Proceedings of The Institute of Acoustics

THE WIGNER DISTRIBUTION AS A SPEECH SIGNAL PROCESSING TOOL.

David Lowe, M.J.Tomlinson and R.K. Moore

Speech Research Unit, R.S.R.E, Malvern, Worcs.

1. INTRODUCTION

In the analysis of multicomponent signals with a strong time varying spectrum content, it is desirable to try and describe the distribution of energy in the signal as a simultaneous function of time and frequency. The conventional analysis tool employed in speech research is based upon the sound spectrograph introduced in 1946 [1]. The present day realisation of the original sound spectrograph is performed either by a filter bank analysis of the speech or by a Fourier transform of a windowed version of the time series. Although these two approaches have been shown to be equivalent [2] it is primarily with the second approach that this paper will contrast. Consequently in order to make this distinction the output of a Fourier transform of a windowed version of the signal will be refered to as a *sliding window representation*.

Although this technique has much intuitive appeal, it is nevertheless not the only way of representing a non–stationary signal as a joint distribution of both time and frequency. The history of attempts to produce an *instantaneous spectrum* can certainly be traced back to the end of the nineteenth century [3]. More recently there have been many distinct, proposed joint distribution functions in fields as diverse as radar [4], loudspeaker design [5], speech [6], laser physics [7] and quantum mechanics [8]. The question of what is the *best* way of representing a signal in any given application area has hardly been raised and a detailed analysis comparison is rare.

The aim of this paper is to consider an alternative analysis of speech signals based on the *Wigner distribution function.*

2. HISTORICAL BACKGROUND

The Wigner distribution is an analysis technique which was originally introduced by Dirac [9] and subsequently credited to Eugene Wigner [10] from a paper attempting to describe quantum systems close to thermodynamic equilibrium. It was realised that this method closely paralleled classical techniques and so it found application in the fundamental foundations of quantum theory [11], non–equilibrium quantum statistical mechanics [12], the dynamical transport behaviour of electrons in small semiconductor devices [13] and optical systems theory [14]. In these areas the conjugate variables are position and momentum which are related by the *Heisenberg Uncertainty relation*. This essentially states the impossibility of measuring how fast an object is moving at a specific location arbitrarily accurately.

In signal processing theory there is an analogous relation between the conjugate variables of time and frequency. This information–theoretic inequality proclaims that the bandwidth of a signal of finite duration Δt cannot be less than the spread in frequency Δf determined by the expression $\Delta t \Delta f > \mu$ [15, 16, 17] where the constant μ is of the order of unity (the explicit value depends precisely upon the definitions of signal duration and frequency spread).

Considering this correspondence between signal processing theory and quantum mechanics, it would be hoped that techniques developed in one field may be of some use in the other. Therefore it would appear reasonable to make the Wigner distribution a function of time and frequency instead of position and momentum, and apply the analysis to selected problems involving non–stationary signals. This is precisely what Ville attempted in 1948 [18]. However his efforts received very little attention at that time. Indeed it was not until 1980 in a seminal series of three articles by Claasen and Mecklenbraüker [19] that the technique received a proper signal processing basis being generalised from the continuous case to that of discretely sampled signals. Since that work various authors have realised the potential of the technique for analysing phenomena whose principle frequency components vary rapidly with time. Selected examples of applications in the signal processing area appear in loudspeaker design [20], time-variant filtering theory [21], target ranging [22], biological time keeping activity [23], and a preliminary foray into speech analysis [24].

3. DEFINITIONS AND PROPERTIES

The Wigner distribution may be considered in mathematical terms as a Fourier transform on the off diagonal elements of the autocorrelation matrix of a random process. Specifically for a (possibly non-stationary, complex valued) random process $x(t)$ the autocorrelation matrix is $K(t_1,t_2)=\langle x(t_1)x^*(t_2)\rangle$ where $\langle ...\rangle$ indicates an ensemble average. Then the Wigner distribution of this process is defined as

$$W_X(t,\omega) = \int d\tau \; e^{-j\omega\tau} \; K(t-\tau/2, t+\tau/2) \quad (1)$$

(note that unless otherwise stated, all integrals will be assumed to be over $R = [-\infty,\infty]$).
It is assumed that the autocorrelation matrix has a bi-dimensional Fourier Stieltjes transform and hence we are dealing with harmonisable random processes.

