHz. The frequency 1850 Hz is set to be the FFT bin the nearest of the antiresonance.

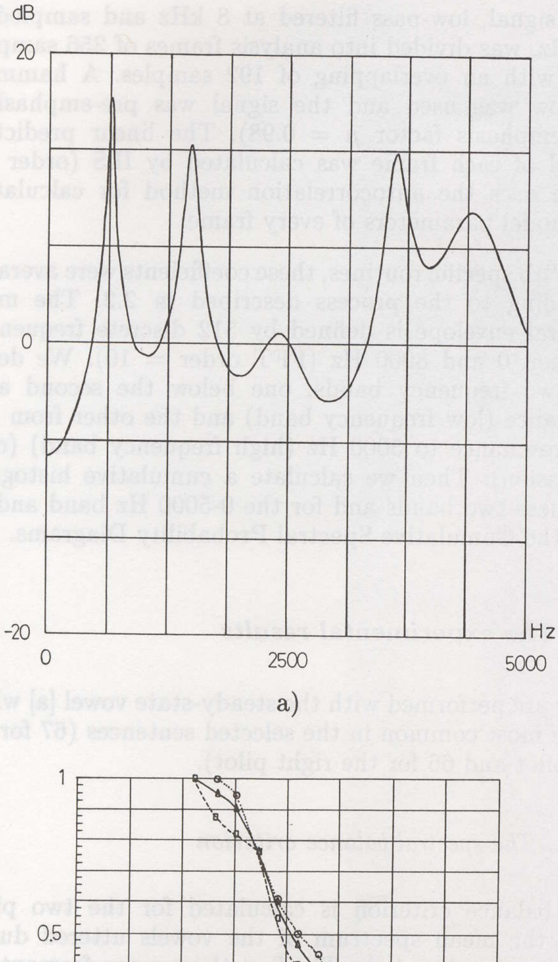
The I_0 index is related to the level L_1 of the first formant both in the low frequency band and in the whole band; but also to the third formant in the high band (L_3). So, L_1 is the highest level in the spectrum and in the low band. The amplitude differences between the first and the third formants ($L_1 - L_3$) are from 15 to 30 dB. Their frequencies are separated by about two octaves.

This general result has been verified during the non-emotional phases of the study, that is to say during the time just preceeding the occurence of incidents.

When the signal is pre-emphasized (6 dB/octave), this difference is less (3 to 18 dB). So differences less than 3 dB are unusual and those which are greater than 3 dB can be considered as usual.

The spectral balance criterion is as follows: when the distance between the low frequency CSPD and the high frequency CSPD is less than 3 dB the spectrum is balanced (see Figure 7b).

If not, the spectrum is unbalanced (see Figures 8a and 8b).



a)

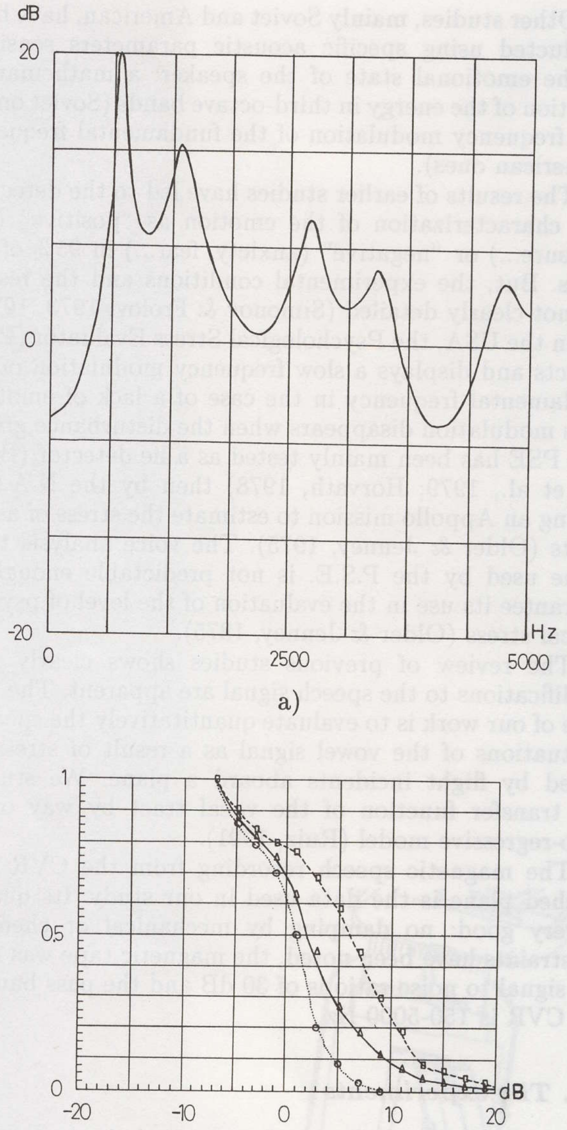
b)

