Hearing Research, 5 (1981) 123–145
Elsevier/North-Holland Biomedical Press

THE PHONOCHROME: A COHERENT SPECTRO-TEMPORAL REPRESENTATION OF SOUND

PETER JOHANNESMA, AD AERTSEN, BERT CRANEN * and LEON VAN ERNING

Workgroup on Neurophysics, Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

(Received 9 March 1981; accepted 5 May 1981)

Representation of simple stationary sounds can be given either in the temporal form by display of the waveform as function of time or in the spectral form by intensity and phase as function of frequency. For complex nonstationary sounds, e.g. animal vocalisations and human speech, a combined spectro-temporal representation is more directly associated with auditory perception. The well-known sonogram or dynamic power spectrum has a fixed spectro-temporal resolution and neglects phase relations of different spectral and temporal sound components.

In this paper the complex spectro-temporal intensity density CoSTID) is presented as a coherent spectro-temporal image of a sound, based on the analytic signal representation. The CoSTID allows an arbitrary form of the spectro-temporal resolution and preserves phase relations of different sound components. Since the CoSTID is a complex function of two variables, it leads naturally to the use of colour images for the spectro-temporal representation of sound: the phonochrome. The phonochromes are shown for different technical and natural sounds. Applications of this technique for study of phonation and audition and for biomedical signal processing are indicated.

Key words: sound; spectro-temporal; coherent; phase; colour representation.

INTRODUCTION

Spectro-temporal images of sound have a long tradition: the notation of music notes on a staff are used to display both spectral and temporal structure. The melody of a song, animal or human, should be described in terms of both frequency and time; the same holds for intonation in speech. Simple elements of sound can be described in either of the two domains: in the temporal domain by display of the waveform as function of time, and in the spectral domain by either real and imaginary part or by amplitude and phase as function of frequency. Temporal and spectral representation of a signal are related one-to-one by Fourier transformation: they are mathematically isomorphic. Perceptually, however, the different representations may lead to quite different impressions. For complex nonstationary sounds either temporal or spectral representation is insufficient. An underlying reason may well be that auditory perception is based on both spectral and temporal aspects of sound.

* Present address: Institute of Phonetics, University of Nijmegen, Nijmegen, The Netherlands.

In musical notation intensity is indicated in a separate form. The timbre is not indicated at all, only the fundamental tone is shown; knowledge of the instrument should suggest the timbre. The sonogram or dynamic spectrum does not have these shortcomings. The intensity of each spectral component is shown as it develops in the course of time: both fundamental and higher spectral components, giving rise to pitch and timbre, are presented. The sonogram has been widely used for the study of animal vocalisations [9,27] and human speech [11,20]. Various forms of analog, digital or hybrid spectral analysers produce the combined spectro-temporal representations of sound [2,3]. An example is given in Fig. 1. A 'B-call' of a male grass frog [3,14] is shown. At the left-hand side the vocalisation is shown on three different time scales; at the right-hand side the dynamic spectrum is displayed in a quasi three-dimensional form and in a grey coding.

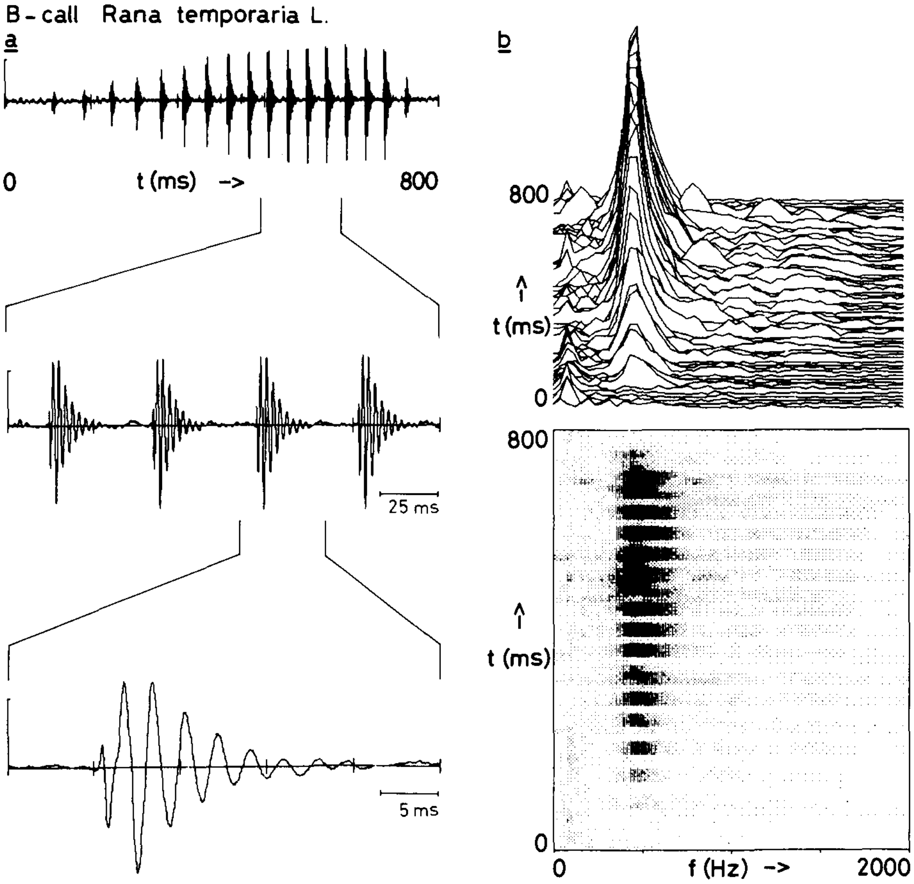

Fig. 1. B-call of male grass frog (*Rana temporaria* L.). (a) The signal as function of time presented on different time scales. (b) Dynamic spectrum or sonogram indicating distribution of amplitude as function of frequency (horizontal axis) and time (vertical axis). Both a three-dimensional display and a grey coding of the sonogram are shown. (From Aertsen et al. [3].)

Some of the sonographs have the advantage of real time operation; all, however, have two fundamental drawbacks. The filters analysing the spectral content of the signal are selected a priori with a given center frequency and a fixed spectral resolution. The spectral resolution again induces a limitation of the temporal resolution; the product of these two resolutions has a lower limit for any physical measuring device [13]. A second limitation of a spectral analyser is the independence of the different filters. Each filter measures only the intensity of the spectral components within its own spectral sensitivity region; as a consequence coherence of different spectro-temporal components will be lost. Since phase relations of spectral components in different frequency regions are neglected the spectral analyser may fail to show, for example, the difference of Gaussian and Poisson types of sounds.

An extension of the sonogram is given in the form of the coherent spectro-temporal intensity density (CoSTID) representation of sound [17,18,21]. While the CoSTID, in a strict sense, can no longer be constructed in real time, it has the advantage that no a priori choice of the spectro-temporal sensitivity regions has to be made; these regions can be chosen a posteriori. Moreover, the CoSTID does preserve spectro-temporal phase relations. In contrast with the real and positive valued sonogram it is a complex function of frequency and time. As a consequence colour enters in a natural way for the representation of the CoSTID.

In this paper a general theory and some examples will be presented. The theory is based on the analytic signal [13,24]. It will be shown that commonly used signal characteristics can be expressed as nonlinear functionals of the analytic signal. An alternative approach is based on a second-order description giving rise to the CoSTID. Signal characteristics are again derived but now as linear functionals of the CoSTID. The CoSTID will be visualized by two different types of chromatic images: one representing the signal characteristics by the local colour contrast, the other giving these characteristics by global aspects of the image.

