The Phonochrome –
A Coherent Spectro-Temporal Representation of Sound

P.I.M. JOHANNESMA and A.M.H.J AERTSEN[1]

1 Introduction

In order to find an optimal visual image of sound three general desiderata for an isomorphic representation are formulated.

Formal equivalence. The relation of sound and image should be unique. In the image all information contained in the sound should be preserved including phase relations. The optical-acoustic map should be one to one.

Operational definition. The representation can be implemented instrumentally on the base of the sound without additional information regarding context or meaning and without human interference. Preferentially the representations are made on a digital computer with conventional displays. Real time operation is a desirable aspect.

Perceptual congruence. Continuity with the tradition of representation of sound in music (notes on a staff) and in vocalisations (sonogram). Moreover a simple correspondence should exist between perceptual elements with associated distances and relations in the original sound (audition) and the perceptual elements with associated distances and relations in the resulting image (visual).

2 Theory and Representation

The mathematical solution starts from the analytic signal $\xi(t)$ associated with the original sound $x(t)$

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

where

$$\tilde{x}(t) := \frac{1}{\pi} \int ds \, \frac{x(s)}{t-s}$$

is the Hilbert transform or quadrature signal of $x(t)$. In the spectral domain

---

[1] Present address: Max-Planck-Institut für biologische Kybernetik, Tübingen, FRG

Localization and Orientation in Biology and Engineering
ed. by Varjú/Schnitzler

$$\hat{\xi}(\omega) = \begin{cases} 2\,\hat{x}(\omega) & \omega > 0 \\ 0 & \omega < 0 \end{cases} \tag{2}$$

where $\hat{x}(\omega)$ is the spectral transform of x(t) and $\hat{\xi}(\omega)$ of $\xi(t)$.

The presentation is based upon a second order functional of the analytic signal. Four equivalent functionals exist: (1) product function $\Pi(\tau, t)$; (2) ambiguity function $\Delta(\tau, \nu)$; (3) CoSTID $\Xi(\omega, t)$; (4) bispectrum $\Gamma(\omega, \nu)$. Mathematically these four representations are equivalent, perceptually a spectro-temporal form is desired. Therefore only ambiguity function and CoSTID remain. The similarity with sonogram representing frequency and time, not their differences, eliminates the ambiguity function. As a consequence we propose the Coherent Spectro-Temporal Intensity Density (CoSTID) as the formal base for the visual representation of sound. Mathematically it is defined on the base of the product function

$$\Pi(\tau, t; a) := \xi^*[t + (a - 1/2)\tau]\,\xi[t + (a + 1/2)\tau] \quad -\frac{1}{2} \leqslant a \leqslant \frac{1}{2}\,. \tag{3}$$

The CoSTID is now the spectral transform with respect to the temporal difference τ

$$\Xi(\omega, t; a) := \int d\tau\, e^{-i\omega\tau}\, \Pi(\tau, t; a)\,. \tag{4}$$

For $a = 0$ the product function Π is Hermitic in τ and the CoSTID is a real valued function, not necessarily positive, of frequency and time.

In physical optics the CoSTID with $a = 0$ has been introduced by Wigner (1932) and used by Bastiaans (1979) and Wolf (1982). In acoustics this function is used by Claasen et al. (1980) and Flandrin and Escudie (1980). For $a = \pm\, 1/2$ the product function Π is not Hermitic and the CoSTID becomes a complex function. However, it can now be written as a product of spectral and temporal aspects with an intermediating spectro-temporal (de)modulation.

$$\Xi(\omega, t) := \Xi(\omega, t; -1/2) \tag{5}$$

$$(4) \rightarrow \quad = \int d\tau\, e^{-i\omega\tau}\, \Pi(\tau, t; -1/2)$$

$$(3) \rightarrow \quad = \int d\tau\, e^{-i\omega\tau}\, \xi^*(t - \tau)\, \xi(t)$$

$$= \hat{\xi}^*(\omega)\, e^{-i\omega t}\, \xi(t)\,. \tag{6}$$

This function has been proposed by Rihaczek (1968).