The definition (1) corresponds to the definition in quantum mechanics in terms of the statistical density matrix where the state of the system is given only as a weighted statistical mixture of pure states. The equivalent of a pure state is a deterministic signal which will be dealt with exclusively in this paper. In the case of deterministic signals, the definition (1) simplifies to:-

$$W_X(t,\omega) = \int d\tau \; e^{-j\omega\tau} \; x^*(t-\tau/2)x(t+\tau/2) \quad (2)$$

Contrast the Wigner distribution with the definition of the sliding window representation :-

$$S_X(t,\omega) = \left| \int d\tau \; e^{-j\omega\tau} \; x(\tau)h(t-\tau) \right|^2 \quad (3)$$

where h is the observational time window.

Evidently the properties of the sliding window representation depend crucially upon the choice of a window function and in particular on the duration of this observational window. Since the product of $x(\tau)h(t-\tau)$ is an effective signal of finite duration then we may deduce from the information-theoretic inequality $\Delta t\Delta f \rangle \mu$ that the minimum bandwidth of the windowed signal is restricted. Consequently if one is interested in analysing fine frequency details such as harmonics of the fundamental pitch period in voiced speech then the effective duration of the signal has to be long which implies a very broad observational window. Unfortunately this means that fine temporal details in the original signal will be smeared out due to the weighting process inherent in the definition (3) of the sliding window representation. Alternatively if the aim of the speech analysis is to follow the behaviour of transients such as the onset of the glottal pulse closure or the transitions between different acoustic features, then it is important to restrict the observational window to the region of interest. The effect of this constraint is a short durational signal whose minimum bandwidth is large and hence all fine spectrum details will have been lost. This dichotomy of time-frequency resolution trade-off is perhaps the major disadvantage of the sliding window representation for speech analysis. The Wigner distribution does not have this window function (in the continuous, infinite time ideal case which we consider here for clarity; practical applications of course require the use of sampling and windowing of data discussed shortly) and hence does not suffer from the same problems.

Frequency domain construction.

Sometimes the analysis of the Wigner distribution from the frequency domain is more convenient, particularly when the signal representation is more easily expressed in terms of frequency such as is the case in filter systems. In such a situation, the Wigner distribution may be obtained from the Fourier transform $X(\omega)$ of a signal $x(t)$ as :-

$$W_X(\omega,t) = \frac{1}{2\pi} \int d\zeta \; X^*(\omega-\zeta/2)X(\omega+\zeta/2) \; e^{j\zeta t} \tag{4}$$

from which follows the self dual nature of the Wigner distribution :-

$$W_X(t,\omega) = W_X(\omega,t) \tag{5}$$

Properties.

Further desirable properties which follow from the definitions are

(i) the Wigner distribution is real valued (even for complex valued signals)

(ii) a time shift of the signal corresponds to a time shift of the Wigner distribution

$$W_{x(t-a)}(t,\omega) = W_x(t-a,\omega) \tag{6}$$

(iii) modulating the signal with $\exp(j\Omega t)$ results in a frequency shift

$$W_{x(t)\exp(j\Omega t)}(t,\omega) = W_x(t,\omega-\Omega) \tag{7}$$

(iv) integrating the Wigner distribution over time gives the energy density spectrum of the signal

$$\int dt \; W_X(t,\omega) = \left|S(\omega)\right|^2 \tag{8}$$

(v) integrating the Wigner distribution over frequency yields the instantaneous signal power

$$\int d\omega \; W_X(t,\omega) = 2\pi \left|s(t)\right|^2 \tag{9}$$

(note that the sliding window representation can only ever give an *averaged* version of signal power and energy density)