THEORY

Analytic signal and signal characteristics

The coherent spectro-temporal intensity density (CoSTID) of a sound or a signal is based on the notion of complementarity of time and frequency [13,24]. The mathematical formulation makes use of the analytic signal $\xi(t)$ associated with a given signal $x(t)$. It is defined as

$$\xi(t) = x(t) + i\tilde{x}(t) \tag{1}$$

where $i = \sqrt{-1}$. The quadrature signal or Hilbert transform $\tilde{x}(t)$ of $x(t)$ is given by

$$\tilde{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \mathrm{d}s \frac{x(s)}{t-s} \tag{2}$$

where the integral has to be taken as the Cauchy principal value. The Hilbert transform can be considered as the convolution of a signal with the kernel $(\pi t)^{-1}$. Apart from a minus sign, the quadrature of the quadrature signal is the original signal:

$$\tilde{\tilde{x}}(t) = -x(t) \tag{3}$$

The analytic signal $\xi(t)$ associated with the real signal $x(t)$ is a complex signal with real and imaginary parts related by Hilbert transformation. For a harmonic oscillation $\cos \omega t$ the quadrature signal is $\sin \omega t$ and the analytic signal equals $\exp i\omega t$. The Fourier transform $\hat{\xi}(\omega)$ of the analytic signal $\xi(t)$ equals zero for negative frequencies; for positive frequencies it has the same form as the spectrum of the real signal:

$$\begin{aligned} \hat{\xi}(\omega) = \int \mathrm{d}t\, \mathrm{e}^{-i\omega t}\, \xi(t) &= 2\hat{x}(\omega) \qquad && \omega > 0 \\ &= 0 && \omega < 0 \end{aligned} \tag{4}$$

The relations given in Eqns. 1–4 are illustrated in Fig. 2 for an amplitude-modulated harmonic oscillation [3].

For an arbitrary sound or signal a number of characteristic parameters can be computed in both the temporal and the spectral domains: intensity, amplitude, amplitude modulation, phase and phase modulation or frequency. Since the analytic signal is complex it can be represented in the form

$$\xi(t) = A(t) \exp i\, b(t) = \exp\{a(t) + i\, b(t)\} \tag{5}$$

and in the spectral domain

$$\hat{\xi}(\omega) = \mathcal{A}(\omega) \exp - i\beta(\omega) = \exp\{\alpha(\omega) - i\beta(\omega)\} \tag{6}$$

The signal parameters a, b, α and β are nonlinear functionals of the analytic signal and its Fourier transform. They are systematically represented in Table I.

Temporal phase modulation $\dot{b}(t)$ is a precise definition of instantaneous frequency. Relative spectral envelope modulation $\alpha'(\omega)$ corresponds with the slope of the logarithmic amplitude characteristic; spectral phase modulation $\beta'(\omega)$ is the slope of the phase characteristic as used in harmonic analysis of linear systems. Combination of Fig. 2 and Table I allows the formulation of algorithms for unequivocal computation of the signal characteristics.

For simple signals, the temporal parameters $(a, \dot{b})$ and the spectral parameters (α, β') vary slowly as functions of time and frequency, respectively. An appropriate sampling of these parameters supplies a satisfactory reduced description of the original signal $x(t)$. However, for complex signals, the parameters may fluctuate as fast as the signal itself. Smoothing of the parameters may give misleading results suggesting, for example, the presence of 600 Hz while tones of 400 and 800 Hz were presented simultaneously.

The coherent spectro-temporal intensity density

The choice between temporal and spectral representation of a signal is difficult to solve. Each domain has its own advantages and drawbacks; moreover, separate represen-

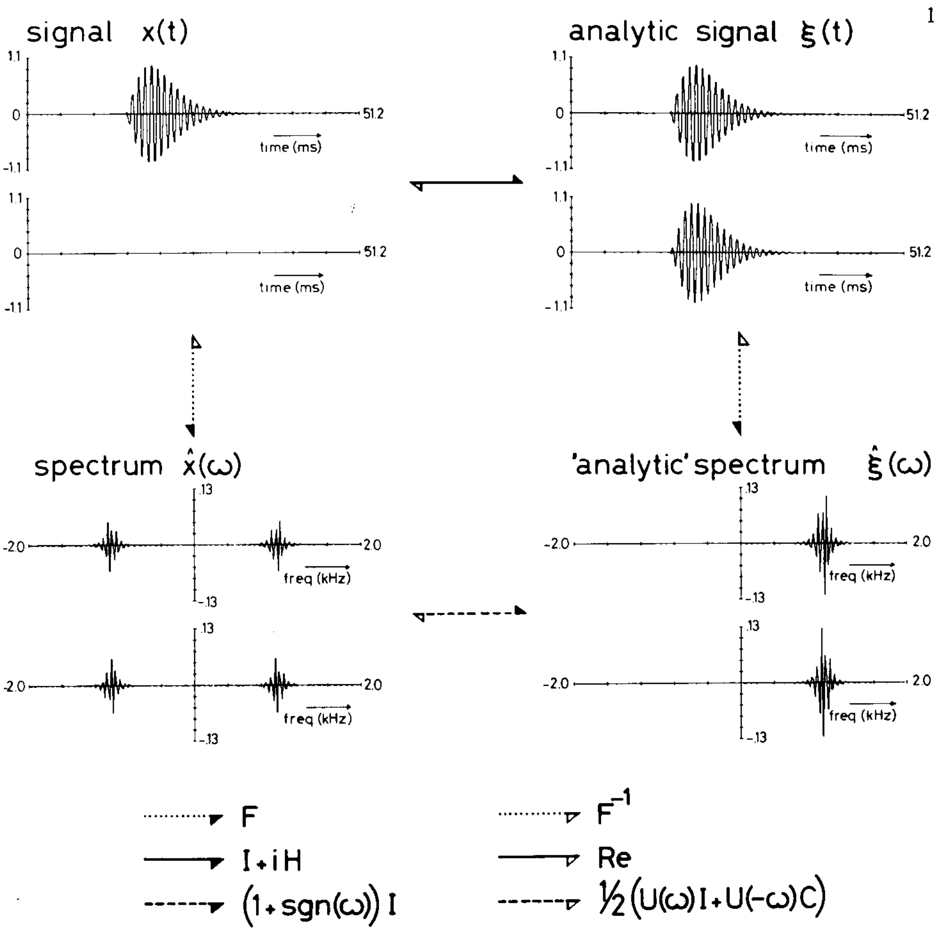

Fig. 2. Signal and analytic signal in temporal and spectral representation. F, Fourier transformation; H, Hilbert transformation; I, identity; Re, real part; $U(\omega)$, unit step function at $\omega = 0$; C, complex conjugation; $\mathrm{sgn}(\omega)$, $U(\omega) - U(-\omega)$.

TABLE I

SIGNAL CHARACTERISTICS IN THE TEMPORAL AND SPECTRAL DOMAIN AND THEIR RELATIONS WITH THE ANALYTICAL SIGNAL

Symbol	Definition	Description	Definition	Symbol
$I(t)$	$\xi^*(t)\,\xi(t)$	intensity	$\hat{\xi}^*(\omega)\,\hat{\xi}(\omega)$	$J(\omega)$
$A(t)$	$\lvert\xi(t)\rvert$	envelope	$\lvert\hat{\xi}(\omega)\rvert$	$\mathcal{A}(\omega)$
$a(t)$	$\mathrm{Re}\ln\xi(t)$	logarithm envelope	$\mathrm{Re}\ln\hat{\xi}(\omega)$	$\alpha(\omega)$
$b(t)$	$\mathrm{Im}\ln\xi(t)$	phase	$\mathrm{Im}\ln\hat{\xi}^*(\omega)$	$\beta(\omega)$
$\dot{a}(t)$	$\frac{d}{dt}\mathrm{Re}\ln\xi(t)$	relative envelope modulation	$\frac{d}{d\omega}\mathrm{Re}\ln\hat{\xi}(\omega)$	$\alpha'(\omega)$
$\dot{b}(t)$	$\frac{d}{dt}\mathrm{Im}\ln\xi(t)$	phase modulation	$\frac{d}{d\omega}\mathrm{Im}\ln\hat{\xi}^*(\omega)$	$\beta'(\omega)$

tation of the signal in the two domains impedes a unified impression. An integrated spectro-temporal signal representation is desirable. As such the dynamic spectrum or sonogram supplies a rough approximation: finite and fixed resolution in time and frequency and neglect of the coherent aspects of signal and sound. The CoSTID may supply a synthesis of signal characteristics present in temporal and spectral form. The CoSTID is defined on the base of the analytic signal as

$$\Xi(\omega, t) = \hat{\xi}^*(\omega)\, e^{-i\omega t}\, \xi(t) \tag{7}$$

where * denotes complex conjugation. This function has been introduced in signal theory [21] under the name 'complex energy density'. It has been proposed for application to the auditory system [5,12,17,18].

