WIGNER DISTRIBUTION - A NEW METHOD FOR HIGH-RESOLUTION TIME-FREQUENCY ANALYSIS OF SPEECH SIGNALS

W. Wokurek, G. Kubin, F. Hlawatsch
Institut für Nachrichten- und Hochfrequenztechnik, TU Wien
Gusshausstr. 25/389, A-1040 Vienna, Austria

ABSTRACT

Two methods for the time-frequency analysis of speech signals are compared: the traditionally used Spectrogram and the Smoothed Pseudo Wigner Distribution (SPWD). It is shown that the time and frequency resolutions of the Spectrogram are restricted by the uncertainty relation while SPWD allows arbitrarily high resolutions. If the analysis parameters are chosen carefully SPWD yields more accurate signal representations than the Spectrogram. This is exemplified by a "microscopic" analysis of vowels and voiced stop consonants.

1. INTRODUCTION

The Wigner Distribution (WD) is a method for the time-frequency analysis of signals. Along with the Spectrogram it is a member of a special class of bilinear, shift-invariant signal representations (Cohen class [1] p.315). Within this paper we compare a somewhat modified WD, i.e. the Smoothed Pseudo Wigner Distribution (SPWD) to the Spectrogram, first with respect to the basic features of time- and frequency resolutions (Sections 2-4). In sections 5 and 6 we compare the results of representing speech signals through SPWD and Spectrogram.

Observing the fact that SPWD enables high-resolution signal representation, we analyze short speech segments of a few pitch periods of length. Therefore it is pointless to compare the SPWD to Spectrogram of high frequency resolution (45 Hz) because they do not display the fine structure in time that we will see in the WD. As a compromise we compare Spectrogram of high frequency resolution with those of higher resolution (which do not show the time structure) and those of high resolution (which means the same as we have already in (3) and (4)).

A second interpretation of (1) is band-pass filtering. If the signal contains oscillations of periods less than the time resolution \( \Delta t \), these components of the signal will be suppressed in the representation.

3. COUPLING OF THE RESOLUTIONS OF THE SPECTROGRAM

The Spectrogram is defined as the square magnitude of the Short-Time Fourier Transform (STFT). The signal is multiplied by a window \( w(t) \) that is shifted to the instant of analysis \( t \). The Fourier Transform of this product is associated with the instant \( t \).

4. INDEPENDENCE OF THE RESOLUTIONS OF THE SPWD

The Wigner Distribution (WD) of a signal \( x(t) \) is defined by

and its features are described in [13]. The WD does not show any effect of limited resolution, but in the case of fairly complex signals such as speech the result is quite unreadable owing to the occurrence of interference terms, described in [13] (see section 5 also). Therefore we consider the SPWD of the signal which is defined as a WD with arbitrary smoothing:

4. Smoothing in both the time and frequency direction is performed by two independently chosen arbitrary window functions \( w(t) \) and \( v(f) \), respectively. Because of the independence of the underlying functions, the resolutions of the SPWD are not restricted by the uncertainty relation (5) (see Fig 3: point D).

5. ANALYSIS OF VOWELS BY SPWD AND SPECTROGRAM

Figure 4 shows a contour plot of the SPWD of three successive pitch periods extracted from the German vowel [a:] spoken by a male subject. This representation displays the following features:

- (1) Quasi-periodic excitation of the vocal tract by wide band narrow-time impulses every 10 msec. The time resolution of approx. 0.5 msec is sufficient to prove that these impulses have a time width of 1 msec or less.
- (2) Exponential decay of three formants at the frequencies \( F_1 = 0.7 \text{ kHz}, F_2 = 1.25 \text{ kHz}, F_3 = 2.6 \text{ kHz} \).
- (3) Besides these signal terms (formants, impulses), the SPWD contains interference terms. They are governed by a simple geometric rule [2], i.e. they always lie half-way between two signal terms and oscillate in the direction perpendicular to the line connecting the two signal terms. These oscillations have a period in the time-frequency plane that is inverse proportional to the distance of the signal terms. The oscillator nature of interference terms is the key to their suppression in any bilinear time-frequency representation. In SPWD, this is achieved by smoothing with the two independent window functions according to (7).

The smoothing must be matched to the signal structure:
The frequency resolution $\Delta f = 100$ Hz is just great enough to damp interferences between successive pitch periods (remember interference oscillations to occur perpendicu-
lar to the line from one excitation impulse to the next, i.e. parallel to the frequency axis). The time resolution $\Delta t = 0.5$ msec is great enough to damp most of the inter-
ferences between neighboring formants. Accordingly, oscillations in the time direction can only be observed be-
tween F1 and F2, the two formants closest to each other.