(vi) invertibility: the signal and its spectrum may be uniquely recovered up to an overall phase factor from the Wigner distribution

$$x(t)2\pi x(0) = \int d\omega \; W_X(t/2,\omega) \; e^{j\omega t} \tag{10}$$

$$X(\omega)X(0) = \int dt \; W_X(\omega/2,t) \; e^{-j\omega t} \tag{11}$$

In addition, for finite energy signals the Wigner distribution is bounded, normalisable and square integrable. Finally, for signal processing applications it is useful to know that:-

(vii) the Wigner distribution of the convolution of two signals is equivalent to the convolution in the time domain of the Wigner distributions of the individual signals

(viii) the Wigner distribution of the product of two signals is equal to the convolution in the frequency domain of the Wigner distributions of the individual signals, and

(ix) the Wigner distribution of the sum of two signals is given by the sum of the Wigner distributions of the component signals, plus the Wigner distribution of the cross correlation between the signals.

These properties make the Wigner distribution an interesting mathematical object. However they also open up the possibility of a useful representation of non-stationary signals such as speech.

4. EXPERIMENTAL PROCEDURE

In a *digital* signal processing environment, a transition of definitions has to be made between the continuous and the discretely sampled case. There have been several interpretations of how this transition should be performed [25, 26, 27]. However the most straightforward way for the application in this paper is to employ the intuitively obvious discrete version of equation (2):—

$$W_x(n,\theta) = 2\sum_{-\infty}^{\infty} e^{-j2k\theta}\, x^*(n-k)x(n+k) \qquad (12)$$

or, in terms of the spectrum of the discrete signal

$$W_X(\theta,n) = \frac{1}{\pi}\int_{-\pi}^{\pi} e^{j2n\zeta}\, X^*(\theta-\zeta)X(\theta+\zeta)\, d\zeta \qquad (13)$$

It should be evident from (13) that the effective sample points upon which the Fourier transform is performed are now equivalent to a time spacing of $2\Delta t$ and the Wigner distribution above is periodic in frequency with a period π instead of 2π as in conventional spectrum analysis. Therefore in order to avoid aliasing problems it is necessary to oversample the original signal at twice the Nyquist rate. As long as this requirement of using an oversampled signal is adhered to then most of the properties of the continuous time case carry over to the discrete version.

An alternative approach to using an oversampled real signal is to employ the analytic signal $s(t)= x(t)+jy(t)$ where $x(t)$ is the original signal sampled at the Nyquist rate and $y(t)$ is the Hilbert transform of x. There is a good physically intuitive reason for utilising the analytic signal as opposed to the real signal: a real signal has a spectrum which is symmetric about its origin and hence the expectation value of the frequency is zero whereas the analytic signal is defined to be zero for negative frequencies and hence its expectation value should yield a positive contribution for the frequency. Therefore the analytic signal makes more sense in situations where the concept of an instantaneous frequency proves to be useful (albeit mathematically impossible!). For example consider a simple tone, $x(t)= \cos(\omega_0 t)$. The expectation value of the frequency of this signal is zero due to the cancelling contributions of positive and negative frequency components. However the frequency expectation value of the corresponding analytic signal, $s(t)= \exp(i\omega_0 t)$ is ω_0 as we would intuitively expect.

