Comm. Math. Sci.
Vol. 3, No. 2, pp. 251–259

AN ORTHOGONAL DISCRETE AUDITORY TRANSFORM[*]

JACK XIN[†] AND YINGYONG QI[‡]

Abstract. An orthogonal discrete auditory transform (ODAT) from sound signal to spectrum is constructed by combining the auditory spreading matrix of Schroeder *et. al.* and the time one map of a discrete nonlocal Schrödinger equation. Thanks to the dispersive smoothing property of the Schrödinger evolution, ODAT spectrum is a smoother than that of the discrete Fourier transform (DFT) consistent with human audition. ODAT and DFT are compared in signal denoising tests with spectral thresholding method. The signals are noisy speech segments. ODAT outperforms DFT in signal to noise ratio (SNR) when the noise level is relatively high.

Key words. Orthogonal discrete auditory transform, Schrödinger equation.

AMS subject classifications. 94A12, 94A14, 65T99.

1. Introduction

Acoustic signal processing can benefit significantly from utilizing properties of human audition, e.g. perceptual coding in MP3 technology of music compression [12, 13]. In [15], an invertible discrete auditory transform (DAT) is formulated by the present authors to map sound signal to auditory spectrum. DAT is more adapted to the spectral features of the ear than Fourier transform. It incorporates the auditory spreading functions of Schroeder, Atal and Hall [13] to achieve smoother spectrum than that of the discrete Fourier transform (DFT), and a better performance in denoising under spectral thresholding. However, such a transform has redundancy in the sense that the image of a discrete vector lies in a higher dimensional space, similar to tight frames in wavelets [2, 14].

In this paper, such redundancy is removed by constructing an orthogonal (unitary) matrix with spreading property over frequency bands comparable to the critical bands in hearing. Critical bands (Table 10.1, p 309, [12]) characterize the bandwidth of the human auditory filter. The auditory orthogonal matrix is obtained from the time one map of a spatially discrete nonlocal Schrödinger equation. Let space time complex function $u=u(x,t)$ be the solution of the Schrödinger equation, the time one map goes from $u(x,0)$ to $u(x,1)$. The Schrödinger equation conserves the L^2 norm or Euclidean length, implying the orthogonality of the time one map. On the other hand, the dispersive smoothing nature of the Schrödinger evolution leads to the spreading property of the time one map. The auditory functions of Schroeder, Atal and Hall [13] appear as a nonlocal potential in the Schrödinger equation. As a result, a class of orthogonal discrete auditory transforms (ODAT) are generated. In searching for ODATs, an alternative method based on the dilation equation of wavelets is also found, however, such an approach turns out to be too rigid to accomodate auditory properties, e.g. spectral spreading across critical bands.

The paper is organized as follows. In section 2, the ODAT is derived from the general DAT [15], and the ODAT construction is presented based on the discrete Schrödinger equation. A specific ODAT is given by inserting the auditory spreading

[*] Received: February 12, 2005; accepted (in revised version): April 16, 2005. Communicated by Anna Gilbert.

[†] Department of Mathematics and ICES, University of Texas at Austin, Austin, TX 78712, USA (jxin@math.utexas.edu).

[‡] Qualcomm Inc, 5775 Morehouse Drive, San Diego, CA 92121, USA.

functions in [13]. In section 3, auditory spectra of a two tone signal (with frequencies across a critical band) and of a vowel segment are compared with their DFT counterparts to illustrate the auditory spectral spreading. Denoising with spectral thresholding is performed on voiced and unvoiced speech segments. ODAT is found to increase signal to noise ratio beyond DFT when the noise content is relatively high. Concluding remarks are made in section 4.

2. ODAT and Schrödinger

Let $s=(s_0,\cdots,s_{N-1})$ be a discrete real signal, the discrete Fourier transform (DFT) is [1]:

$$\hat{s}_k=\sum_{n=0}^{N-1} s_n\, e^{-i(2\pi nk/N)}. \tag{2.1}$$

The general discrete auditory transform (DAT) is [15]:

$$S_{j,m}\equiv\sum_{l=0}^{N-1} s_l\, K_{j-l,m}, \tag{2.2}$$

where the double indexed kernel function is:

$$K_{l,m}=\sum_{n=0}^{N-1} X_{m,n}\, e^{i(2\pi ln/N)}; \tag{2.3}$$

and the matrix $X_{m,n}$ has square sum equal to one in m:

$$\sum_{m=0}^{M-1} |X_{m,n}|^2=1,\ \forall n. \tag{2.4}$$

Here M is on the order of N.