The expression given by Eqn. 7 supplies a definition of the CoSTID which can be understood in an intuitive way: the product of analytic signal $\xi(t)$ and its complex conjugated spectrum $\hat{\xi}^*(\omega)$ with a demodulation factor exp $-i\omega t$. The CoSTID is related to the lagged product function Π of the analytic signal ξ

$$\Pi(t - \tau, t) = \xi^*(t - \tau) \cdot \xi(t) \tag{8}$$

by Fourier transformation with respect to the time difference τ

$$\Xi(\omega, t) = \int d\tau\, e^{-i\omega\tau}\, \Pi(t - \tau, t) \tag{9}$$

Substitution of Eqn. 8 into Eqn. 9 leads directly to the definition of Ξ given in Eqn. 7. The CoSTID is also simply related to two other second-order signal characteristics: bispectrum and ambiguity function. Fourier transformation of $\Xi(\omega, t)$ with respect to t produces the bispectrum Γ. Fourier transformation of $\Xi(\omega, t)$ with respect to t and inverse Fourier transformation with respect to ω gives the ambiguity function Δ. For a signal consisting of a mixture of an emitted and a reflected signal the ambiguity function may show the temporal and spectral difference of these two signals, indicating in this way distance and velocity of the reflecting object. The ambiguity function has been used

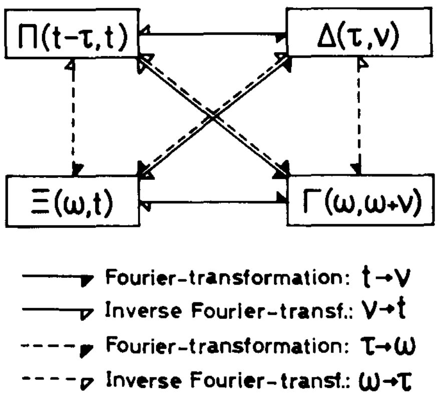

Fig. 3. Diagram of second-order functionals and their relations by Fourier transformation.

in radar and sonar [22,26]. In biology it has been applied to the study of echolocation in bats [4,6,8]. A schematic summary of the relations of these second-order signal functionals is given in Fig. 3.

Properties of the CoSTID

Arguments for the introduction of the CoSTID are the elegant and fundamental mathematical properties as well as the intuitive perceptual value when presented in the form of two-dimensional colour displays. These images, the phonochromes, are shown under Results; mathematical aspects are presented below.

Property I: The CoSTID gives a complete representation of the signal apart from a phase factor

From the definition in Eqn. 7 follows directly that the CoSTIDs of the signals ξ_1 and ξ_2 are equal:

$$\Xi_2(\omega, t) = \Xi_1(\omega, t)$$

if and only if the analytic signals are related by

$$\xi_2(t) = e^{-i\phi}\xi_1(t)$$

which is correct if and only if the real signals are related in the form

$$x_2(t) = \cos\phi \cdot x_1(t) + \sin\phi \cdot \tilde{x}_1(t)$$

where ϕ may assume an arbitrary value. The quadrature signal $\tilde{x}(t)$ differs $\pi/2$ in phase with the original signal $x(t)$. Temporal and spectral envelopes of x_1, $\tilde{x}_1$, x_2 and $\tilde{x}_2$ are identical. While absolute phase of the signal is not retained in the CoSTID, phase relations between different spectro-temporal components are incorporated in the CoSTID.

Property II: Spectral and temporal characteristics of the signal can be derived from the CoSTID by appropriate differentiation

Substitution of Eqns. 5 and 6 in the definition of the CoSTID given in Eqn. 7 leads to

$$\Xi(\omega, t) = \exp[a(t) + \alpha(\omega) + i\{b(t) + \beta(\omega) - \omega t\}] \tag{10}$$

This equation indicates that the CoSTID is directly related to the signal characteristics given in Table I. Differentiation of the logarithm of Eqn. 10 with respect to frequency ω and time t leads to the set of relations

$$\begin{aligned} \dot{a}(t) &= \mathrm{Re}\frac{\partial}{\partial t}\ln\Xi(\omega, t) & \alpha'(\omega) &= \mathrm{Re}\frac{\partial}{\partial\omega}\ln\Xi(\omega, t) \\ \dot{b}(t) &= \mathrm{Im}\frac{\partial}{\partial t}\ln\Xi(\omega, t) + \omega & \beta'(\omega) &= \mathrm{Im}\frac{\partial}{\partial\omega}\ln\Xi(\omega, t) + t \end{aligned} \tag{11}$$

In a compact form this can be written as

$$\{\dot{a} + i\dot{b}, \alpha' + i\beta'\} = \left\{\frac{\partial}{\partial t}, \frac{\partial}{\partial \omega}\right\} \{\ln \Xi(\omega, t) + i\omega t\} \tag{12}$$

which leads to the conclusion that spectral and temporal signal characteristics are in a direct way related to the spectro-temporal gradient of the logarithm of the CoSTID.

Property III: Spectral and temporal characteristics of the signal can be derived from the CoSTID by appropriate integration

Direct integration of Eqn. 7 over frequency or time gives for the intensity density

$$I(t) = \frac{1}{2\pi} \int d\omega\, \Xi(\omega, t) \qquad J(\omega) = \int dt\, \Xi(\omega, t) \tag{13}$$

while the total intensity of the signal is

$$\int dt\, I(t) = \frac{1}{2\pi} \int d\omega\, J(\omega) = \frac{1}{2\pi} \int d\omega \int dt\, \Xi(\omega, t) \tag{14}$$

The CoSTID is considered as the complex distribution of signal intensity in the (ω, t) plane since $I(t)$ is the real and positive valued distribution of intensity with respect to time and $J(\omega)$ the distribution of intensity with respect to frequency. In this way $I(t)$ and $J(\omega)$ are the marginal distributions associated with $\Xi(\omega, t)$. Further signal characteristics are given by the real and imaginary parts of the average value of frequency or time over the CoSTID [1]:

$$\dot{a}(t) = \mathrm{Im} \frac{\frac{1}{2\pi} \int d\omega\, \omega\, \Xi(\omega, t)}{I(t)} \qquad \alpha'(\omega) = \mathrm{Im} \frac{\int dt\, t\, \Xi(\omega, t)}{J(\omega)}$$

$$\dot{b}(t) = \mathrm{Re} \frac{\frac{1}{2\pi} \int d\omega\, \omega\, \Xi(\omega, t)}{I(t)} \qquad \beta'(\omega) = \mathrm{Re} \frac{\int dt\, t\, \Xi(\omega, t)}{J(\omega)} \tag{15}$$

In a compact form this reads

$$\{\dot{a} - i\dot{b}, \alpha' - i\beta'\} = \left\{\frac{-i}{I(t)} \frac{1}{2\pi} \int d\omega\, \omega, \frac{-i}{J(\omega)} \int dt\, t\right\} \Xi(\omega, t) \tag{16}$$

where $\frac{1}{2\pi} \int d\omega\, \omega$ and $\int dt\, t$ are integral operators applied on $\Xi(\omega, t)$.

Comparison of Eqns. 12 and 16 indicates that the signal characteristics as given in Table I can be derived both from the local structure of the CoSTID by differentiation and from the global structure by integration. Some of the redundancy of this spectro-temporal representation of signals becomes apparent here.

Equations for the higher order derivatives of the signal characteristics

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^m \{a(t) - ib(t)\} \qquad \text{and} \qquad \left(\frac{\mathrm{d}}{\mathrm{d}\omega}\right)^n \{\alpha(\omega) - i\beta(\omega)\}$$

can be given in terms of the moments of the CoSTID with respect to frequency or time:

$$\frac{1}{2\pi}\int \mathrm{d}\omega\, \omega^m\, \Xi(\omega, t) \qquad \text{and} \qquad \int \mathrm{d}t\, t^n\, \Xi(\omega, t)$$

If cumulants are chosen instead of moments then the nth order derivative of the signal parameter is related to the nth order cumulant of the CoSTID.

Property IV: The imaginary part of the CoSTID is determined completely by the real part; the converse is not necessarily true

$$\mathrm{Re}\, \Xi(\omega, t) \rightarrow \mathrm{Im}\, \Xi(\omega, t)$$

The proof of this relation can be given in the following way. Making use of Eqn. 13 the temporal intensity $I(t)$ and therefore also the logarithm of the temporal envelope $a(t)$ can be found from the real part of the CoSTID by integration over the frequency ω. In an analogous way Eqn. 15 does supply the instantaneous frequency $\dot{b}(t)$, again making use of the real part of the CoSTID. Returning now to Eqn. 5 it becomes clear that, apart from a phase factor $e^{i\phi}$, the analytic signal can be derived from the real part of the CoSTID. Since the phase does not contribute to the CoSTID (cf. Property I) it follows that the CoSTID as a whole is determined by its real part. As a consequence the imaginary part of the CoSTID can be derived unequivocally from the real part. It should be realised that this property does not hold for a sum or an average value of CoSTIDs; in this case real and imaginary parts may contain independent information.

