数理解析研究所講究録
第 880 巻 1994 年 1-9

Wavelets and Acoustical Signal Analysis

Mei Kobayashi and Masaharu Sakamoto, IBM Tokyo Research Laboratory
小林メイ 阪本正治
1623-14 Shimotsuruma, Yamato-shi, Kanagawa-ken 242 Japan
e-mail: mei or sakamoto@trlvm.vnet.ibm.com

Abstract: This report describes results from acoustical signal processing experiments using wavelets.

1. Introduction

Wavelets are *"families of functions $h_{a,b}$,*

$$h_{a,b} = |a|^{-1/2} h\left(\frac{x-b}{a}\right) \qquad ; \qquad a,b \in R \quad , \quad a \neq 0 \; ,$$

generated from a single function h by dilations and translations [D1]." One of the applications of the theory is to construct a basis set $\{h_{a,b}\}$ for efficient and accurate approximation of signals. In signal analysis the parameters a and b are restricted to a discrete sublattice; the dilation step $a_0 > 1$ and translation step $b_0 \neq 0$ are fixed. The corresponding wavelet family is $h_{m,n}(x) = |a_0|^{-m/2}\, h(a_0^{-m}x - nb_0)$, where $m,n \in Z$, $a = a_0^m$ and $b = nb_0a_0^m$ [D2]. We note that if the translation parameter b_0 is small, then the basis elements lie closer together, and the approximation is of a finer resolution. In some cases, a small value of b_0 may lead to overlap or redundancy.

Transforms are used in wavelet methods to encode the approximation of a function in much the same way as in Fourier methods. For wavelets with mother function h, the continuous wavelet transform (WT) for $f \in L^2(R)$ is defined as

$$(Uf)(a,b) = < h_{a,b}, f > = |a|^{-1/2} \int dx \cdot h\left(\frac{x-b}{a}\right) \cdot f(x)$$

for $a,b \in R$, $a \neq 0$, and the discrete WT

$$(Tf)_m = < h_{m,n}, f > = |a_0|^{-m/2} \int dx \cdot h(a_0^{-m}x - nb_0) \cdot f(x)$$

for $a_0 > 1$, $b_0 \neq 0$. Computation of transforms of wavelets with compact support automatically yields the same windowing effect as convolution with a time window function used in short-time spectral analysis and synthesis; however, wavelet computations are better because of a constant frequency-bandwidth ratio. Further details on the theory, examples, and a discussion of the error, as well as practical advice on computing transforms and their inverses, are given in [D1],[D2], and [D3].

Use of the WT shows promise in a variety of scientific and engineering applications; however, its advantages over conventional methods have not been clearly established, and further study is needed.[1] There are many types of wavelets, and it appears that success in using the method depends on the choice of the family and the associated computational algorithms. Shanon, Fourier, Gabor, and various WTs and some of their properties are reviewed in [Hu]. A detailed account of wavelets used in our earlier experiments is given in [KS]. We present a brief review of related work by others in the next section before discussing our speech signal processing experiments using wavelets.

2. Review of Acoutical Signal Studies using Wavelets

Early acoustical signal processing work using wavelets was conducted by Wickerhauser et al. at Yale [W1] and Kronland-M., Grossman, Morlet, et al. [KMG] in Marseille. Wickerhauser generalized the wavelet transform to *"produce a library of orthonormal bases of modulated wave packets, where each basis comes with a fast transform."* The Marseille group used a modulated Gaussian analyzing wavelet

$$h(t) = e^{i\omega_0 t} \cdot e^{-t^2/2}$$

multiplied by a normalization factor plus some negligible correction terms to generate songrams and phase diagrams of five signals: a single $\delta-$ function spike, two succesive $\delta-$ function spikes, sawtooth spikes, the syllables *pap* in

[1] The Japan IEICE established a closed working group in 1994 to issue a report on wavelet technologies in 1995.

papy and *tat* in *taty*, and notes on a clarinet. The paper concludes that *"the preliminary results (indicate) the combined information on modulus and on the phase of the WTs is useful for the segmentation of speech sounds* [KMG]."

In acoustical analysis experiments, the WT of a signal is calculated for a different series of wavelet lengths to generate a time-frequency diagram. S-traightforward computation becomes very intense and perhaps even impossible, depending on the application and data size. The *Algorithme à Trous* was introduced by Holschneider, Kronland-Martinet, Morlet, and Tchamitchian [Ho] to speed up the process, and may be used on any arbitrary signal. Wickerhauser [W1],[W2] has also developed and applied fast wavelet algorithms for acoustic and multimedia signal processing.