DFT is recovered from DAT by setting $j=0$, $M=N$, and $X_{m,n}$ the $N\times N$ identity matrix. In case that $X_{m,n}$ is a nontrivial orthogonal matrix, let us still set $j=0$ in (2.2) to find:

$$\begin{aligned} S_{0,m}\equiv S_m &=\sum_{l=0}^{N-1} s_l \sum_{n=0}^{N-1} X_{m,n}\, e^{-i(2\pi ln/N)} \\ &=\sum_{n=0}^{N-1} X_{m,n}\, \hat{s}_n. \end{aligned} \tag{2.5}$$

The mapping from s_l to S_m is orthogonal. The problem reduces to finding an orthogonal matrix $(X_{m,n})$ with auditory features.

Such a matrix acts on complex numbers $\hat{s}_n$ (except the modes $n=0$ and $n=N/2$, so called DC and Nyquist modes). Let us consider the time one map of the following spatially discrete Schrödinger equation:

$$i\,u_{n,t}=\sigma_1\,(u_{n+1}-2u_n+u_{n-1})+\sigma_2\sum_{m=1}^{N_h} V_{m,n}\,u_m, \tag{2.6}$$

where σ_1 and σ_2 are positive real numbers, $N_h=N/2-1$, $(V_{m,n})$ is a symmetric $N_h\times N_h$ matrix to carry certain auditory information of the ear. For simplicity, Dirichlet boundary condition is imposed for the evolution of equation (2.6). The discrete equations (2.6) can be cast in the matrix form:

$$iU_t = (\sigma_1 A + \sigma_2 B)U, \tag{2.7}$$

where $U = (u_1, u_2, \cdots, u_{N_h})^T$, A the tridiagonal matrix (-2 on the diagonal, 1 on the two off-diagonals), B the real symmetric matrix with entry $V_{m,n}$ at (m,n). The time one map of (2.7), denoted by T_w, is simply $\exp\{i(\sigma_1 A + \sigma_2 B)\}$ which is clearly orthogonal, $T_w T_w^{'} = Id_{N_h}$, where the prime denotes the conjugate transpose.

The matrix B is built from auditory spreading functions [13] denoted by $S(b(f_m), b(f_n))$, where f_m is the frequency to spread from, f_n is the frequency to spread to, and b is the standard mapping from Hertz (Hz) to Bark scale [7]. The functional form of $S(\cdot,\cdot)$ is given in [13]. Define $V_{m,n} = 1/2 \cdot (S(b(f_m), b(f_n)) + S(b(f_n), b(f_m)))$, so B is the symmetric part of the matrix $(S(b(f_m), b(f_n)))$. Numerical results based on this choice of B will be reported in the next section.

The matrix $X = (X_{m,n})$ takes the block diagonal form:

$$X = \text{diag}\{1, T_w, 1, \widehat{T_w}^*\}, \tag{2.8}$$

where the tilde denotes the reverse permutation of columns of T_w so that the spreading occurs symmetrically on the DFT components ($\hat{s}_l$, $N_h + 2 \leq l \leq N-1$) to preserve the conjugate symmetry of the spectrum. The matrix X is clearly orthogonal and leaves invariant the DC and Nyquist modes. The ODAT matrix is the product of X and DFT matrix.