Property V: Sonograms with different forms of spectro-temporal resolution can be derived from the CoSTID

The sonogram of $x(t)$ can be considered as a set of K functions

$$y_k(t) = \left[\int \mathrm{d}s\, f_k(s)\, x(t - s)\right]^2 + \left[\int \mathrm{d}s \tilde{f}_k(s)\, x(t - s)\right]^2 \qquad k = 1, \ldots K \tag{17}$$

where $f_k(s)$ is the impulse response of filter k [3]. Normally f_k is a narrow band-pass filter with center frequency ω_k and $y_k(t)$ is then the intensity of the signal in the frequency region around ω_k during a time interval preceding t. The spectro-temporal resolution of the sonogram is determined by the choice of the set of filters f_k. For any real-time spectral analyser the product of spectral resolution $\Delta\omega$ and temporal resolution Δt has a lower limit [13]:

$$\Delta\omega \cdot \Delta t \geqslant \tfrac{1}{2} \tag{18}$$

In the process of filtering and squaring, which is inherent in any form of sonogram, information is destroyed. As a consequence sonograms with different choice of filters f_k cannot be transformed into one another.

The sonogram as defined in Eqn. 17 can also be expressed in terms of CoSTID of signal and filter:

$$y_k(t) = \int ds \frac{1}{8\pi} \int d\omega \, \Phi_k(\omega, s) \, \Xi(\omega, t - s) \qquad k = 1, \ldots K \tag{19}$$

where

$$\Phi_k(\omega, s) = \hat{\phi}_k^*(\omega) \, e^{-i\omega s} \, \phi_k(s) \tag{20}$$

and

$$\phi_k(s) = f_k(s) + i\tilde{f}_k(s) \tag{21}$$

The use of inverse Fourier transformation leads directly from Eqn. 19 back to Eqn. 17. From Eqn. 19 it follows that the sonogram is a linear functional of the CoSTID: multiplication in the spectral and convolution in the temporal domain of CoSTIDs of signal and filter. Since an arbitrary sonogram can be derived from the CoSTID, this function forms a generating function or underlying structure for the sonogram. As such the CoSTID is not physically realisable, but arbitrary, realisable spectro-temporal representations can be derived from it. As a consequence no a priori choice has to be made concerning temporal versus spectral resolution. Conservation of coherence is the basic explanation for this generality.

Property VI: For a mixture of signals the CoSTID incorporates interference products of the elementary signals

Given a signal consisting of a sum of elementary signal components

$$\xi(t) = \sum_i \xi_i(t) \tag{22}$$

the CoSTID of this signal then becomes

$$\Xi(\omega, t) = \sum_{i,j} \Xi_{ij}(\omega, t) \tag{23}$$

where

$$\Xi_{ij}(\omega, t) = \hat{\xi}_i^*(\omega) \, e^{-i\omega t} \, \xi_j(t) \tag{24}$$

The contribution Ξ_{ii} originates from the signal ξ_i and does not contain the phase of ξ_i. The contribution $\Xi_{ij}(i \neq j)$ is created by the combination of ξ_i and ξ_j and is dependent on the phase difference of these two elementary signals. In the sonogram the contribution based on $\Xi_{ij}(i \neq j)$ is not present as long as ξ_i and ξ_j are located in nonoverlapping spectro-temporal regions. In the phonochrome representation given under Results the contribution Ξ_{ij} leads to 'ghost images' as can be seen clearly in Fig. 6e, f. Smoothing of the contributions Ξ_{ij} in a manner given by Eqn. 19 does completely eliminate Ξ_{ij} for $i \neq j$ but not for $i = j$.

If the total signal $\xi(t)$ is an element of an ensemble of comparable signals, e.g. frog vocalisations of a certain type, then the average CoSTID for this ensemble of signals may be determined. The phase relations of different signal components influence the appearance of the average CoSTID. If the signal $\xi(t)$ is incoherent, i.e., the phase relations among the $\xi_i(t)$ differ for different realisations, then the terms $\Xi_{ij}(i \neq j)$ producing the virtual images will cancel in the average CoSTID. On the other hand, if the signal $\xi(t)$ is coherent, i.e., the phase relations among the $\xi_i(t)$ are identical in different realisations, then the virtual images caused by $\Xi_{ij}(i \neq j)$ will be present in the average CoSTID. The boundaries of the spectro-temporal coherence region can then be estimated from the values of the intermodulation products averaged over an ensemble of comparable signals or sounds. Segments of animal and human vocalisation or music which are reproducibly coherent form indivisible elements of the phonation. These acoustic elements may be indicated as phonons.

Definition: A phonon is a spectro-temporal component of a homogeneous ensemble of signals or phonations such that within the phonon phase relations are deterministic and between phonons the phase relations vary stochastically. In other words: a phonon is the content of the spectro-temporal coherence region of an ensemble of signals.

From the average CoSTID of an ensemble of signals the phonons can be determined. If the spectro-temporal regions for the phonons forming a given signal are known, then the boundaries $(-\infty, \infty)$ in the definition of integrals leading to the CoSTID can be replaced by the phonon boundaries. Following this approach, Eqn. 22 can be rewritten as a sum of phonons and Eqn. 23 then no longer contains any interference products. As a consequence phonons form the elementary constituents of an ensemble of signals or sounds. In solid-state physics the phonon is defined as a harmonic acoustic oscillation in a crystal. Since phonons interact only to a small extent, the definition given here may be considered as a generalisation of the one used in physics [25]. In statistical signal analysis use is made of the principal components or Karhunen-Loève expansion for the description of an ensemble of signals [19]. The eigenfunctions of the time-dependent autocorrelation function of the signal ensemble form an uncorrelated, orthogonal and complete set of functions optimally suited for representation and approximation of the signals under consideration. The relation of the principal components analysis and signal representation in terms of phonons appears worthy of investigation.

The difference of phonon and phoneme should clearly be realised. A phonon can be determined from a homogeneous ensemble of different realisations of a single phonation. A phoneme is the minimal discriminative element for different phonations. As such, at least two different types of phonations should be involved. For the determination of a phoneme, semantic and pragmatic aspects of the phonation are relevant; this is not the case for the phonon. A phoneme may contain more than one phonon.

The theoretical considerations of this chapter can be summarised as follows. The CoSTID representation of a signal gives a coherent spectro-temporal representation. Differentials and integrals of the CoSTID generate both the well known signal characteristics and the dynamic spectrum or sonogram with arbitrary form of the spectro-temporal resolution. Ensemble averaging of the CoSTID for different realisations of the signal indicates the spectro-temporal coherence regions of the signal. Knowledge of these regions may allow an expansion of the signal or sound in terms of phonons.

RESULTS

Chromatic representation

In this chapter examples are presented of the CoSTID for nine different signals varying from a simple amplitude-modulated tone to an element of the vocalisation of the grass frog. The signals shown in Fig. 4 have been chosen in order to illustrate the concepts and mathematical properties of the previous chapter.

The CoSTID is a function of two variables: frequency and time. Moreover the CoSTID is a complex function; this implies that in each point of the (ω, t)-plane two values have to be presented. These values can be real and imaginary parts or amplitude and phase of the CoSTID. In Fig. 5 these four aspects of the CoSTID are presented for the single gamma-tone given in Fig. 4a by means of grey-coding of the (ω, t)-plane. The horizontal axes of the figures correspond with time, the vertical axes with frequency. A drawback of this representation is the necessity to display two separate pictures for one CoSTID. Perceptually it is difficult to integrate either real and imaginary or amplitude and phase display into one impression. A representation of a complex variable can also be given by means of a vector field defined on the (ω, t)-plane [18]. This procedure, however, leads to a sort of optical congestion at positions where the amplitude of the CoSTID is large; also, the direction of the vector is difficult to represent.