We apologize for the incompleteness of our discussion below of more recent acoustical studies using wavelet methods, as it is limited to work by researchers with whom we have had contact. Dorize, Gram-Hansen, Upton, and Daimon [DG],[U1] compared the capabilities of the short-time Fourier Transform (FT), Wigner-Ville distribution, and Gabor WT for analyzing noises such as car door slams. Isei and Kunimatsu used the WT package developed by Upton et al. to study blasting noises from underground explosions [U2]. Kikuchi, Nakashizuka, H. Watanabe, S. Watanabe, and Tomizawa calculated Gaussian and Mexican Hat WT and phase space diagrams of vehicle engine sounds to successfully detect detonations *"over some dilation scales and during a particular period. Detonations create strong pressure waves and can destroy an engine body if they* (occur frequently, particularly in succession. Detection and control of the phenomena is) *critically important for ignition advancement control* [Ki]." Nakashizuka, Kikuchi, Makino and Ishii have also studied the compression of ECG data using the wavelet zero-crossing representation [N]. Lee examined the compression and analysis of chirp signals. Rapid changes in chirps make them very difficult to compress, restore, and analyze by means of conventional methods. He found that *"the Gabor transform cannot separate three component signals, while the adaptive chirplet transform can* [L]." Tewfik, Sinha, and Jorgensen, [SiT],[TSJ] developed an audio synthesis/coding method based on an optimization of the WT of a signal.

A variety of commercial and public domain wavelet software is now available for specialized uses. Wickerhauser developed a software package for acoustical signal processing that handles denoising, compression, and parameter reduction for recognition as well as time-frequency analysis with the best-basis wavelet and Malvar transforms [W2],[CMW]. Cody presents an algorithm and code for the Fast WT and compares it with the Fast FT [Cd1],[Cd2]. AWARE markets the *UltraWave Explorer.* A system by Bruel and Kjaer analyzes nonstationary signals by using a variety of transforms, one of which is the Gabor WT [BK]. Donoho [Do] and Sakakibara [Sk] have made wavelet based, signal de-noising packages that run on a PC.

Work on wavelet analysis of speech signals has become directed towards more specific applications. Liénard and Alessandro [LA] studied wavelets and the granular analysis of speech events. Irino and Kawahara [I],[Kw] developed an algorithm to reconstruct a signal from an analyzing wavelet based on the impulse response of an auditory peripheral model. The algorithm was used to study the time-scale modification of speech. Takafumi Sakamoto and Tominaga have been studying speaker recognition using wavelets [SaT]. Gram-Hansen and Dorize [DG] compared short-time Fourier and Gaussian wavelet sonograms of speech signals: "*The WT with a 1/12-octave wavelet turned out to be favourable for analysis of a speech signal, since both transients and harmonic components appear clearly in the same representation* [DG]." Cairns [Ca] at Duke University has been investigating word boundary identification. Tan, Lang, Schroder, Spray and Dermody [T] have been studying the identification of the four major categories of speech — voiced speech, plosives, fricatives, and silence. They propose a segmentation scheme for application to hearing aid devices. Preliminary results obtained by using the algorithm are very good.

3. Our Experiments

The development of accurate and reliable speech recognition systems has been and remains a challenging task. An overview of the associated difficulties and Fourier-based approaches to tackling them is given by Picone [P].

The WT appears to be a good alternative to the FT, because of a constant frequency-bandwidth ratio [T]. This section describes our experiments on speech event analysis, which served as a basis for work in wavelet based speech recognition [SN]. We note that closely related work by Tan, Lang, Schroder, Spray, and Dermody [T] came to our attention as we were writing up our results. Our findings and theirs are consistent, although the data used by Tan et al. are much cleaner (i.e. noise-free) than ours. The application targets, however, have different requirements; we are looking for an alternative approach to be used in speech command/recognition systems, and they are seeking to develop improved hearing aid devices.

