Proc. 1987 Int. Conf. on Digital Signal Processing,
Florence, Italy, Sept. 1987, pp. 294–298

WIGNER DISTRIBUTION ANALYSIS OF SPEECH SIGNALS

Wolfgang Wokurek, Franz Hlawatsch, Gernot Kubin

Institut für Nachrichtentechnik und Hochfrequenztechnik
Technische Universität Wien
Gusshausstrasse 25/389, A-1040 Vienna, Austria

The *spectrogram*, traditionally used for time-frequency analysis of speech signals, is a smoothed version of another time-frequency representation, *Wigner distribution* (WD). WD is thus the basis for time-frequency analysis with superior resolution. This paper compares a modification of WD, the *Smoothed Pseudo Wigner Distribution* (SPWD), with conventional spectrograms. SPWD is shown to allow representation of *both* the time (pitch) and frequency (formant) structure of vowels as well as accurate analysis of highly nonstationary features of plosive sounds.

1. INTRODUCTION

Due to their nonstationarity, speech signals are often analyzed using a time-frequency signal representation. Usually *spectrograms* are employed for this purpose; they are mostly displayed in a highly time-compressed form which allows whole words or phrases to be visualized in a single picture [1].

The spectrogram can be shown to be a smoothed version of another time-frequency representation, *Wigner distribution* (WD) [2]. Due to the smoothing process involved, many details of the signal's time-frequency structure which are shown by WD are blurred or altogether invisible in the spectrogram. WD is thus the basis for time-frequency signal analysis with substantially improved resolution. Application of WD, however, is complicated by *interference terms* [3] which have to be partially suppressed. This paper discusses WD's problems and advantages as well as some important aspects of implementation. Typical results of WD and spectrogram in the case of speech signals are finally given and compared.

2. THE WIGNER DISTRIBUTION

Signal theoretical aspects of WD

$$WD_x(t,f) = \int_\tau x(t+\tau/2)x^*(t-\tau/2)\, e^{-j2\pi f\tau}\, d\tau \quad \epsilon\ R$$

($x(t)$=signal, t=time, f=frequency) are extensively analyzed in [2]. We here content ourselves with discussing two aspects of WD important for applications, *resolution* and *interference*. Consider the WD of a superposition of time-domain impulses $x(t)=\delta(t-t_1)+\delta(t-t_2)$,

$$WD_x(t,f)= \delta(t-t_1)+ \delta(t-t_2)+ 2\delta(t-t_m)\cos(2\pi t_d f)$$

where $t_m=(t_1+t_2)/2$ and $t_d=t_2-t_1$ (*Fig. 2.a*).

The first two WD terms ("signal terms") retain the ideal time concentration of the two signal components, which shows the *ideal time resolution* of WD: WD does not introduce any analysis uncertainty. The third WD term (cross or *interference term* [3]) reflects WD's quadratic nature. This WD interference term is located at the center ($t=t_m$) between the two "interfering" signal terms; it *oscillates* in the frequency direction with period $f_d=1/t_d$, where $t_d=t_2-t_1$ is the distance between the signal terms.

A dual situation (*Fig. 2.b*) exists in the case of two complex sinusoids (frequency-domain impulses) $x(t)=\exp(j2\pi f_1 t)+\exp(j2\pi f_2 t)$. This example shows WD's ideal *frequency* resolution. There is again an interference term at the center $f_m=(f_1+f_2)/2$ oscillating in the time direction with period $t_d=1/f_d$, where $f_d=f_2-f_1$ is again the distance between the signal terms.

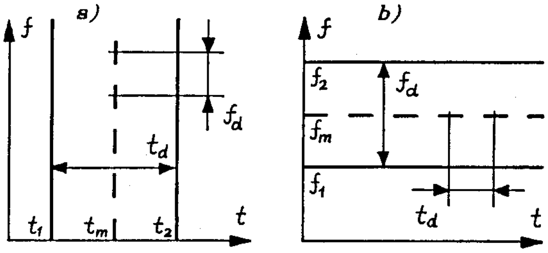


Fig.2. Resolution and interference in WD

3. SMOOTHING WD INTERFERENCE TERMS

WD interference terms are a problem in applications as they tend to obscur WD plots. Because of their oscillation, they can be (partially) suppressed by *time-frequency smoothing*. We next discuss two smoothed WD versions.