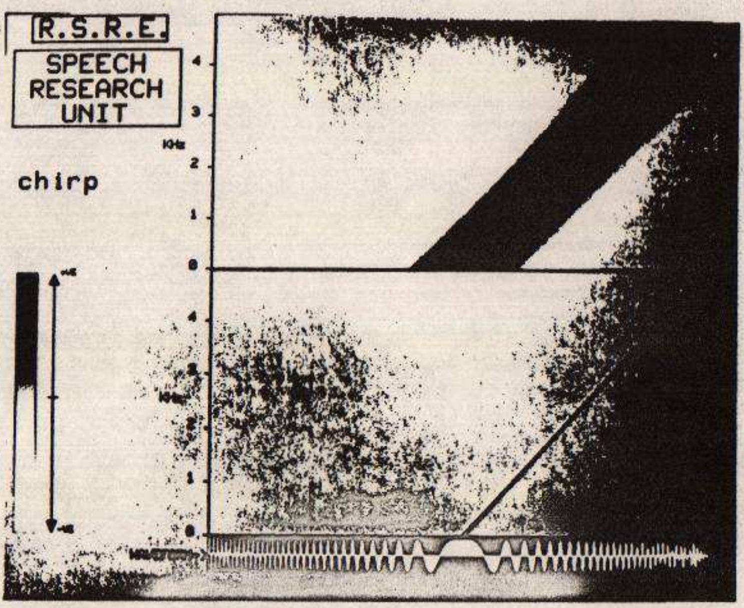
In this paper the real signals are converted into analytic signals which are also oversampled by a factor of two. Consequently the aliasing problems are certainly avoided and spurious correlation effects between symmetric positive and negative frequency components in the signal are neglected. Furthermore the scaling of both the sliding window representation and the Wigner distribution to the same frequency range may be automatically ensured by sampling the analytic signal half as often for the sliding window case as the Wigner distribution case. Finally a comment on windowing of the data. In theory the Wigner distribution does not have an explicit window function. In practice with the constraints of the discrete Fourier transform and memory limitations the Wigner distribution is only calculated using a finite number of data points (511 per time slice). Therefore for each point in time, n, we form the Wigner kernel $s^*(n-k)s(n+k)$ for $k \epsilon [-255,255]$, multiply by an optional (Hamming) window function $w(k)$ (the window should have the property that $w^*(k)=w(-k)$) centred at $k=0$ and apply an FFT algorithm. The practical effect of windowing is to distort the frequency values slightly which is not a problem in our case. However an arbitrarily accurate frequency resolution may be obtained, if desired, by using more points in the FFT routine.

Thus the procedure followed in obtaining an output from a real signal is as follows. Take the original real signal sampled at the Nyquist rate. Convert into an analytic function oversampled by a factor of two. The analytic function is then used as input to both the sliding window routine and the Wigner distribution calculator. The sliding window representation employs a step length twice as large as used by the Wigner distribution, and prior to being Fourier transformed is windowed by a Gaussian shape of user specified bandwidth (the bandwidth is defined such that by half the bandwidth the power window has fallen to 1/e of its maximum value) and cut off at 1% of its peak value.

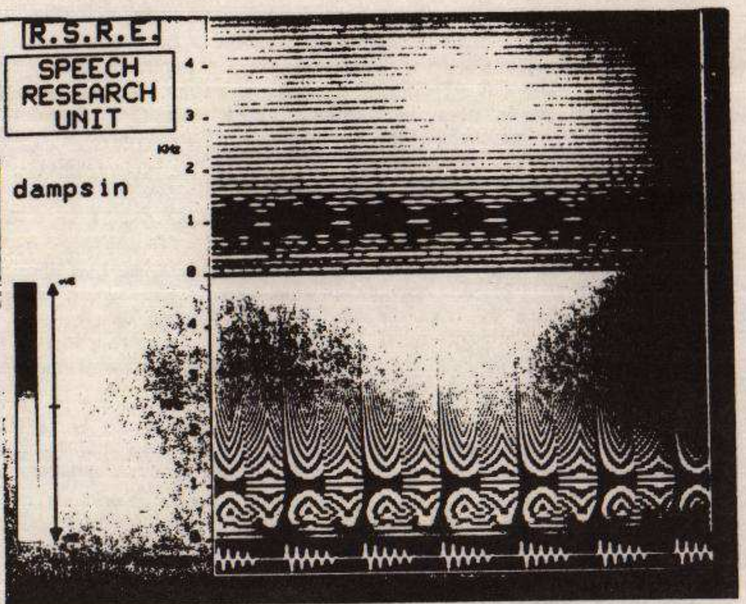
There are 256 frequency values calculated in each time slice, and 512 time slices per picture. The real output values are logarithmically scaled and linearly quantised into 120 levels for positive values and 120 levels for negative values (only relevant for the Wigner distribution) prior to being displayed on a high resolution graphics screen. The results of photographing the screen are displayed in the subsequent pictures.