The continuum version of (2.6) is:

$$iu_t = \Delta_x u + V(x) * u, \; x \in R^n, \; n \geq 1, \tag{2.9}$$

where $*$ is convolution, $V(x)$ is real and even. The L^2 norm of u is conserved in time. Schrödinger equations analogous to (2.9) have been much studied regarding smoothing (scattering) properties and derivation from particle dynamics, [5, 6, 8, 10, 11] among others. When the convolution term is cubically nonlinear in u, the equation is known as Schrödinger-Hartree [5, 8, 6]. In [8, 10], the smoothing and spreading property is measured in the weighted norm $\|\psi\|_{m,s} = \|(1+|x|^2)^{s/2}(1-\Delta)^{m/2}\psi\|_2$, Δ the spatial Laplacian. Solutions at time $t \neq 0$ satisfy the bound:

$$\|u(t)\|_{1,-1} \leq C(\|u(0)\|_{0,1})(|t| + |t|^{-1}). \tag{2.10}$$

We shall see in the next section that the time one Schrödinger map T_w inherits the smoothing and spreading property of the continuum case.

3. Numerical Tests

The computation is carried out in Matlab, with ODAT parameters $(\sigma_1, \sigma_2) = (0.6, 0.04)$. Discrete signal (frame) length $N = 256$. First consider a two tone signal consisting of sinusoids of frequencies 3 kHz (kilo-Hertz) and 4.3 kHz with identical amplitudes. The two frequency values span a critical band. Figure 3.1 compares the ODAT (dashed) and DFT (solid) log-magnitude spectra. The ODAT spectral peak regions are lower and wider than DFT's. Also there is more spreading in ODAT spectrum towards higher frequency, consistent with upward masking property of human ear [18]. This can be explained by the weighted norm estimate (2.10), where large x

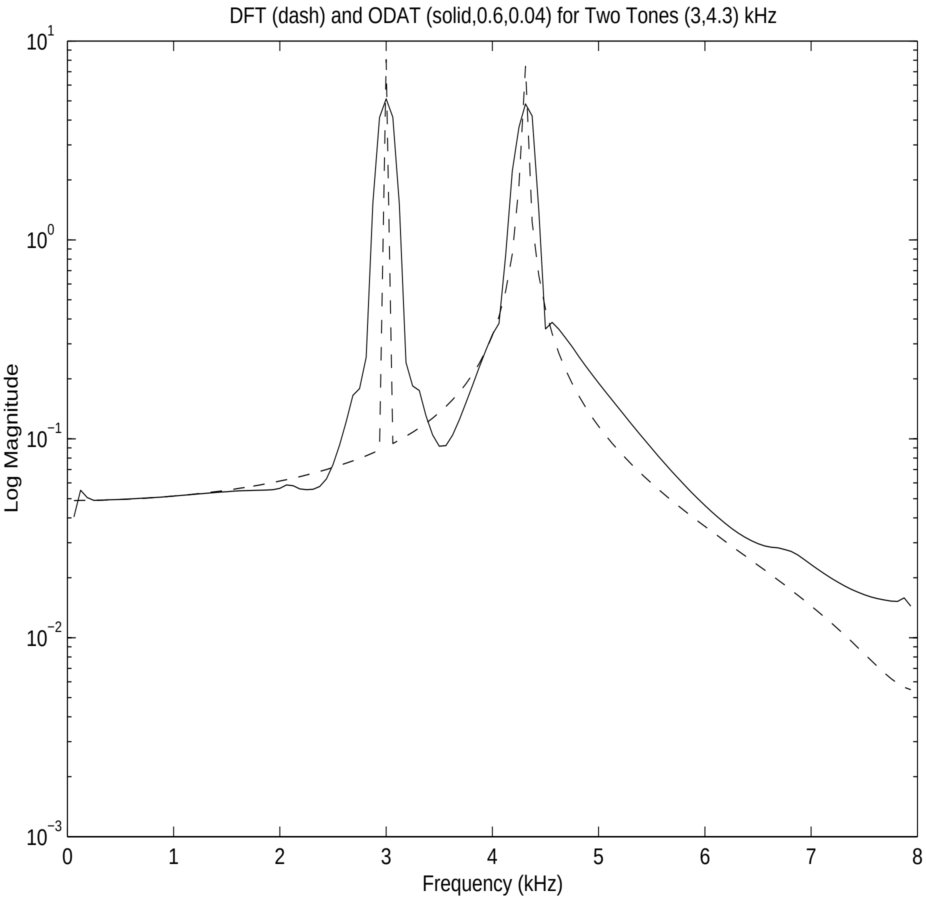

FIG. 3.1. *Comparison of ODAT spectrum (solid) and DFT spectrum (dash) of a two tone signal of frequencies (3,4.3) kHz and identical amplitudes. ODAT's spectral spreading appears near the peak areas and towards the higher frequencies. ODAT parameters* $(\sigma_1, \sigma_2) = (0.6, 0.04)$.

corresponds to large frequency. Figure 3.2 shows ODAT and DFT spectra of a vowel segment containing multiple harmonics, where spectral smoothing is observed again.