3.1 The SPWD

A versatile and computationally efficient smoothed WD version is the "*smoothed pseudo Wigner distribution*" (SPWD) [2],[4]

$$\mathrm{SPWD}_x(t,f) =$$

$$\left[\int_{\tau} x(t+\tau/2)x^*(t-\tau/2)w^2(\tau/2)e^{-j2\pi f\tau}d\tau \right] \underset{t}{*} u(t)$$

$$= \mathrm{WD}_x(t,f) \underset{t}{*} \underset{f}{*} u(t)v(f) \ . \qquad (3.1)$$

We see from (3.1) that two different (finite-length) windows w(t),u(t) are used in the calculation of SPWD. The window w(t) truncates time integration and causes *frequency smoothing* of WD by $v(f)=\mathrm{WD}_w(0,f)=\mathcal{F}\, w^2(t)$: a *shorter* window w(t) will produce a *broader* v(f) and thus more frequency smoothing. The second window, u(t), is directly used for *time smoothing* of WD; a longer window u(t) will produce more time smoothing.

An important feature of SPWD is that the amounts of smoothing in the time and frequency direction can be *arbitrarily and independently* controlled through the choice of the "smoothing windows" u(t) and v(f). This choice should be based on the interference term geometry of the signal to be analyzed. To illustrate this important point, we recur to the two basic examples of Section 2: the WD interference term of two time-domain impulses with distance t_d (Fig. 2.a) oscillates in the frequency direction with period $f_d=1/t_d$; to suppress this interference term, the (effective) length of the frequency smoothing window v(f) of (3.1) should be in the order of f_d (or greater). In the dual case of two frequency - domain impulses f_d apart (Fig. 2.b), the WD interference term oscillates in the time direction with period $t_d=1/f_d$; it will be suppressed if the length of the time smoothing window u(t) of (3.1) is approximately t_d (or greater).

The other side of smoothing is, of course, the reduced resolution of SPWD. Contrary to WD itself, SPWD does introduce analysis uncertainties Δt, Δf which equal the (effective) lengths of windows u(t), v(f). Like any smoothed WD, SPWD is thus a tradeoff between good resolution and good interference term suppression: more smoothing will suppress more interference terms but will yield poorer resolution.

We may summarize resolution and interference properties of SPWD by the following rule-of-the-thumb: if t_u, f_v are the (effective) lengths of the smoothing windows u(t), v(f), then (1) SPWD time and frequency resolution are $\Delta t=t_u$, $\Delta f=f_v$ respectively, and (2) interference terms are suppressed if the interfering signal components have time distance $t_d>1/f_v$ or frequency distance $f_d>1/t_u$.

3.2 The spectrogram

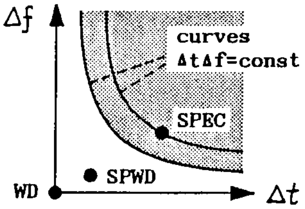
Just as SPWD, the traditionally used *spectrogram* [1],[2]

$$\mathrm{SPEC}_x(t,f) = \left| \int_{\tau} x(\tau)w(\tau-t)e^{-j2\pi f\tau}d\tau \right|^2$$

$$= \mathrm{WD}_x(t,f) \underset{t}{*} \underset{f}{*} \mathrm{WD}_w(-t,f) \qquad (3.2)$$

is a time-frequency smoothed WD. Comparing (3.1) and (3.2), however, we notice a fundamental difference between SPWD and SPEC: while the SPWD time-frequency smoothing kernel u(t)v(f) is a separable function of t and f whose factors are two independent windows, the SPEC smoothing kernel $\mathrm{WD}_w(-t,f)$ is a WD (of the analysis window w(t)) whose time and frequency characteristics are coupled by the uncertainty relation. Contrary to SPWD, therefore, *the amounts of time and frequency smoothing cannot be chosen independently in the SPEC*; rather, they are reciprocally coupled through the uncertainty relation $\Delta t\Delta f \gtrsim 1$, where Δt,Δf are the "effective lengths" of $\mathrm{WD}_w(-t,f)$ in time and frequency. This means that good frequency resolution (Δf small, obtained with long window w(t)) comes at the cost of poor time resolution (Δt large), and vice versa. The overall amount of smoothing (as expressed by the product $\Delta t\Delta f$) cannot be made smaller than a fundamental constant and is so large that most of WD's interference terms are typically suppressed in SPEC.