5. RESULTS

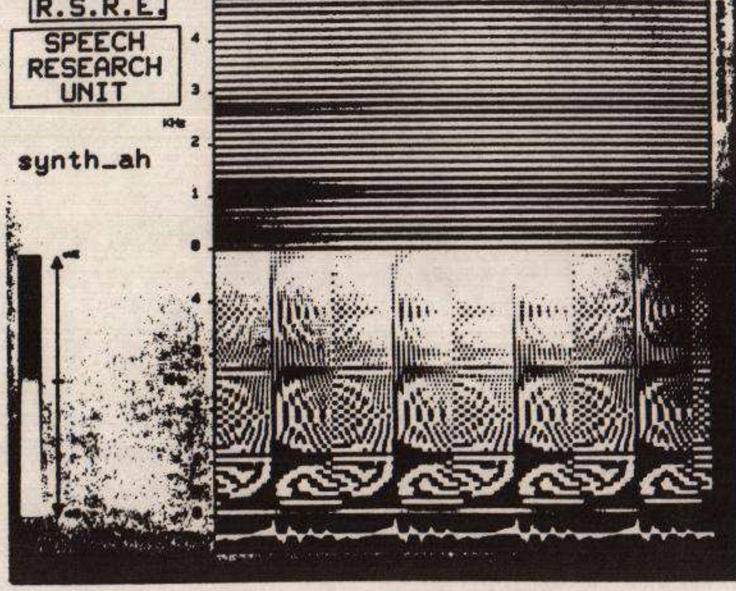
Picture 1.:This depicts a chirp (a signal whose frequency increases linearly with time). The smearing of the fine resolution in the Wigner distribution by the wide band analysis of the sliding window is evident. Clearly the mean squared deviation from the instantaneous frequency is greatest for the sliding window representation.



THE WIGNER DISTRIBUTION AS A SPEECH SIGNAL PROCESSING TOOL.



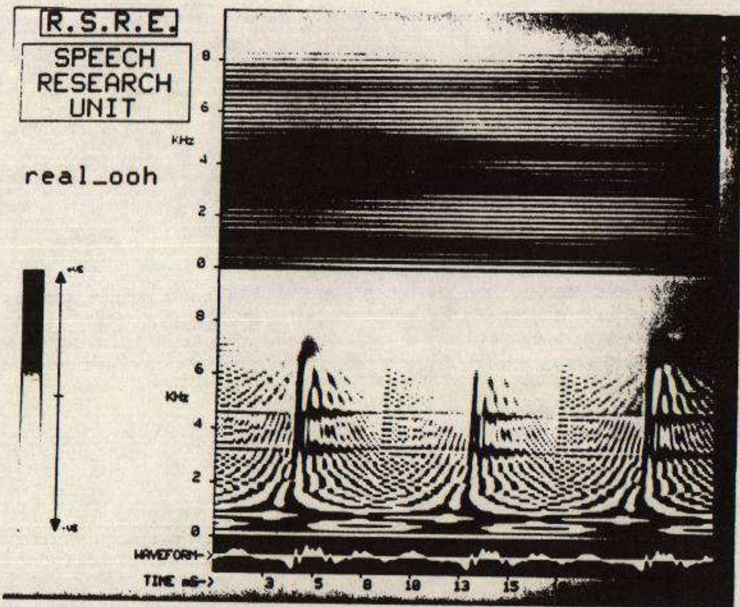
Picture 2.:A sequence of damped sine waves, each having two damping factors, 30 Hz and 90 Hz. The signal is meant to convey the pressure variations induced by the flapping vocal folds in voiced speech. In the top narrow band analysis the harmonics of the fundamental periodicity are evident but there is no indication of temporal structure. The Wigner representation shows the horizontal striations of harmonics *and* vertical striations associated with the onsets of the waves.