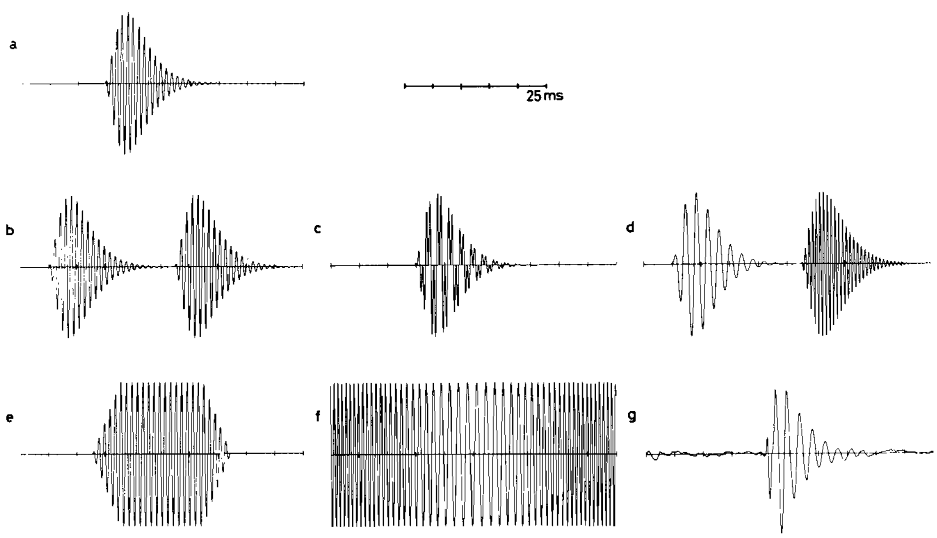

Fig. 4. Examples of different acoustic signals: (a) single γ-tone; (b) sequence of two γ-tones with identical spectral content; (c) two simultaneous γ-tones with different spectral content; (d) sequence of two γ-tones with different spectral content; (e) tone burst with trapezoid envelope; (f) frequency-modulated tone; (g) element of frog vocalisation.

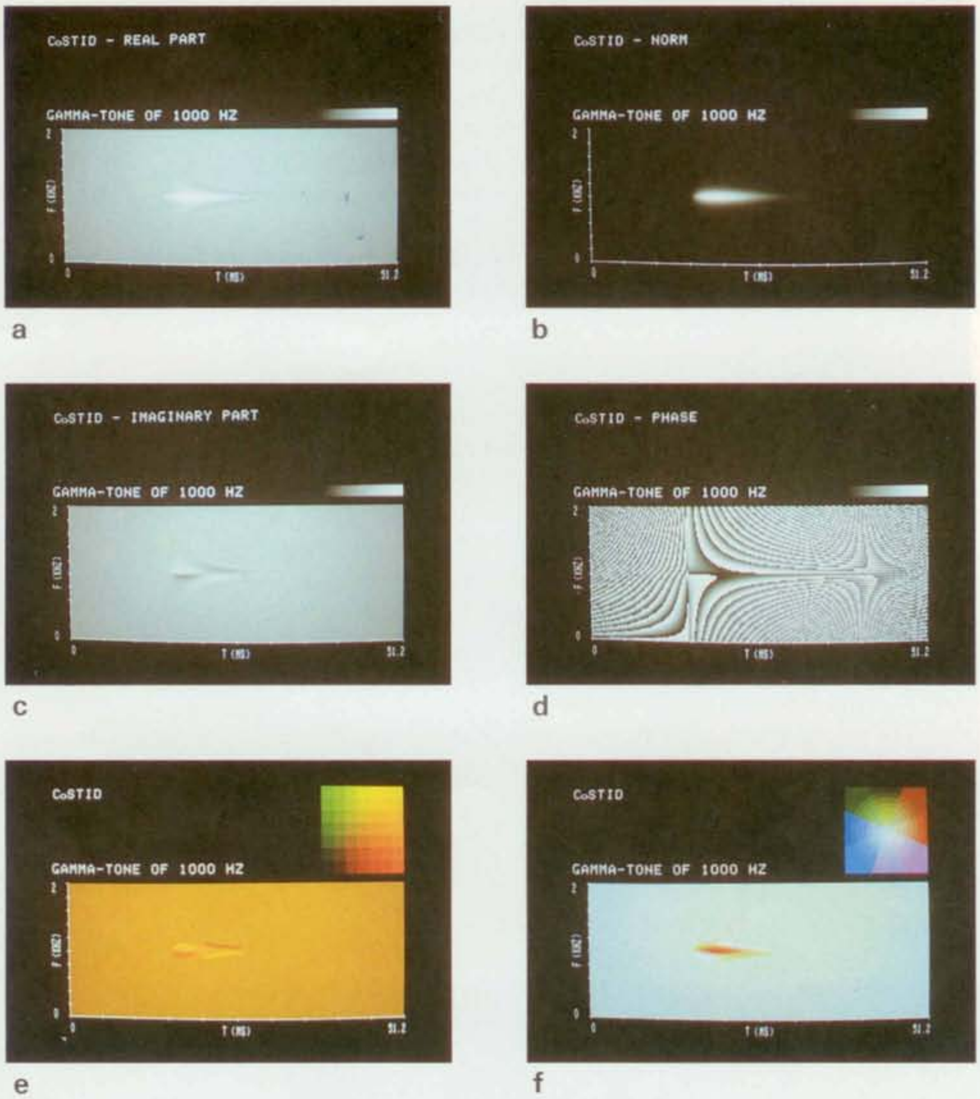

Fig. 5. Different representations of the CoSTID of a single γ-tone as given in Fig. 4a.

A solution of these problems can be found by the introduction of colour. Since the colour or chromatic value of a point is specified by three independent aspects, this gives the possibility to represent three independent variables at each point in the (ω, t)-plane. Two rather straightforward choices can be made for the chromatic variables: the triplet (*Red, Green, Blue*) as well as the triplet (*Intensity, Saturation, Hue*) [28]. For the representation of a complex number only two variables are needed. For the chromatic representation of the CoSTID $\Xi(\omega, t)$ two variables are chosen from the triplet.

The chromatic representation of the CoSTID is then made in the following way. The spectro-temporal plane is mapped onto the image plane by

$$(\omega, t) \rightarrow (y, x)$$

The CoSTID in each spectro-temporal point (ω, t) is transformed into a chromatic value for the associated (y, x) point in the image by either

$$\{\mathrm{Re}\,\Xi(\omega, t), \quad \mathrm{Im}\,\Xi(\omega, t)\} \rightarrow \{R(y,x), \quad G(y,x)\}$$

or

$$\{\mathrm{Norm}\,\Xi(\omega, t), \quad \mathrm{Arg}\,\Xi(\omega, t)\} \rightarrow \{S(y,x), \quad H(y,x)\}$$

where $R(y, x)$ is the intensity of red at the position (y, x), G stands for green, S for saturation and H for hue.

The precise quantitative form of the chromatic map can now be given. For the arguments there exist at least two possibilities

$$y = \omega, \quad x = t \tag{25a}$$

or

$$y = \ln \omega, \quad x = t \tag{25b}$$

Different choices of $\omega \rightarrow y$ may be derived from perceptual resolution of frequency based on measurements of pitch and timbre. For the complex variable some simple choices for the chromatic coding are

$$R - R_0 = \mathrm{Re}\,\Xi \qquad G - G_0 = \mathrm{Im}\,\Xi \tag{26a}$$

$$\ln R/R_0 = \mathrm{Re}\,\Xi \qquad \ln G/G_0 = \mathrm{Im}\,\Xi \tag{26b}$$

$$S = \mathrm{Norm}\,\Xi \qquad H = \mathrm{Arg}\,\Xi \tag{26c}$$

$$\frac{S}{1-S} = \mathrm{Norm}\,\Xi \qquad H = \mathrm{Arg}\,\Xi \tag{26d}$$

Again a comparison of perceptual resolution (j.n.d.) of auditory and visual systems may suggest different or more refined forms of Eqn. 26.

Definition: A PHONOCHROME of a sound or signal x(t) is the chromatic image of the coherent spectro-temporal intensity density $\Xi(\omega, t)$ of the sound based on a mapping as defined by Eqns. 25 and 26.

The results presented in this paper were produced by a computer-controlled video colour display (Ramtek RH-2300, 6 bitplanes; Barco display CDCT 2/66). Frequency and time scale are chosen linear as given in Eqn. 25a. For the complex variable Ξ both the linear form of the (R, G) coding of Eqn. 26a and the linear form of the (S, H) coding of Eqn. 26c are shown. The lower part of Fig. 5 displays these two phonochromes of the amplitude-modulated oscillation given in Fig. 4a. The chromatic code which has been used is included as a separate key in each figure.