Compared to SPWD, therefore, SPEC allows only very restricted control in the tradeoff between good resolution and good interference term suppression. This is illustrated by ***Fig. 3***: while SPWD can realize any point in the (Δt,Δf) quadrant, SPEC can only assume points on curves $\Delta t\Delta f$=const.



▨ SPEC region

Fig.3.
Time-frequency resolution
of SPWD and SPEC

Table 3 compares smoothing kernels and resolutions of SPWD and SPEC.

	smoothing kernel	resolutions Δt, Δf
WD	$\delta(t)\delta(f)$	$\Delta t=\Delta f=0$
SPWD	u(t)v(f)	arbitrary & independent
SPEC	$\mathrm{WD}_w(-t,f)$	$\Delta t\Delta f \gtrsim 1$

Table 3

4. TIME-FREQUENCY ANALYSIS OF SPEECH SIGNALS

Speech signals are partly "quasi-stationary" (voiced sounds, fricatives) and partly extremely non-stationary (explosion phase of stop consonants). Conventional SPEC analysis gives a satisfactory overview of the quasi-stationary speech portions but is inadequate for displaying fast temporal changes of the signal. SPWD is an ideal tool for displaying such changes. The good time resolution of SPWD is, of course, incompatible with a highly time-compressed representation. An SPWD plot will hence capture some pitch periods (some tens of milliseconds) while a whole word or phrase may be pressed into a SPEC plot. This "microscopic" character of SPWD is a basic difference to conventional spectrogram analysis.

Successful application of SPWD requires the observation of some implementation rules. Apart from a suitable choice of time-frequency smoothing (see Section 3), we suggest the following mode of implementation:

(1) Use of the *analytic signal* is mandatory to avoid SPWD aliasing [2] and interference between positive and negative frequencies.

(2) For voiced speech, some sort of *preemphasis* must be applied to the signal in order to accentuate higher frequencies.

(3) SPWD may be *decimated in time* depending on the amount of time smoothing used. This yields a significant reduction of computation and storage requirements.

(4) *Contour line plots* showing *positive* heights only are better adapted to SPWD's interference structure than axonometric 3D plots.

(5) SPWD *height compression* (e.g. logarithmic) is often useful for increasing dynamic range, i.e. emphasizing low-energy signal terms.

The plots shown in the next two sections were produced in accordance to the above rules.

4.1 Vowels

The signals corresponding to *vowels* are produced by (quasi-) periodic excitation of the vocal tract and can be modeled as a periodic superposition of impulse responses. The signals have a regular time structure (*excitation* or *pitch periodicity*) as well as a regular frequency structure (*formants* = resonance frequencies of the vocal tract) [1]. In contrast to SPEC, SPWD displays both the time and the frequency structure simultaneously. This is seen from ***Fig. 4.1.b*** which shows the SPWD (resolution $\Delta t \approx 1$ms, $\Delta f \approx 100$Hz) of three pitch periods of the sound [a] spoken by a male speaker; the corresponding signal is plotted in ***Fig. 4.1.a.*** Within one pitch period, SPWD shows a broad-band, time-concentrated signal term *E* (impulse corresponding to *excitation)* and three narrow-band signal terms *F1, F2, F3* (corresponding to the three *formants*). An interference term oscillating in the time direction (comp. Fig. 2.b) exists between formant signal terms F1 and F2 (which are closest); all other interference terms (e.g. those between excitation terms which would oscillate in the frequency direction, comp. Fig. 2.a) are suppressed by the SPWD smoothing. Further interference terms of SPWD quickly show up if the resolution is improved; this is illustrated by ***Fig. 4.1.c.***

For comparison, ***Fig. 4.1.d*** shows the SPEC of the same signal segment, with frequency resolution equal to that of SPWD ($\Delta f \approx 100$Hz). As a consequence, time resolution of SPEC is comparatively poor ($\Delta t \approx 6$ms), which results in the excitation terms (E) being spread over several milliseconds in SPEC. The intervening formant terms (F_i) are thus largely "buried" by the broadened excitation terms. *The superior time resolution of SPWD is thus also the basis for a clearer representation of formant frequencies and formant bandwidths.* The "burying effect" of SPEC may be mitigated by the choice of a better time resolution. ***Fig. 4.1.***e shows a SPEC with $\Delta t \approx 2$ms; unfortunately, frequency resolution is now considerably worse ($\Delta f \approx 300$Hz, according to $\Delta t \Delta f$=const) so that formants F1 and F2 cannot be distinguished at first sight. We also see that the *interference term* between F1 and F2 shows up in the SPEC due to the insufficient time smoothing. Interference terms are thus seen to occur not only in SPWD, but also in SPEC (indeed, in any smoothed WD).