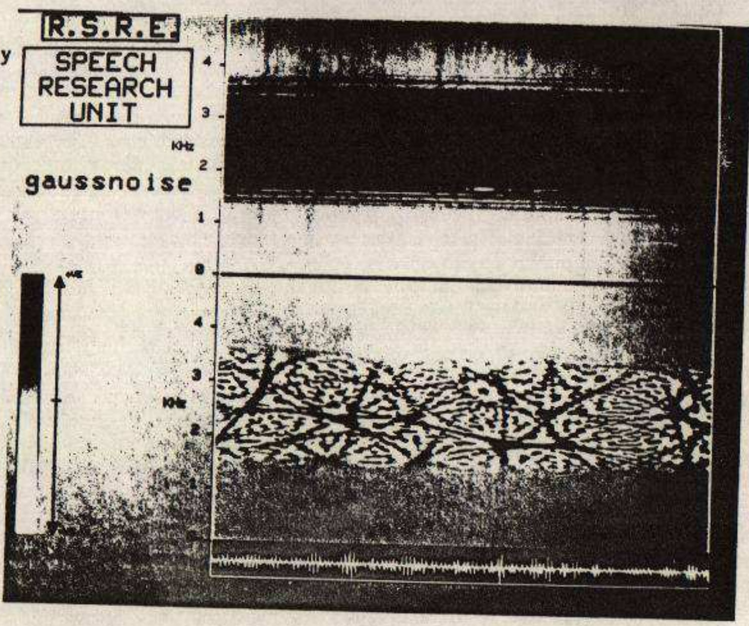
Picture 3.:An example of synthetic speech (an open neutral vowel). The horizontal lines of the harmonics are clear in both cases. The similarity with picture 2. is clear since picture 2 essentially represents a single formant case. Again the temporal structure of the onset of the impulses are evident as vertical lines in the Wigner distribution. The formants are also evident as horizontal structure in both cases.

THE WIGNER DISTRIBUTION AS A SPEECH SIGNAL PROCESSING TOOL.

<u>Picture 4.</u>:A segment of real voiced speech (the /u:/ part of a male saying 'two'). This picture should be compared to 2 and 3. The Wigner picture displays explicitly the impulse associated with the sudden closure of the glottal folds, the decay of the impulse into the first four formants (the formants being the pronounced horizontal bands) and the harmonics of the fundamental frequency.

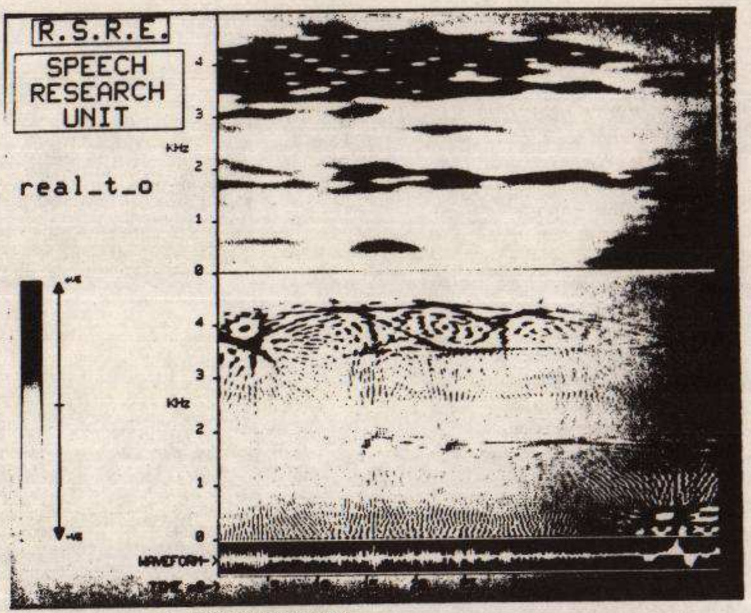


<u>Picture 5.</u>:As a contrastive example, this shows the different time−frequency structure produced by noise. The signal is uniform, band limited noise smoothed with a Gaussian envelope. The narrow band analysis of the top picture has clearly missed much of the structure displayed by the Wigner function. Also apparent is the compactness in frequency of the Wigner representation. This picture should be compared with the example of a periodic signal in picture 2.