Some comments should be given on the choice of the chromatic mapping defined in Eqn. 26. Real and imaginary parts of a complex variable are real variables in the range $(-\infty, \infty)$ while intensity of a colour component is a real variable in the range $(0, \infty)$. This suggests the logarithmic relations of Eqn. 26b. For a finite segment of sound, however, maximal and minimal values of real and imaginary parts of the CoSTID are limited. It is possible to apply the linear (R, G) coding given in Eqn. 26a by a shift of the (R, G) origin from $(0, 0)$ to (R_0, G_0). Saturation S of a colour is a real variable limited to the range $[0, 1]$, $S = 0$ corresponding with completely unsaturated, i.e., white, and $S = 1$ describing the fully saturated colour. Hue H can be defined on $[0, 2\pi]$: it may be considered as a variable defined on a circle; $H(0)$ is identical with $H(2\pi)$. As a consequence, saturation S may be related to norm or amplitude of the CoSTID, while hue H is a proper variable to represent argument or phase. While Eqn. 26d can be used for arbitrary sounds, a finite segment with a maximal value of $|\Xi|$ can be coded by Eqn. 26c. Because of the finite colour resolution of the video display, the intensities of red and green have only seven different levels each; saturation is sectioned into eight values, and hue is divided into eight segments apart from the smallest set of saturation values where differentiation is no longer made with respect to hue.

For the signals presented in Fig. 4b–g the (Red, Green) and (Saturation, Hue) phonochromes are shown in Figs. 6 and 7.

Properties of the phonochrome

The chromatic coding defined by Eqns. 25 and 26 is defined from point to point: one spectro-temporal value of the sound is mapped into one chromatic point of the image. *The goal of the construction of the phonochrome is a simple correspondence of visual qualities of a picture and auditory features of sound.* Distinctive auditory features in animal communication may be related to identification and localisation of the acoustic source: in human communication recognition of spoken words and music is an important aspect. Some perceptual qualities of sound are loudness, pitch, timbre, harmony, rhythm and melody. As a first approximation these features may be related to the characteristic functions $a(t)$, $\alpha(\omega)$, $b(t)$, $\beta(\omega)$ and their derivatives given in Table I.

For colour pictures two main types of features can be indicated: *local* and *global* aspects of the image. It will be shown that for a phonochrome both aspects correspond directly with the characteristic features of an acoustic signal as proposed in Table I. The colour or chromatic value of a single point of the phonochrome can be taken as a position in a colour plane. Its coordinate is given by either $R + i\,G$ or by $S\,e^{iH}$. As a consequence

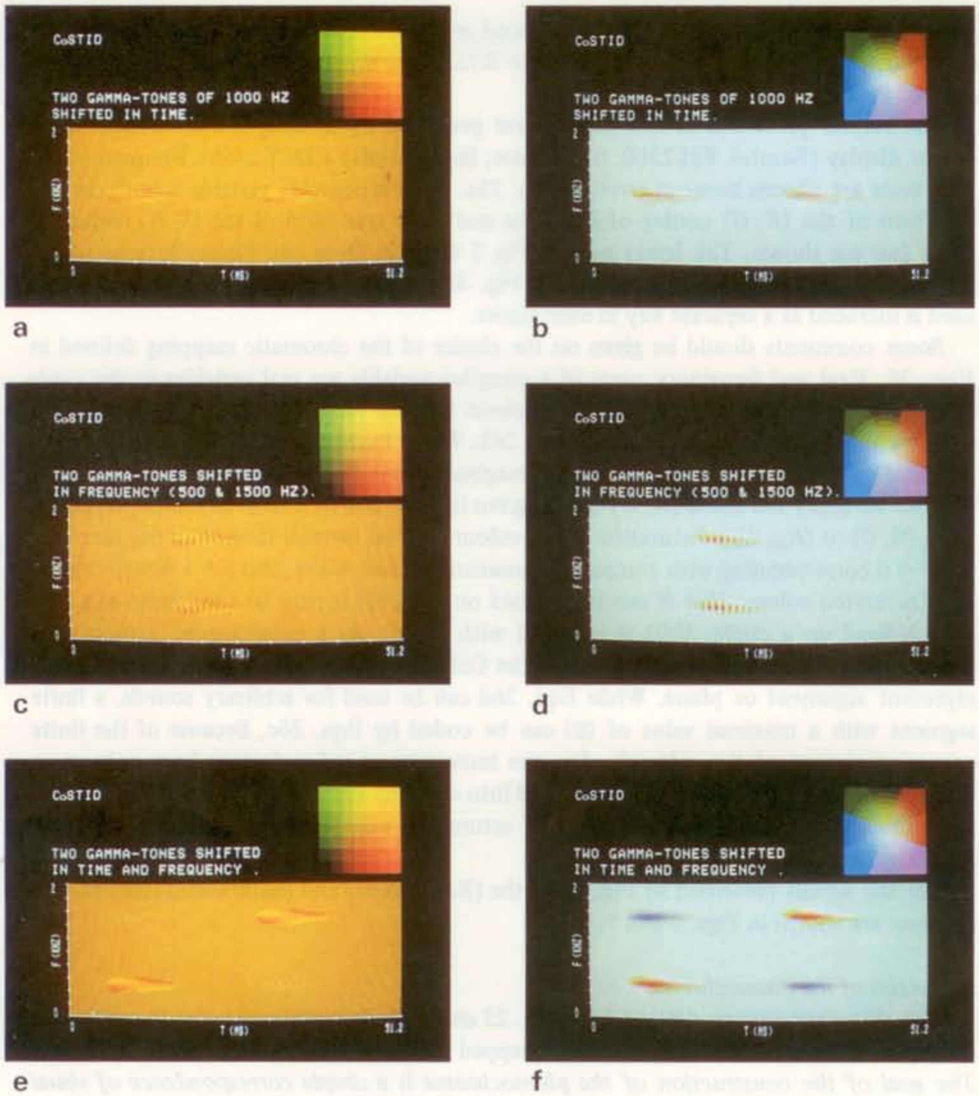

Fig. 6. Phonochromes with two different colour codes of three different combinations of two γ-tones as given in Fig. 4b–d.

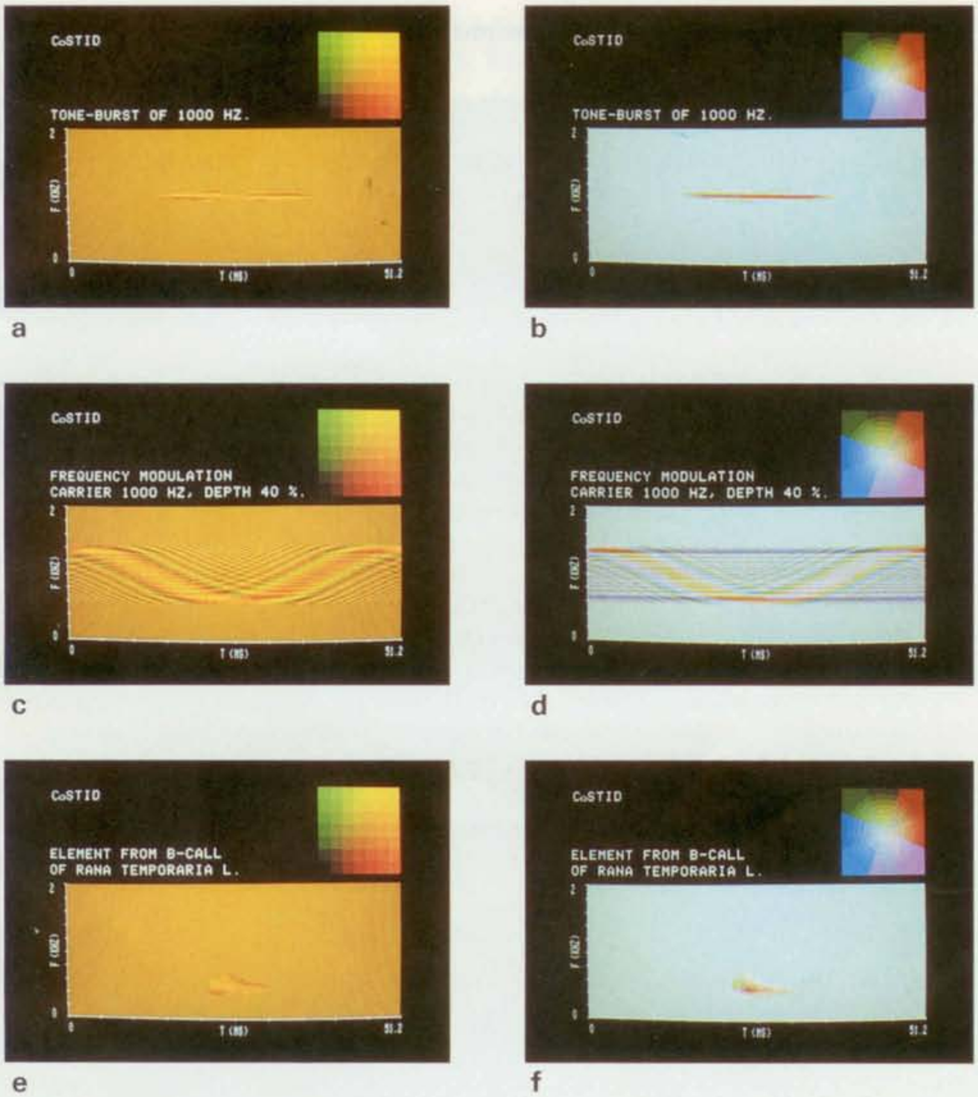

Fig. 7. Phonochromes with two different colour codes of tone burst, frequency-modulated tone and an element of the vocalisation of the grass frog. The signals are given in Fig. 4e–g.