In our experiments we considered seven types of wavelets: the Daubechies [D1], three types of splines [MH],[Ch2], Gabor [G],[Ch2], chirp [I],[L] and the Mexican hat [D3]. WT sonograms were generated for the isolated syllables *aba, ada, aga, aka, apa,* and *ata* at a 10–KHz sampling rate, and for the phrases *ta* in the words ki*ta*mi, *ta*kefumi, rikuzen*ta*kada, *ta*keo, and ohmu*ta* and *ka* in the words *ka*zo, gushi*ka*wa, rikuzen*ta*kada, *ka*mo, and mo*ka* at a 20–KHz sampling rate. We calculated the transforms for 6 octaves with 12 half-steps per octave, using all types of wavelets except for the Daubechies, whose fractal nature was not amenable to the half-step calculations. The locations of the burst signals can clearly be seen for the syllables *aga* and *aka*, but less so for *aba, ada, apa* and *ata.* A finer sampling rate would have given a more pronounced mark for the latter four. When we doubled the sampling rate, the bursts for both *ka* and *ta* could clearly be seen in the complete words and phrases listed above. Results from further studies using the English words *stop* and *call up* spoken by a British male sampled at a 20–KHz rate are given in the figures below; clearly they varied depending on the wavelet type. The wavelet shape is more important than the mathematical definition in determining the sonogram features. Although many more samples must be examined before any definite statements can be made, we noticed the following trends. The coarse nature of the Daubechies diagram did not allow for any meaningful analysis. Splines and the Mexican hat appear to be the best at identifying the locations of bursts. Sonograms for the simple polynomial and quadratic and cubic cardinal splines, as well as for the Mexican hat, show a similar overall pattern. For *stop*, the transform values are small for the fricative noise *ssss.* The Gabor and Chirp sonograms show markedly different patterns from those of splines and the Mexican hat. The fricative noise yields high transform values.

We are currently shifting from speech event studies to assessing whether WT-based speech recognition is a realistic possibility. Very early results from phoneme recogition experiments using 5,240 Japanese words (male voice) from the ATR database MAU indicate a slightly worse rate for WT methods; however, the algorithm has been tuned for FT methods [SN]. The extent of overlap in the error sets from the WT and FT methods must be determined to assess the usability of WTs in speech and/or phoneme recognition systems. Furthermore, modification of the current algorithm for FT methods must be made or a WT specific algorithm developed.

References

[BK] *Bruel & Kjaer's WT9362*, Matsushita-Intertechno catalog.

[Ca] Doug Cairns, *private communication on M.A. work,* 1993.

[Cb] Combes et al., eds., *Wavelets*, Springer 1989.

[Cd1] Cody, *Dr. Dobb's Journal*, Apr. 1992, 16-28, 100-103.

[Cd2] Cody, *Dr. Dobb's Journal*, Apr. 1993, 44-54, 80-91.

[Ch1] Chui, ed., *Wavelets*, Academic 1992.

[Ch2] Chui, ed., *An Intro. to Wavelets*, Academic 1992.

[CMW] Coifman, Meyer, Wickerhauser, 153-178 in [R].

[D1] Daubechies, *Comm. Pure Appl. Math.*, XLI, 1988, 909-996.

[D2] Daubechies, *IEEE Trans. Inf. Th.*, 36, 1990, 961-1005.

[D3] Daubechies, *Ten Lectures on Wavelets*, SIAM 1992.

[DG] Dorize, Gram-Hansen, 77-92 in [My2].

[Do] Donoho, *TeachWave0550*, ftp wavelab@playfair.stanford.edu.

[FHV] Farge, Hunt, Vassilicos, *Wavelets, Fractals & FTs*, Oxford, 1993.

[G] Grossman, Kronland-M., Morlet, 2-20 in [Cb].

[Ho] Holschneider et al., *the Algorithme à Trous*, CNRS Report, May 1988.

[Hu] Hunt, Kevlahan, Vassilicos, Farge, 1-38 in [FHV].

[I] Irino, *Tech. Rep. IEICE*, Sp92-81, DSP92-66, 59-66. *

[Ki] Kikuchi et al., *IEICE Trans. Fund.*, Vol. E75-A, No. 8, Aug. 1992.

[Kw] Kawahara, *JSIAM Wavelet Seminar I*, Ch.4 (text) 1991. *

[KMG] Kronland-M., et al., *Int'l J. Pat. Recog.*, 1(2) 1987, 273-302.

[KS] Kobayashi, M. Sakamoto, *JSIAM Wavelet Seminar II*, Ch.3 (text) 1993.

[L] Daniel Lee, *Wavelets, Course Notes*, HP Labs Japan, May 1992.

[LA] Liénard, Alessandro 158-163 in [Ch1].

[MH] Mallat, Hwang, *IEEE Trans. Inf. Th.*, 38(2) March 1992, 617-643.

[N] Nakashizuka et al., *IEICE Tech. Rep.*, CAS93-63 (1993-09).

[P] Picone, *Proc. IEEE*, 79(4), April 1991, 1214-1246.

[R] Ruskai et al., eds., *Wavelets*, Jones and Bartlett 1992.

[SaT] T. Sakamoto, Tominaga, *Proc. IEICE 1993*, A-221 Spring. *

[SiT] Sinha, Tewfik, *IEEE Sig. Proc.*, June 1993, submitted.