In both SPWD and SPEC with high time resolution, an awkward situation arises when an interference term coincides in frequency with a formant signal term (see ***Fig. 4.2.a*** for an SPWD example); this will be the case if a formant lies midway between two other ones. More time smoothing will solve this problem at the cost of poorer time resolution (***Fig. 4.2.b***).

SPWD can also be used for comparing natural and synthetic speech. ***Fig. 4.2.c*** shows an attempt to imitate a segment of natural speech (sound [e], see Fig. 4.2.b) by the output of a parallel-resonator model. The synthetic signal is $s(t)=p(t)*h(t)$, where the "vocal tract filter", with impulse response

$$h(t) = \begin{cases} \sum_{i=1}^{4} e^{-\lambda_i t} \cos(2\pi f_i t) \ , & t>0 \ , \\ 0 \ , \quad t<0 \ , & \end{cases}$$

is excited by a gaussian glottis pulse train

$$p(t) = \sum_{n=-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-nT}{\tau}\right)^2} \ .$$

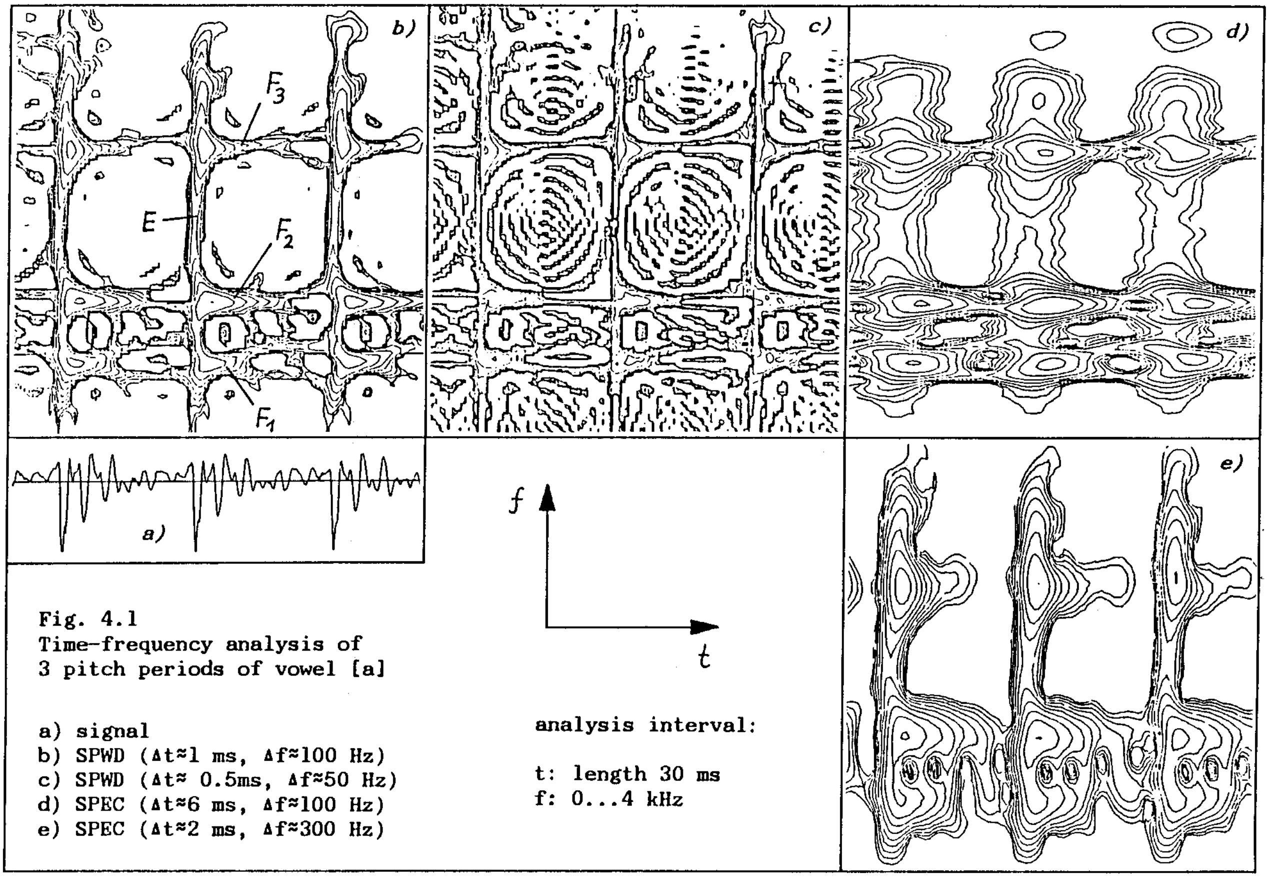


Fig. 4.1
Time-frequency analysis of
3 pitch periods of vowel [a]

a) signal
b) SPWD (Δt≈1 ms, Δf≈100 Hz)
c) SPWD (Δt≈ 0.5ms, Δf≈50 Hz)
d) SPEC (Δt≈6 ms, Δf≈100 Hz)
e) SPEC (Δt≈2 ms, Δf≈300 Hz)

analysis interval:

t: length 30 ms
f: 0...4 kHz

4.2 Unvoiced stop consonants

Stop consonants (plosive sounds) are characterized by a closure in the vocal tract which suddenly breaks open, releasing the air pressure built up during the closure phase and causing a turbulent air flow corresponding to a noise-like signal [1]. *Fig. 4.3* compares SPWD and SPEC for the explosion phase of the first [t] from the German word [ta:t] (male speaker). Frequency resolution of both SPWD and SPEC is 100Hz; time resolution is 1ms in SPWD and 