Eqns. 25 and 26 can be combined to a compact notation for the linear (R, G) code:

$$\Xi(\omega, t) \leftrightarrow R(y,x) + i\, G(y,x) \tag{27}$$

where R and G are the amount of red and green with respect to some background level. For the linear (S, H) code the mapping can be written in the form

$$\Xi(\omega, t) \leftrightarrow S(y,x) \exp iH(y,x) \tag{28}$$

These two mappings form the base for the audio-visual relations of characteristic features of sound and images.

Local or differential aspects of a chromatic picture are determined by the relative change in colour: the chromatic contrast given by the gradient of the logarithm of the chromatic variables. The differential properties of the CoSTID as expressed in Eqn. 11 given in connection with Property II, combined with the linear (S, H) code given by Eqn. 28, lead to the relation of acoustic and chromatic variables given in Table II.

Global or integral aspects of a chromatic picture are related to the amount of colour present in certain regions. In Property III it has been shown that integrals over the CoSTID are simply related to the spectro-temporal acoustic characteristics. Use of the linear (R, G) phonochrome as expressed in Eqn. 27 leads to a relation of acoustic parameters and global aspects of its chromatic image. The relations are given in Table III.

The results presented in Tables II and III allow the conclusion: *Spectro-temporal change in relative envelope and in phase of the acoustic signal correspond with contrast in saturation and change in hue in the (*S, H*) phonochrome and with average location of red and green in the (*R, G*) phonochrome.*

Apart from the local and global features described here, colour pictures can also be characterised by their set of isochromes. An isochrome is defined as the set of points

TABLE II

RELATION OF CHARACTERISTICS OF ACOUSTICAL SIGNAL AND LOCAL ASPECTS OF THE (S, H) PHONOCHROME

Acoustical		Optical	
$\dot{a}(t)$	relative temporal envelope modulation	$\frac{\partial}{\partial x} \ln S$	horizontal relative change in saturation
$\alpha'(\omega)$	relative spectral envelope modulation	$\frac{\partial}{\partial y} \ln S$	vertical relative change in saturation
$\dot{b}(t) - \omega$	difference of temporal phase modulation and frequency	$\frac{\partial}{\partial x} H$	horizontal change in hue
$\beta'(\omega) - t$	difference of spectral phase modulation and time	$\frac{\partial}{\partial y} H$	vertical change in hue

TABLE III

RELATION OF CHARACTERISTICS OF ACOUSTICAL SIGNAL AND GLOBAL ASPECTS OF THE (R, G) PHONOCHROME

Acoustical		Optical	
$I(t)$	temporal intensity	$R(x) = \int dy\, R(y,x)$	amount of red on vertical line at $x = t$
$J(\omega)$	spectral intensity	$R(y) = \int dx\, R(y,x)$	amount of red on horizontal line at $y = \omega$
$\dot{a}(t)$	relative temporal envelope modulation	$\dfrac{\int dy\, y\, G(y,x)}{R(x)}$	average location of green on horizontal line at $x = t$
$\alpha'(\omega)$	relative spectral envelope modulation	$\dfrac{\int dx\, x\, G(y,x)}{R(y)}$	average location of green on vertical line at $y = \omega$
$\dot{b}(t)$	temporal phase modulation	$\dfrac{\int dy\, y\, R(y,x)}{R(x)}$	average location of red on horizontal line at $x = t$
$\beta'(\omega)$	spectral phase modulation	$\dfrac{\int dx\, x\, R(y,x)}{R(y)}$	average location of red on vertical line at $y = \omega$

which have the same value of the chromatic variables. Both for the (R, G) and the (S, H) coding an isochrome in the phonochrome is the set of points for which the CoSTID $\Xi(\omega, t)$ has the same value. Taking account of Eqn. 10, the following conclusion can be formulated: *on or along an isochrome of the phonochrome the acoustic features given in Table I obey the relations*

$$a(t) + \alpha(\omega) = \text{constant} \tag{29a}$$

$$b(t) + \beta(\omega) - \omega t = \text{constant} \tag{29b}$$

In experimental practice only the finite number of different colours shown in the colour code are available. As a consequence the isochromes are determined by the condition that the left-hand sides of Eqn. 29 are approximately constant; this leads to the results that for continuous, differentiable signals each part of an isochrome always covers a finite area. Mathematically it can be shown by direct computation that for a tone of constant frequency and Gaussian-modulated envelope the regions of constant saturation given by the condition of Eqn. 29a are ellipses, while the regions of constant hue defined by Eqn. 29b are hyperboles. The isochromes are then formed by the intersection of regions of constant saturation and constant hue. For a γ-tone as given in Fig. 4a these regions have approximately the same form; this can be clearly seen in the grey-display of norm and phase of the CoSTID of this signal as shown in Fig. 5. The iso-norm regions in

Fig. 5b are approximately ellipsoidal, the iso-phase regions in Fig. 5d hyperbolical. The form of the (S, H) phonochrome can be understood from these data.

The two γ-tones shifted in time presented in Fig. 6a, b show periodic spectral modulation induced by the temporal repetition of the signal. In an analogous way the two γ-tones shifted in frequency given in Fig. 6c, d show periodic temporal modulation reflecting the interference or 'beating'. In the case of the low-frequency tone followed by a high-frequency tone given in Fig. 6e, f the combination of these two effects is visible in the cross-terms or 'ghost images' as discussed in connection with Property VI. In this context it should be noted that visual inspection of the (R, G) phonochrome in Fig. 6e with the eyes half closed results in the disappearance of the 'ghost images'. This effect can be understood from the integral properties of the $(R; G)$ phonochrome. The same effect can be observed in Fig. 7c. As is expected, this effect does not occur in the corresponding (S, H) photochromes. The elements of a frog vocalisation shown in Fig. 4g closely resembles a γ-tone. Its phonochrome given in Fig. 7e, f does not differ significantly from that of the γ-tone shown in Fig. 5e, f; only a slight deviation is present in the onset.

DISCUSSION

For the understanding of audition and phonation both temporal and spectral aspects of sound have to be considered. Vocalisation apparatus and cochlea are constructed in such a way that an interplay of spectral and temporal features may occur in the synthesis and analysis of sound. An integrated spectro-temporal representation should be based on second-order characteristics of the signal. There exist four types of these second-order characteristics mutually related through Fourier transformation: product function $\Pi(t - \tau, t)$ CoSTID $\Xi(\omega, t)$, ambiguity function $\Delta(\tau, \nu)$ and bispectrum $\Gamma(\omega, \nu)$. For the characterisation of an ensemble of signals the average product function or autocorrelation matrix can be used to generate uncorrelated orthogonal signal components: the Karhunen-Loève expansion which is closely related to principal components analysis [19]. This approach should be connected with the definition of phonons as elementary acoustic signals proposed in the Theory section. Determination of the elementary constituents of a vocalisation in terms of phonons may contribute to a basic understanding of form and function of acoustic communication.

The bispectrum can be used in the same way; it is also useful for the characterisation of nonlinear systems. Both ambiguity function and CoSTID are defined in a spectro-temporal domain. The ambiguity function is expressed in terms of spectral difference ν and temporal difference τ. It appears an appropriate function when signals and their reflections are involved; with a well chosen test signal the ambiguity function may show distance and velocity of a sound-reflecting object. The CoSTID appears to be closely related to auditory perception; it can be displayed as a colour image with simple relations of visual features of the image and auditory features of the sound.