Johannesma et al. (1981) introduced it as a coherent spectro-temporal image of sound. Altes (1980) discussed it in relation to echolocation. Hermes et al. (1981) applied it to the spectro-temporal sensitivity of neurons. The relation of CoSTID to other second order functionals is discussed in Aertsen et al. (1983).

The CoSTID cannot be regarded as a physical entity but should be interpreted as a formal structure defined on a signal which by application of appropriate operators produces physical entities.

The phonochrome is a representational structure of a signal based on the CoSTID and intended for visual perception. Its realisation is formed through a representation of a complex function of two variables by a chromatic image. For an extensive discussion and presentation of phonochromes of different signals see Johannesma et al. (1981).

3 Identification and Localization

For identification and localization of an acoustic source of which only the air pressure variations can be observed, two receivers are needed. Not directly considering biological realisation or technical implementation it is possible to evaluate the CoSTID for this purpose. If the microphones receive signals $x_1(t)$ and $x_2(t)$ then their difference will be mainly in the phase of the signals. Now two signals are formed:

$$x_{\pm} = \frac{1}{2}(x_1 \pm x_2) \tag{7}$$

and for each the associated CoSTID $\Xi_{\pm}$:

$$\Xi_{\pm} = \frac{1}{2}\left\{\Xi_{11} + \Xi_{22} \pm (\Xi_{12} + \Xi_{21})\right\}. \tag{8}$$

Now take even and odd part

$$\Xi^{e} = \Xi_{+} + \Xi_{-} = \frac{1}{2}(\Xi_{11} + \Xi_{22}), \tag{9a}$$

$$\Xi^{o} = \Xi_{+} - \Xi_{-} = \frac{1}{2}(\Xi_{12} + \Xi_{21}). \tag{9b}$$

Then Ξ^{e} is based only on the autoCoSTID's
and Ξ^{o} is based only on the crossCoSTID's.

As a consequence Ξ^{e} is weakly and Ξ^{o} strongly dependent on the phase relations of x_1 and x_2 and as a consequence on the position of the source.

For active localization and identification an analogous way of reasoning may be applied. However now x_1 is the emitted signal and x_2 is the reflected signal. Comparison of Ξ_{22} with Ξ_{11} may lead to identification while evaluation of Ξ_{12} and Ξ_{21} may supply the clues for localization.

In this context the function of EE- and EI-neurons in the central auditory nervous system should be considered.

Acknowledgements. This investigation forms a part of the research program "Brain and Behaviour: Neural base of Localization and Orientation". The work has been supported by the Netherlands Organisation for Advancement of Pure Research (Z.W.O.).

References

Aertsen AMHJ, Johannesma PIM, Bruijns J, Koopman PWM (1983) Second order representation of signals. In: Varjú D (ed) Orientation and localization in engineering and biology. Proc 8th Cybernetics Congress, Tübingen, B.R.D.

Altes RA (1980) Models for echolocation. In: Busnel RG, Fish JF (eds) Animal sonar systems. Plenum Publ Corp, New York, pp 625–671

Bastiaans MJ (1979) Wigner distribution and its application to first order optics. J Opt Soc 69: 1710–1716

Claasen TACM, Mecklenbräuker WFG (1980) The Wigner distribution – A tool for time frequency signal analysis. I. Continuous time signals. II. Discrete time signals. III. Relations with other time-frequency signal transformations. Philips J Res 35:217–250; 276–300; 372–389

Flandrin P, Escudie B (1980) Time and frequency representation of finite energy signals: a physical property as a result of an Hilbertian condition. Signal Processing 2:93–100

Hermes DJ, Aertsen AMHJ, Johannesma PIM, Eggermont JJ (1981) Spectrotemporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grassfrog (Rana temporaria L.) investigated with noise stimuli. Hear Res 5:145–179

Johannesma PIM, Aertsen AMHJ, Cranen B, Erning L v (1981) The phonochrome: a coherent spectro-temporal representation of sound. Hear Res 5:123–145

Rihaczek AW (1968) Signal energy distribution in time and frequency. IEEE Tr on Information Theory IT-14:369–374

Wigner E (1932) On the quantum correction for thermodynamic equilibrium. Phys Rev 40: 749–759

Wolf E (1982) New theory of partial coherence in the space-frequency domain I. JOSA 72–3: 343–351