Figure 7. a) An example of a balanced spectrum (steady-state vowel [a], male speaker). b) Cumulative spectral probability diagrams of the Figure 7a spectrum: (—) band (0, 5000)Hz; (– – –) band of the two first formants (0, 1875)Hz; (- - - -) band (1875, 5000)Hz.

3.2.2. *The application of spectral balance criterion*

The steady-state vowels extracted from the pilots' speech are gathered into two parts: those which present a balanced spectrum and the others. Such a classification leads to 45% of balanced spectrum for the left pilot and 25% for the right pilot. These vowels are all uttered during stressfull situations. A balanced spectrum never belongs to quiet periods of flight.

To explain the differences between the two pilots, it can be noted that the left pilot is in charge of the plane. He notes the incidents, but he is not the chief of the crew. The right pilot is more quiet, and tries to explain the origin of the failures.



a)

b)

Figure 8. a) An example of an unbalanced spectrum (steady-state vowel [a], male speaker). b) Cumulative spectral probability diagrams of the Figure 8a spectrum: (—) band (0, 5000)Hz; (– – –) band of the two first formants (0, 1875)Hz; (- - - - - -) band (1875, 5000)Hz.

The spectral balance seems to be a good indicator of the emotional disorders induced by the occurence of incidents during the flight. It is always expressed by an increase of high-frequency energy while the low-frequency energy never leads to spectral balance.

4. Discussion

- The first part of this discussion concerns the choice of the sound level class width ΔL and the frequency resolution Δf for the calculation and the plot of the CSPD.

ΔL and Δf must be of the same size and less than 1 Hz and 1 dB for an acceptable bandwith estimation (error less than 2.5 Hz (Ruiz & Legros, 1994)). In such conditions, the CSPD fits better the cumulative distribution of sound level.

For the experimental application, we have chosen $\Delta f = 15.625$ Hz and $\Delta L = 2$ dB. For Δf, the choice is due to the maximal possible order of the Fast Fourier Transform on a 16 bits computer (order 10 with a 0-8000 Hz spectrum). This leads to an imprecise bandwith estimation. But here, the CSPD have not been intended to do this: only comparative measurements of complementary CSPD have been done. Thus, the accuracy of the plots is less important.

With a 32 bits computer the frequency resolution Δf can be of 1.2 Hz (order 13 of FFT with a 0-5000 Hz spectrum). Precise calculation and plot of CSPD is therefore possible.

Although the experimental choices are not completely satisfactory for bandwith estimation, the conclusions about the spectral balance criterion remain viable because they are based on CSPD comparison.

- The division of spectra into two complementary bands comes from the formant analysis of the monophtongs [a] uttered during the flight. The first two formants are not modified by the emotional disturbances but it appears that high-order resonances are correlated with stress occurence (Ruiz, 1991). This is why we have defined a first frequency band including the first two resonances and a complementary one up to the upper limit of the spectrum. The logical limit between them is the second anti-resonance frequency.

Considering the steady-state vowel [a] of the pilots of the experimentation, we have measured the mean frequency of this anti-resonance: it is 1875 Hz. For other speakers or another vowel this value can change.

Two experimental procedures can be tested depending on the manner to use the second anti-resonance frequency to evaluate the spectral balance.

Firstly, the second anti-resonance frequency is constant. It is the one measured from the mean spectrum of the monophtongs uttered in a rest state by the speaker considered. For every new utterance, this frequency allows us to separate the low and the high frequency bands in the spectrum and to plot the corresponding CSPD for studying the spectral balance.

Secondly, an estimation of the second anti-resonance frequency can be done for every utterance of the vowel. Its value is therefore variable.

In our experimental application of the CSPD we have chosen the first procedure. Indeed, we were interested in defining a neutral vocal emotional state and then comparing it with utterances in stress situations.