ODAT and DFT were used to denoise speech signals via the thresholding method in the transformed domain [15]. The aim is to improve the signal-to-noise ratio (SNR) of noisy speech. The premise of the method is that low level components in the transformed domain are more likely to be noise than signal plus noise. So thresholding could improve the overall SNR of the signal. The simple thresholding method serves to illustrate the difference between ODAT and DFT in signal processing. A vowel and a consonant speech segments were selected, each segment has 512 data points. Noisy speech was created by adding Gaussian noise to the selected segments. The level of noise was set to produce the SNR ranging from -12 decible (dB) to +12 dB with a 3 dB step size. ODAT and DFT were applied to the noisy speech signals. The magnitude of transformed components were then compared to a threshold. All components with magnitude smaller than the threshold were ignored for the reconstruction of the signal.

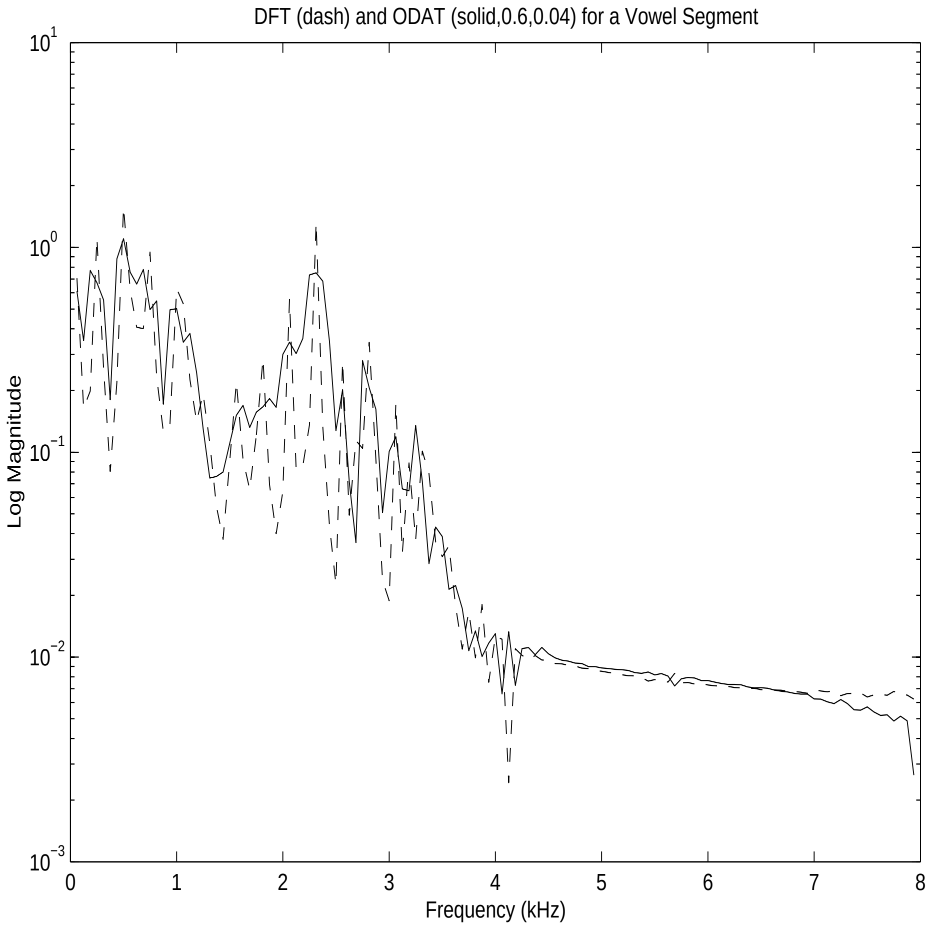

FIG. 3.2. *Comparison of ODAT spectrum (solid) and DFT spectrum (dash) of a vowel segment. ODAT parameters* $(\sigma_1, \sigma_2) = (0.6, 0.04)$.