A somewhat different spectro-temporal signal representation can be derived from the symmetrical form of the product function

$$\Pi(t - \tau/2, t + \tau/2) = \xi^*(t - \tau/2) \cdot \xi(t + \tau/2) \tag{8a}$$

The Fourier transform of Eqn. 8a with respect to τ is a function of frequency ω and time t: $W(\omega, t)$. Since the product function given in Eqn. 8a is a Hermitian matrix its Fourier transform $W(\omega, t)$ is real valued. In quantum mechanics this function is known as the Wigner distribution function. Defined with spatial coordinate $\vec{x}$ and wave vector $\vec{k}$ as arguments $W(\vec{k}, \vec{x})$ supplies a measure for the coherence of light [7]. For the spectro-temporal representation of sound W has been proposed by de Bruijn [10]. Since W is real valued it can be represented in grey scaled image. However, the $W(\omega, t)$ function cannot be factorised in a spectral and a temporal part; also, no simple relation has been found between integrals or derivatives of $W(\omega, t)$ and the signal parameters a, α, b and β given in Table I. For these reasons we tend to prefer the CoSTID.

Since the real-valued Wigner distribution function is mathematically isomorphic with the complex valued CoSTID it is expected that some relation exists between the real and imaginary parts of the CoSTID. That this is in fact the case has been formulated in Property IV given in the Theory section. The imaginary part of the CoSTID is determined by the real part. For this reason it may appear that for a single signal the imaginary part of the CoSTID is superfluous. From this point of view the phonochrome might be reduced to a grey image. However, the relations of the acoustical parameters with the local aspects of the phonochrome as given in Table II are then not valid any more. Moreover it should be realised that the existence of mathematical relations between real and imaginary parts does not imply that one of the two should not be included in the phonochrome: mathematical isomorphism does not necessarily imply perceptual isomorphism, redundancy may support perception. Moreover, for the average CoSTID the imaginary part is not determined by the real part.

The function of the auditory system in a natural environment is both the identification and localisation of an acoustic source. For localisation two acoustic signals are needed. The introduction of the cross-CoSTID for the two signals gives the possibility to represent the sound as present at the two ears in the form of a phonochrome; in the cross-CoSTID real and imaginary parts are independent. Since the CoSTID is sensitive for phase relations of different signals it seems probable that the spatial position of a sound source has its influence on the appearance of the phonochrome. For a given sound given by N sampled values, the resulting phonochrome includes N^2 chromatic points. As a consequence N different signals can be stored in one phonochrome. Making use of the cross-CoSTID we suggest the possibility that N binaural sounds of N samples each can be stored in a $N \times N$ chromatic image.

The purpose of the phonochrome is the creation of a physical one-to-one map of acoustical signal and optical image such that there exists a simple perceptual correspondence of auditory features of the sound and visual features of the image. Two different mappings have been introduced. The Red–Green phonochrome, of which the global features, and the Saturation–Hue phonochrome, of which the local features correspond with characteristics of the sound. In both representations the isochromes relate to acoustical features. In this paper the mathematical formulation has been given together with some simple examples. The coherent spectro-temporal representation of signals and sounds based on the CoSTID has been applied in two different areas. For analysis of the electrocardiogram CoSTIDs have been determined for single ECGs [15] as well as the average CoSTID for an ensemble of ECGs [23]. In a neurophysiological investigation of the auditory midbrain of the frog the CoSTID has been used for the determination of the spectro-temporal receptive field of single neurons [16].

ACKNOWLEDGEMENTS

This investigation forms part of the research program of the Neurophysics Workgroup of the Department of Medical Physics and Biophysics, University of Nijmegen. The work has been supported by the Netherlands Organization for Advancement of Pure Research (ZWO). Technical support has been given by J. Bruijns and W.T. van Deelen; J.J. Eggermont gave constructive suggestions on a preliminary draft of the manuscript.

REFERENCES

[1] Ackroyd, M.H. (1970): Instantaneous spectra and instantaneous frequency. Proc. I.E.E.E. 58, 141.

[2] Aertsen, A.M.H.J. and Johannesma, P.I.M. (1979): Spectro-temporal analysis of auditory neurons in the grass frog. In: Hearing Mechanisms and Speech, pp. 87–93. Editors: O. Creutzfeldt, H. Scheich and Chr. Schreiner. Exp. Brain Res. Suppl. II.

[3] Aertsen, A.M.H.J. and Johannesma, P.I.M. (1980): Spectro-temporal receptive fields of auditory neurons in the grass frog. I. Characterization of tonal and natural stimuli. Biol. Cybernetics 38, 223–234.

[4] Altes, R.A. (1978): Angle estimation and binaural processing in animal echolocation. J. Acoust. Soc. Am. 63, 155–173.

[5] Altes, R.A. (1978): Possible reconstruction of auditory signals by the central nervous system (abstract). J. Acoust. Soc. Am. 64 Suppl. I, S137.

[6] Altes, R.A. and Titlebaum, E.T. (1970): Bat signals as optimally Doppler tolerant waveforms. J. Acoust. Soc. Am. 48, 1 014–1 020.

[7] Bastiaans, M.J. (1978): The Wigner distribution function applied to optical signals and systems. Optics Commun. 25, 26–30.

[8] Beuter, K.J. (1976): System-theoretische Untersuchungen zur Echo-ortung der Fledermäuse. Dissertation, University of Tübingen.

[9] Bullock, T.H. (ed.) (1977): Recognition of Complex Acoustic Signals. Life Sciences Res. Rep., Vol. 5. Springer-Verlag, Berlin.

[10] De Bruijn, N.G. (1973): A theory of generalized functions with applications to Wigner distribution and Weyl correspondence. Nieuw Arch. Wiskd. 3-XXI, 205–280.

[11] Flanagan, J.L. (1965): Speech Analysis, Synthesis and Perception. Springer-Verlag, Berlin.

[12] Flaska, M.D. (1976): Crosscorrelation of short-time spectral histories. J. Acoust. Soc. Am. 59, 381–388.

[13] Gabor, D. (1946): Theory of communication. J.I.E.E. (London), Part III, 93, 429–457.

[14] Gelder, J.J. van, Evers, P.M.G. and Maagnus, C.J.M. (1978): Calling and associated behaviour of the common frog, *Rana temporaria,* during breeding activity. J. Anim. Ecol. 47, 667–676.

[15] Heringa, A., Uijen, G.J.H. and Johannesma, P.I.M. (1980): Spectro-temporal representation of the electro-cardiogram. In: Proc. 7th Int. Congress on Electrocardiology, Lisbon.

[16] Hermes, D.J., Aertsen, A.M.H.J., Johannesma, P.I.M. and Eggermont, J.J. (1981): Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with noise stimuli. Hearing Res. 5, 147–178.

[17] Johannesma, P.I.M. (1972): The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: Proceedings of the IPO Symposium on Hearing Theory, pp. 58–69. Editor: B.L. Cardozo, Eindhoven.

[18] Johannesma, P.I.M. and Aertsen, A.M.H.J. (1979): Neural image of sound in the grass frog. In: Hearing Mechanisms and Speech, pp. 79–86. Editors: O. Creutzfeldt, H. Scheich and Chr. Schreiner. Exp. Brain Res. Suppl. II.

[19] Middleton, D. (1960): An Introduction to Statistical Communication Theory. McGraw-Hill, New York.
[20] Potter, R.K., Kopp, G.A. and Green, H.C. (1947): Visible Speech. Van Nostrand, New York.
[21] Rihaczek, A.W. (1968): Signal energy distribution in time and frequency. IEEE Trans. Inf. Theory IT-14, 369–374.
[22] Rihaczek, A.W. (1969): Principles of High-Resolution Radar. McGraw-Hill, New York.
[23] Uijen, G.J.H., Heringa, A. and Johannesma, P.I.M. (1980): Spectro-temporal averaging of electrocardiograms. In: Proc. 7th Int. Congress on Electrocardiology, Lisbon.
[24] Ville, J. (1948): Théorie et applications de la notion de signal analytique. Cables Transm. 2, 61–74.
[25] Wannier, G.H. (1960): Elements of Solid State Theory. Cambridge University Press, London.
[26] Woodward, P.M. (1953): Probability and Information Theory with Applications to Radar. Pergamon Press, London.
[27] Worden, F.G. and Galambos, R. (eds.) (1972): Auditory processing of biologically significant sounds. Neurosci. Res. Prog. Bull., Vol. 10.
[28] Wyszecki, G. and Stiles, W.S. (1966): Color Science: Concepts and Methods, Quantitative Data and Formulas. Wiley, New York.