- The spectral envelope has a zero mean; then, the level of the minima goes to below zero dB and sometimes to positive values. The level class of $\hat{I}_x = 0.5$ is that of 0 dB in the CSPD calculated on $(0, f_e/2)$.

When the CSPD is calculated on a frequency band of the whole spectrum, the $\hat{I}_x = 0.5$ index indicates the mean sound level in that band. Indeed, it is the level reached or overrun for 50% of the discrete frequencies of the band.

The spectrum of the linear prediction model overvalue the sound level of "valleys" (Markel & Gray, 1982). Therefore, the slope of the CSPD should be less than those actually obtained and the CSPD may include anti-resonances between $\hat{I}_x = 0$ and $\hat{I}_x = 0.5$.

The non-uniform spectral weighting of the linear prediction method (Markel & Gray, 1982) leads to a more accurate CSPD for the low indexes (near zero) corresponding to the higher sound levels than for the indexes near one corresponding to the "valleys" of the spectrum.

Let us consider a resonance of the envelope spectrum of a steady-state vowel. The sound level from which the estimation of the envelope by the LPC method becomes less accurate is unknown. Therefore, the study of spectral balance is limited between $\hat{I}_x = 0$ and $\hat{I}_x = 0.5$, i.e. from the maximal to the mean sound level.

- The advantages of computing the CSPD in frequency bands with regard to measures of energy in the same bands are the followings:

The spectral balance is evaluated from a zero mean spectrum. Then energy computations are not correct because sound level has negative values.

Spectral envelope changes may be due to anatomical modifications of the vocal tract or energy variations of the vowel signal or the combined action of the two. The LPC analysis method allows only working with the transfer function of the vocal tract and of course only examining its modifications without the influence of the sound power. Moreover, we have noticed that the sound level fluctuations are too numerous and correlated poorly with the emotional stress (Ruiz, 1991). Because the CSPD is obtained from the "normalized" spectrum $D(\omega)$ where the mean sound level is not present $(20\log_{10}\sigma)$, the CSPD is an acoustical parameter of the influence of the vocal tract on the speech signal.

The study of spectral balance gives the manner the sound level is distributed. It is easier to find two different spectra which have the same energy in lower and upper bands than the same CSPD. Two utterances may be characterized by their spectral balance obtained from their CSPD, but not completely by their energy in the two complementary frequency bands.

The comparison of CSPD in lower and upper bands does not provide the same information about the presence or absence of any spectrum modifications (due to stress for example) as a staightforward comparison of energy.

5. Conclusion

The estimation of an emotional disorder by way of speech analysis has been studied as early as the 1940's, but remained little studied for half a century. The topic of these investigations is again current because of the increasing developement of voice recognition systems whose error rate increases when the speaker is in an emotional state. So, it is necessary to know the modifications of voice production under stressful conditions to take them into account in the systems.

The spectral balance criterion, obtained by the analysis by frequency bands of the Cumulative Spectral Probability Diagram (CSPD), gives promising results which must be confirmed by a greater statistical study.

Today, the only experimental application is the investigation of voice changes under stress conditions. But the CSPD can be applied to others causes of vowel variability.

Furthermore, in order to calculate and plot the CSPD, it is not necessary to use an LPC analysis method to obtain a zero mean spectrum of the steady-state vowel signal. Whatever the method used to obtain a discrete spectrum, it is important to normalize it such as to give it a zero mean. Then, the study of the spectral balance becomes independent of the spectral analysis method.

Further research linked with the application of CSPD can consist in investigating a spectral balance frequency. It is the one for which the balance is perfect, i.e the euclidian distance between the two CSPD high and low frequency it defines is null or near zero. Such a spectrum corresponds to maximal degree of balance and the state of rest spectrum is the minimal one. So, it would be interesting to have a quantitative scale of spectral balance in relation to degrees of stress. It is the object of our actual research: we attempt to correlate acoustical modifications of the voice with variations of physiological measurements.

At present, the study of the balance of sound levels in the spectrum shows unexpected results related with the psycho-physiological state of the speaker. The study is carried on in such a way as to extend to other situations the acoustical results correlated with the emotional state.

Acknowledgements

This research was supported by the Services Techniques de la Navigation Aérienne, France.