6ms in SPEC. The explosion interval consists of three phases: (1) *Plosion phase P* (4ms): impulse-like transient due to the release of the pressure built up behind the vocal-tract closure; (2) *Frication phase F* (25 ms): noise phase due to turbulent air flow at the opening constriction; (3) *Aspiration phase A* (30 ms): noise phase with formant structure due to excitation of vocal-tract resonances by a turbulent air flow at the glottis.

Fig. 4.3 shows two advantages of SPWD over SPEC: (1) The short impulse of the plosion phase P is readily recognized in SPWD as a short-time, broad-band term; it is invisible in SPEC due to SPEC's poor time resolution. (2) In SPWD, noise-like excitation of the vocal tract manifests itself as an irregular, yet characteristic structure (*texture*) which is completely different from the case of "deterministic" or "voiced" excitation studied in the previous section. This difference is by far not so pronounced in SPEC. *Fig. 4.4* shows by way of simulation that the above-mentioned texture is indeed characteristic of noise signals: *Fig. 4.4.a* displays the SPWD of synthetic noise which was filtered and windowed so as to serve as a model of the frication phase F; in *Fig. 4.4.b*, narrow-band synthetic noise is used as a model of the "noise-excited formants" found in the aspiration phase A.

5. CONCLUSION

The Wigner distribution is the basis for high-resolution time-frequency analysis of speech signals. We compared the *smoothed pseudo Wigner distribution* (SPWD) with conventional spectrograms and found SPWD to allow the *simultaneous* representation of the time (pitch, excitation) structure and the frequency (formant) structure of vowels, as well as a refined