Figure 5 shows a Spectrogram of the same signal seg-
ment with resolutions $\Delta f = 200$ Hz and $\Delta t = 2$ msec. Note
that the product of these resolutions equals $300$ Hz - 2 msec = 0.6 which is more than ten times the product of resolutions of SPWD in Figure 4
($100$ Hz - 0.5 msec = 0.05). The Spectrogram's resolu-
tions are already chosen so as to achieve a signal repre-
sentation as close as possible to SPWD. A simultaneous
improvement of the Spectrogram's resolutions is impos-
sible due to (5). Therefore the Spectrogram evidences
much broader excitation impulses in the time dimension as
well as much wider formants in the frequency dimension
due to the inherent smoothing of the Spectrogram. The inherently stronger smoothing of the
Spectrogram renders better suppression of interference
terms (though they are still perceivable between F1 and F2),
yet worse fidelity in signal terms than SPWD. An inter-
fERENCE TERMS are predictable from the above geometr-
ical rule, SPWD is better suited to the analysis of vowels
than the Spectrogram.

One may conjecture that a change of the Spectrogram window function $w(t)$ in (2) may improve its resolutions.
This has to be refuted when studying Figures 6 and 7.
In Figure 6, the time resolution of the Spectrogram is im-
proved to $\Delta t = 1$ msec, thus approaching the value of $\Delta t$
for SPWD in Figure 4. Due to the uncertainty relation (5),
the frequency resolution goes up to $\Delta f = 600$ Hz so that
the two lower formants F1 and F2 are merged Into a single
untouched band stretching over several hundred Hz.
In Figure 7, the frequency resolution of the Spectrogram is improved to $\Delta f = 100$ Hz as is the case for SPWD in
Figure 4. As time resolution has to go up to 6 msec, exci-
tation impulses are broadened drastically and spilled over
the formant structure even into the interval of the pitch
period without glottal excitation. Therefore, formant band-
width measurements are again more difficult than with
SPWD, inspite of the high frequency resolution of Figure 7.

Summarizing we observe that the Spectrogram is not
suited for simultaneous display of both the excitation and
the formant structure of vowels whereas SPWD has this
property notwithstanding its (easily controlled) interfer-
tence terms.

6. ANALYSIS OF UNVOICED STOP CONSONANTS BY
SPWD AND SPECTROGRAM

Figures 8 and 9 show the explosion interval (60 msec)
of the first /t/ in the German word ['tat'] making use of
SPWD and Spectrogram, respectively. For the sake of com-
parison, both displays have a frequency resolution of
100 Hz and associated time resolutions of 1 msec (SPWD)
and 6 msec (Spectrogram). The explosion interval consists of three more or less separable phases:
1. An impulse-like transient (about 4 msec) due to the
release of the pressure built up behind the vocal—tract clo-
cure (plosive phase P).
2. A noise phase extending from approx. 4 kHz to 6 kHz
(25 msec) due to the turbulent air flow at the opening
constriction (friction phase F).
3. A noise phase with a formant structure (30 msec) due
to the reocurrence of the open vocal tract excited by tur-
bulent air flow at the glottis (aspiration phase A).

The advantages of SPWD over the Spectrogram for the
analysis of this type of sounds can be summarized as follows:

(1) The short impulse of the plosion phase P is readily
seen in SPWD whereas the Spectrogram is not able to
resolve this temporal fine structure (at the given
frequency resolution).
(2) The boundary between fricative phase F and aspira-
tion phase A is more pronounced in SPWD than in the
Spectrogram.
(3) Noise-like excitation of the vocal tract manifests
itself as a very specific meshy texture in SPWD which
is clearly distinguishable from deterministic excitation as
seen in Figure 4. With the Spectrogram, noise-like ex-
citation induces no significant changes in the texture if
the contour plots when compared to deterministic ex-
citation as seen in Figures 5, 6, and 7.

7. CONCLUSIONS

From the above discussion, it should be clear that SPWD
is superior to the Spectrogram for the time—frequency
analysis of signal terms as typified by the examples
given in sections 5 and 6. It should be kept in mind,
however, that the comparison was made on the basis of
very short signal segments so as to emphasize SPWD's
character as a time—frequency "microscope". If the
analysis interval is extended to 1 second or more both
the resolutions of video displays and the human eye
become insufficient to realize the differences of the two
methods. Anyway, these long-time displays are only use-
ful for the compressed visualization of slowly time-vary-
ing and global features characterizing whole syllables or
worst. For the detailed high-resolution study of rapidly
time—varying speech phenomena, preference is to be
given to the new method.

8. REFERENCES

(1) T. Claasen, W. Mecklenbrauker, 1980; The Wigner Distribu-
tion — A Tool for the Time—Frequency Signal Analysis,
(2) F. Hlawatsch, 1984; Interference Terms in the Wigner Dis-
tribution, International Conference on Digital Signal
Processing; Florence, Italy