THE WIGNER DISTRIBUTION AS A SPEECH SIGNAL PROCESSING TOOL.

Picture 6.:An example of aspiration (the release of the /t/ from a male saying 'to'). How close band-limited noise can model fricative sounds may be gauged by comparing the Wigner distribution patterns of this, and the previous picture. The contrast in patterns between this and the voiced sound,picture 4, should also be noted.



6. CONCLUSIONS

This paper has depicted examples of speech sounds using two distinct time-frequency representations. From an inspection of the outputs of the corresponding representations it is clear that the Wigner distribution has the ability to elucidate simultaneously fine spectrum details and localised time events, in contrast to the sliding window representation. Also there appears to be a larger contrastive difference between acoustically distinct signals using the Wigner distribution which may prove to be of some use in the separation of similar sounding words and speaker identification. Therefore considering the Wigner distribution as an *in-focus* spectrogram shows its potential applications in very high resolution problems such as speech-in-noise, and speaker separation work. Moreover since any smearing decisions have been delayed (presumably to a later stage of data-reduction when it is decided what is relevant in the signal, either physiologically or mechanically), the Wigner distribution is a natural basis for a theory of time-variant filtering.

Its disadvantages are essentially linked with the retention of maximum information in the picture: current recognition algorithms are unable to succesfully handle the data rate supplied by the Wigner distribution. Moreover because the Wigner distribution makes explict correlations in the signal, this leads to a time-frequency structure (ghosting) appearing at twice the periodicity of repeating structure in the original signal. Although this ghosting is annoying (albeit essential for obtaining correct averages), it does not detract from the interpretational power of the Wigner distribution in the time-frequency domain. Finally although the Wigner distribution is as close as one can get to a joint probability distribution of time and frequency without violating the information-theoretic inequality, it cannot be strictly interpreted as a true probability distribution function since in general it can take on negative values. This is not a detriment since the physically relevant features are the 'observables' derived from the distribution (for example, the *perceived* sound of an acoustic element). The Wigner distribution is set up to calculate the 'observables' correctly whereas the sliding window representation can only reproduce the smeared and distorted equivalent of the observables.

As a closing remark we comment that this study has only scratched the surface of what promises to be a fruitful analysis philosophy for speech work. It is hoped to extend this approach and further exploit some of the links between signal processing theory and quantum mechanics to the benefit of speech recognition and synthesis in the near future.

7. REFERENCES

[1] W. Koenig, H.K. Dunn and L.Y. Lacy, 'The Sound Spectrograph', J. Acoust. Soc. Am. Vol. 18, no.1, 19−49, (1946).

[2] J.B. Allen, L.R. Rabiner, 'A unified approach to short time Fourier analysis and synthesis' Proc. I.E.E.E Vol. 65, 1558−1564, (1977).

[3] A. Sommerfeld, Doctoral thesis (1890), in L. Pimonov, 'Vibrations en régimes transitoire', Dunod, Paris. p.68 (1962).

[4] P.M. Woodward, 'Probability and information theory with applications to radar', Pergammon NY, (1953)

[5] J.M. Berman, L.R. Fincham, 'The application of digital techniques to the measurement of loudspeakers', J. Audio Eng. Soc. Vol. 25, 370−384, (1977).

[6] A.V. Oppenheim, 'Speech spectrograms using the fast Fourier transform', I.E.E.E. Spectrum, 57−62, (1970).

[7] R.F. O'Connell, 'Distribution functions in quantum optics', 3rd New Zealand Symposium on Laser Physics, University Waikato, Hamilton, New Zealand, (1983).