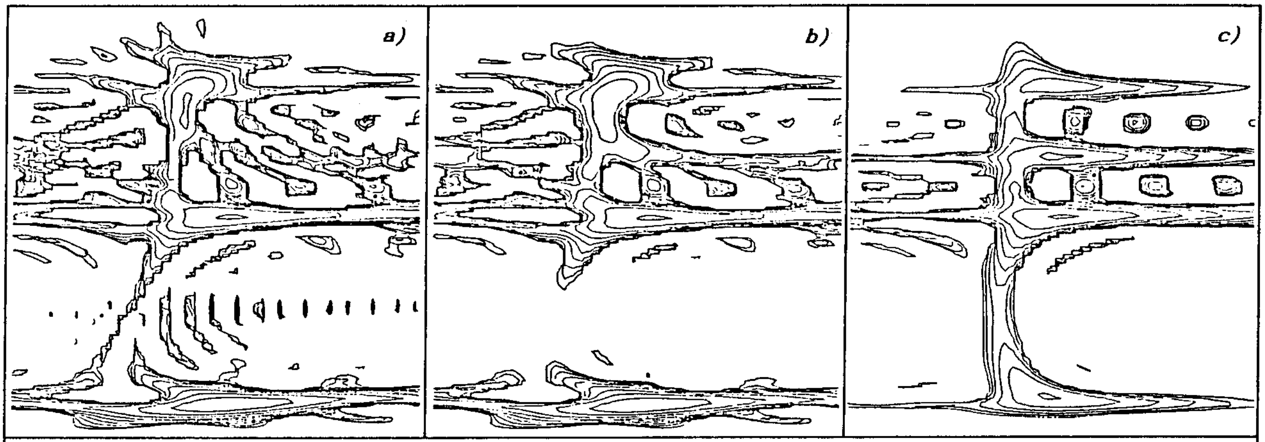


Fig. 4.2. Time-frequency analysis of 1 pitch period of vowel [e]

a) SPWD (Δt≈0.5 ms, Δf≈100 Hz)
b) SPWD (Δt≈0.8 ms, Δf≈100 Hz)
c) SPWD (Δt, Δf as in b)) of synthetic speech signal

analysis interval:

t: length 10 ms
f: 0...4.5 kHz

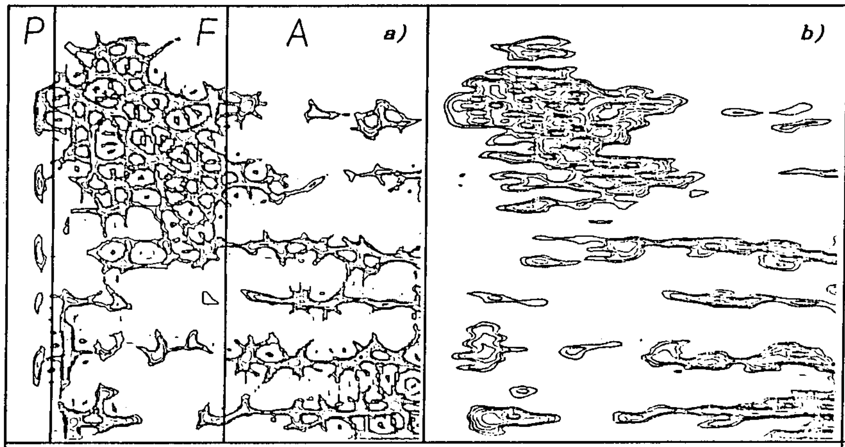


Fig. 4.3
Time-frequency analysis of explosion phase of plosive sound [t]

a) SPWD (Δt≈1 ms, Δf≈100 Hz)
b) SPEC (Δt≈6 ms, Δf≈100 Hz)

analysis interval:
t: length 60 ms, f: 0...9 kHz

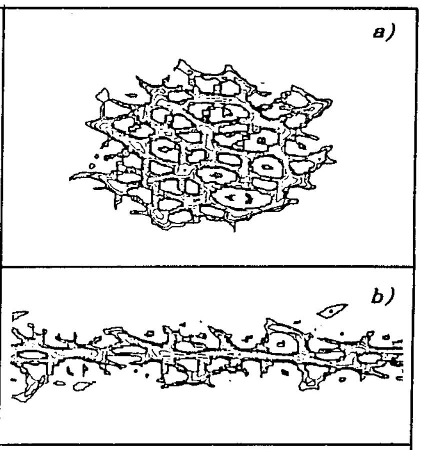


Fig. 4.4
SPWD of synthetic noise

a) broad-band, time-windowed
b) narrow-band

analysis of the highly nonstationary explosion phase of stop consonants. SPWD thus presents substantial advantages over spectrograms whenever an accurate description of a speech signal's time-frequency structure is desired.

References

[1] L.Rabiner, R.Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[2] T.Claasen, W.Mecklenbräuker, *The Wigner Distribution - a Tool for Time-Frequency Signal Analysis*, Philips J. Res. 35, pp. 217-250, 276-300, 372-389, 1980.

[3] F.Hlawatsch, *Interference Terms in the Wigner Distribution*, Proc. Int. Conf. Digital Signal Processing, Florence, pp.363-367, 1984.

[4] P.Flandrin, W.Martin, *Pseudo-Wigner Estimators for the Analysis of Nonstationary Processes*, Proc. IEEE ASSP Spectr. Est. Workshop II, Tampa, FL, pp. 181-185, 1983.