[Sk1] Sakakibara, *SplineWavelet.m*, demo, Messina, Italy, Oct. 1993.

[SN] M. Sakamoto, Nishimura, *Spring Meet. Acoust. Soc. Japan 1994*, (poster).

[T] Tan et al., *SPIE'94 Orlando*, preprint.

[TSJ] Tewfik et al., *IEEE Trans. Inf. Th.*, 38(2) March 1992, 747-765.

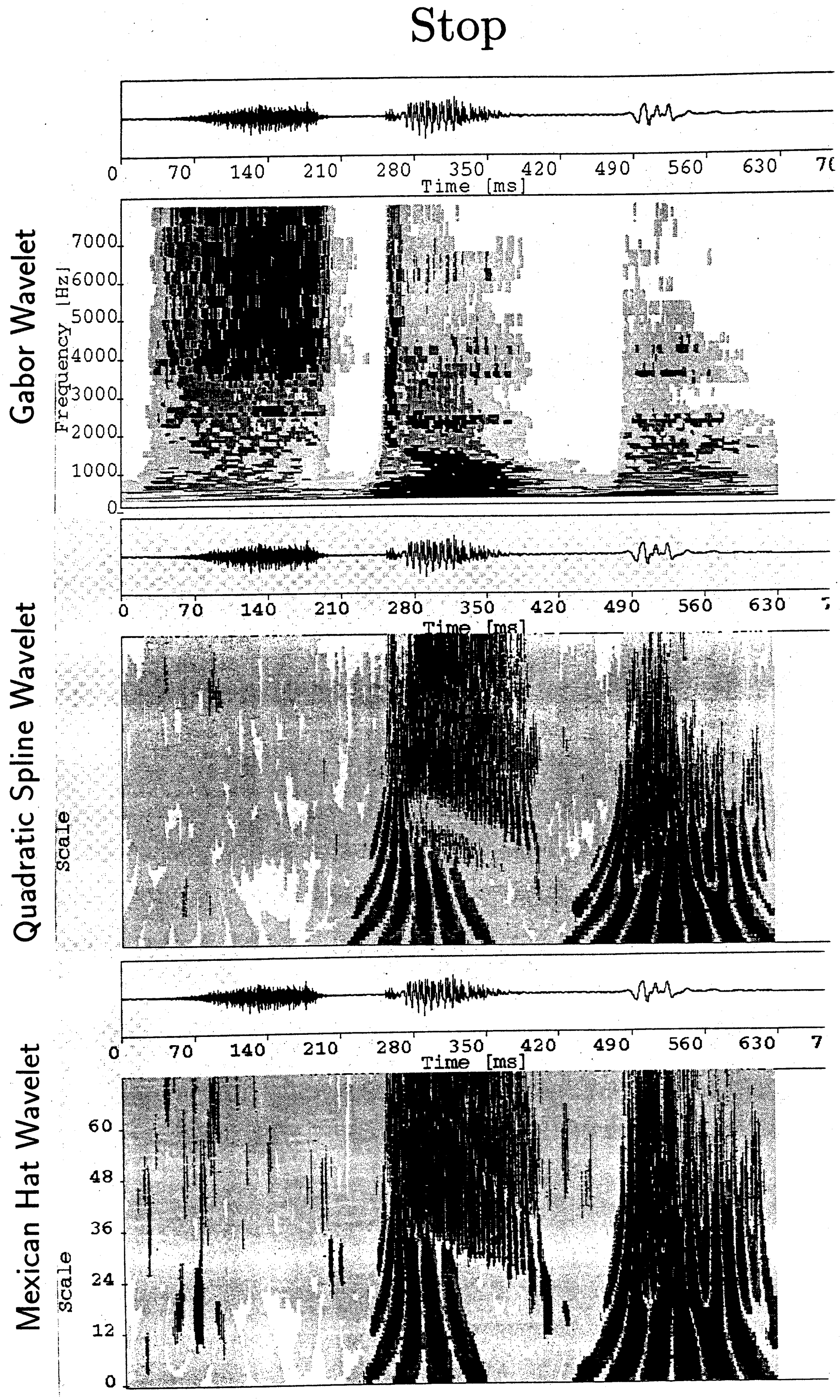
[U1] Upton, Daimon, *Jidousha Gijutsu Kai*, 921, 1992-5 (921013) 49-52.

[U2] Upton et al., *Proc. Acoust. Soc. Japan 1992*, Fall(1-2-14) and (1-2-15).