[8] V.I. Tatarskiĭ, 'The Wigner representation of quantum mechanics', Usp.Fiz. Nauk. Vol. 139, 587−619, (April 1983).

[9] P.A.M. Dirac, 'Note on exchange phenomena in the Thomas atom', Proc.Camb.Phil.Soc. Vol. 26, 376−385, (July 1930).

[10] E.P. Wigner, 'On the quantum correction for thermodynamic equilibrium', Phys.Rev. Vol. 40,749−759, (1932).

[11] J.E. Moyal, 'Quantum mechanics as a statistical theory', Proc.Camb.Phil.Soc, Vol. 45, 99−123, (1949).

[12] L.P. Kadanoff, G. Baym, 'Quantum Statistical Mechanics', W.A. Benjamin Inc. (1962).

[13] J.R. Barker, D. Lowe, S. Murray, 'A Wigner function approach to transport and switching in submicron heterostructures', in 'Physics of Sub−Micron Devices', edited by D.K. Ferry and K.Hess, Plenum Press, NY (1982).

[14] M.J. Bastiaans, 'The Wigner distribution applied to optical signals and systems', Optics Commun. Vol. 25, no.1, 26−30, (1978).

[15] D. Gabor, 'Theory of communication', J.Inst.Electr.Eng. Vol. 93, 429−441, (1946).

[16] D.G. Lampard, 'Definitions of ⟨bandwidth⟩ and ⟨time duration⟩ of signals which are connected by an identity', Trans.I.R.E. C.T.−3 USA, vol 4, 286−288, (1956).

[17] K. Küpfmüller, 'Über Einschwingvorgänge in Wellenfiltern' Elektr.Nachr.tech.Dtsch. Vol. 1, no.5, 141−151, (1924).

[18] J. Ville, 'Théories et applications de la notion de signal analytique'. Cables et transmissions 2ᵉA Vol. 1, 61−74, (1948).

[19] T.A.C.M. Claasen, W.F.G. Mecklenbräuker, 'The Wigner distribution−a tool for time frequency signal analysis', Philips J.Research Vol. 35, Part 1:217−250,Part 2:276−300, Part 3:372−389, (1980).

[20] C.P. Janse, A.J.M. Kaizer, 'Time frequency distributions of loudspeakers: the application of the Wigner distribution', J.Audio Eng.Soc. Vol. 31, no.4, 198−223, (1983).

[21] B.E.A. Saleh, N.S. Subotic, 'Time−variant filtering of signals in the mixed time frequency domain', I.E.E.E. trans on Acoustics, Speech and Signal Processing, Vol.ASSP−33, no.6, 1479−1485, (1985).

[22] B.R. Breed, T.E. Posch, 'A range and azimuth estimator based on forming the spatial Wigner distribution', Proc ICASSP 84, San Diego, California, Vol. 3,41B.9.1−41B.9.2, (1984).

[23] W. Martin 'Measuring the degree of non−stationarity by using the Wigner−Ville spectrum', Proc, ICASSP 84, San Diego, California, Vol. 3, 41B.3.1−41B.3.4, (1984).

[24] D. Chester, F.J. Taylor, M. Doyle, 'The Wigner distribution in speech processing applications', J.Franklin Inst. Vol. 318, no.6, 415−430, (1984).

[25] T.A.C.M. Claasen, W.F.G. Mecklenbraüker,'A note on aliasing in discrete time Wigner distributions',IEEE trans.on Acoustics,Speech and Signal Processing, Vol.ASSP−31,1067−1072,(1983).

[26] D.S.K. Chan, 'A non−aliased discrete time Wigner distribution for time frequency signal analysis',Proc. ICASSP 1982, Paris, France, 1333−1336, (1982).

[27] K.H. Brenner, 'A discréte version of the Wigner distribution function', in Proc.EUSIPCO Erlangen, West Germany, Elsevier Science, 307−310, (1983).

Copyright © Controller HMSO, London, 1986