References

ARMSTRONG J. W. (1980) "The effects of concurrent motor tasking on performance of a voice recognition system", Thesis, Naval Postgraduate School, Monterey CA., USA

ARMSTRONG J. W., POOCK G. K. (1981) "Effect of operator mental loading on voice recognition system performance", Technical Report NPS 55-81-016, Naval Postgraduate School, Monterey CA., USA

ATAL B. S., HANAUER S. L. (1971) "Speech analysis and synthesis by linear prediction of the speech wave", J. Acoust. Soc. Am., 50, 637-655

BRENNER M., BRANSCOMB H. H., SCHWARTZ G. E. (1979) "Psychological Stress Evaluator, two tests of a vocal measure", Psychophysiology., 16, 351-357

GRIFFIN G. R., WILLIAMS C. E. (1987) "The effects of different levels of task complexity on three vocal measures", Aviat. Space Environ. Med., 58, 1165-1170

HECKER M. H. L., STEVENS K. N., VON BISMARK G., WILLIAMS C. E. (1968) "Manifestations of task-induced stress in the acoustic speech signal", J. Acoust. Soc. Am., 44, 993-1001

HORVATH F. (1978) "An experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception", J. Appl. Psychol., 63, 338-344

LEGROS C., RUIZ R. (1988) "Etude de l'influence d'une charge de travail sur les caractéristiques acoustiques de la voix: 1. bibliographie", rapport CNES Toulouse, France

LEGROS C., RUIZ R. (1990) "Etude de l'influence d'une charge de travail sur les caractéristiques acoustiques de la voix: 2. expérimentations en simulateur de vol", rapport CNES Toulouse, France

LIEBERMAN P. (1961) "Perturbations in vocal pitch", J. Acoust. Soc. Am., 33, 597-603

MAKHOUL J. (1973) "Spectral analysis of speech by linear prediction", IEEE Trans. Audio Electroacoustics, AU-21, 140-148

MAKHOUL J. (1975a) "Linear prediction: a tutorial review", Proc. IEEE., 63, 561-580

MAKHOUL J. (1975b) "Spectral linear prediction: properties and applications", IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-23, 283-296

MARKEL J. D., GRAY A. H. (1982) "Linear prediction of speech", Springer-Verlag, Berlin

MOSKO J. D., STEVENS K. N., GRIFFIN G. R. (1983) "Interactive voice technology: variations in the vocal utterances of speakers performing a stress-inducing task", Naval Aerospace Medical Research Laboratory Report 1300, Pensacola FA, USA

OLDER H. J., JENNEY L. L. (1975) "Psychological stress measurement through voice output analysis", NASA CR 141723, N75-19960

RUIZ R., LEGROS C., GUELL A. (1990) "Voice analysis to predict the psychological state of a speaker", Aviat. Space Environ. Med., 61, 266-271

RUIZ R. (1991) "Contribution à l'étude de l'influence d'une charge de travail sur les caractéristiques acoustiques de la voix", Thèse de Doctorat de l'Université Paul Sabatier, Toulouse, France

RUIZ R., LEGROS C. (1994) "Estimation de la largeur de bande des formants à partir de la distribution cumulée du niveau sonore dans le spectre", 3ème Congrès Français d'Acoustique, Toulouse, J. Phys. IV France, 4, C5-461-464

SIMONOV P. V., FROLOV M. V. (1973) "Utilization of human voice for estimation of man's emotional stress and state of attention", Aerospace Med., 44, 256-258

SIMONOV P. V., FROLOV M. V. (1977) "Analysis of the human voice as a method of controlling emotional state: achievements and goals", Aviat. Space Environ. Med., 48, 23-25

STREETER L. A., KRAUSS R. M., GELLER V., OLSON C., APPLE W. (1977) "Pitch changes during attempted deception", J. Personality and Social Psychol., 35, 345-350

WILLIAMS C. E., STEVENS K. N. (1969) "On determining the emotional state of pilots during flight: an exploratory study", Aerospace Med., 40, 1369-1372

WILLIAMS C. E., STEVENS K. N. (1972) "Emotions and speech: some acoustical correlates", J. Acoust. Soc. Am., 52, 1238-1250

WILLIAMS C. E., STEVENS K. N. (1981) "Vocal correlates of emotional states", in Speech Evaluation in Psychiatry, J. K. Darby Jr. Ed., Grune & Stratton, Inc.