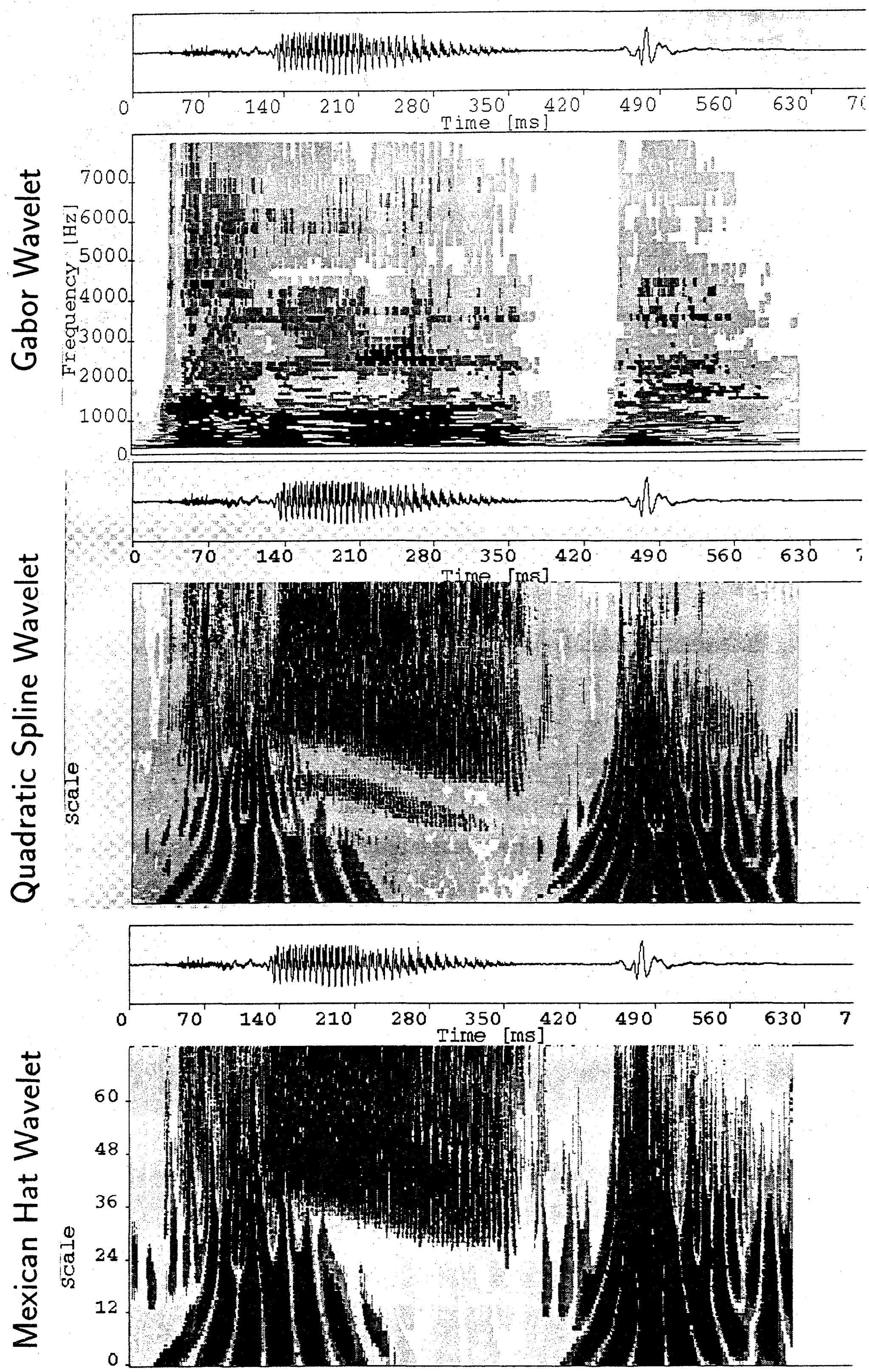
[W1] Wickerhauser, 679-700 in [Ch1].

[W2] Wickerhauser, *Lectures on Wavelet Packet Algorithms*, preprint, 1991.

* in Japanese


Stop
0
70
140
210
280
350
420
490
560
630
Time [ms]
Gabor Wavelet
Frequency [Hz]
7000
6000
5000
4000
3000
2000
1000
0
Quadratic Spline Wavelet
Scale
Mexican Hat Wavelet
Scale
60
48
36
24
12
0

Call Up



0 70 140 210 280 350 420 490 560 630
Time [ms]
Gabor Wavelet
Frequency [Hz]
7000
6000
5000
4000
3000
2000
1000
0
Quadratic Spline Wavelet
Scale
Mexican Hat Wavelet
60
48
36
24
12