The threshold was computed as the average of the DFT magnitude spectrum. Signal was reconstructed by the inverse ODAT and DFT, respectively. The SNRs of the reconstructed vowel signal is plotted vs. input SNRs in Figure 3.3. The SNRs of the reconstructed consonant signal is plotted vs. input SNRs in Figure 3.4. We see that ODAT (solid) improves over DFT (dashdot) in terms of SNR when the noise level is relatively high, particularly in case of consonants which resemble noise more than the vowels.

The noise-reduction advantage can be attributed to the spectral spreading property of ODAT. In Figure 3.5, FFT and ODAT spectrograms are shown for the speech utterance "fairy tales should be fun to write" from the TIMIT database. The spectrograms illustrate the temporal variation of sound spectra. The ODFT spectrogram is visibly smoother. The spectrograms are obtained by dividing the sentence into short-time frames, then performing FFT and ODAT to generate spectra of each frame. The redundant DAT [15] is quite similar in terms of spectral smoothing property. Redun-

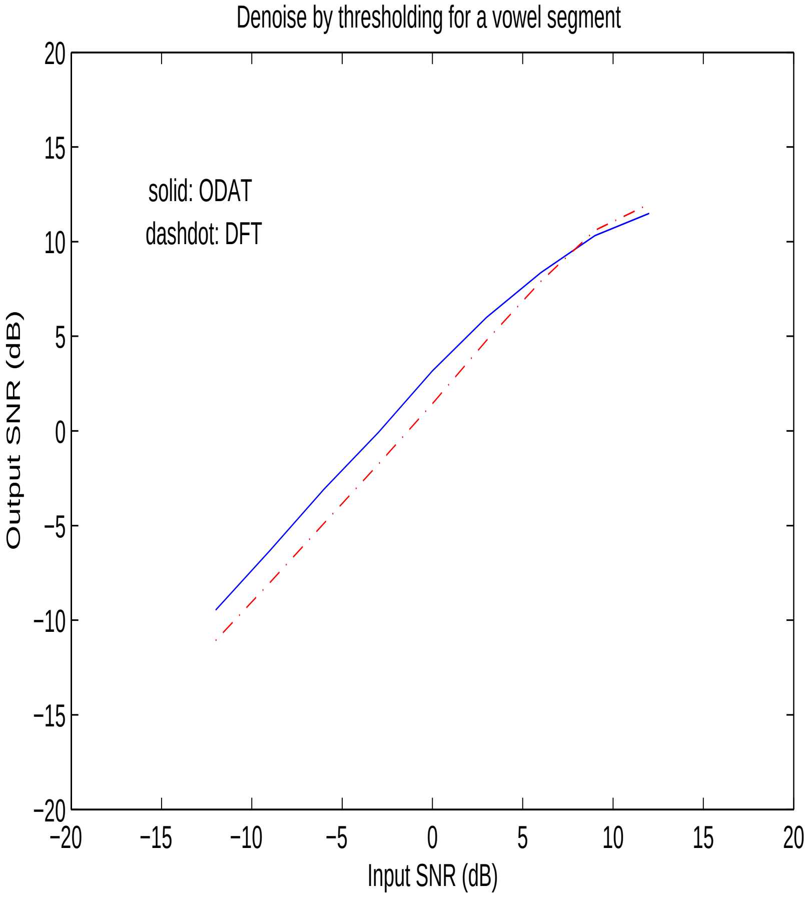

Fig. 3.3. *Comparison of ODAT (solid) and DFT (dashdot) denoising by spectral thresholding for a vowel segment. Spectral spreading property of ODAT helps to increase signal content when noise level is relatively high, e.g. input SNR below 7 decible (dB).*

dancy however renders more modes in the transformed domain, and was observed to provide more SNR gain in denoising tests. It is interesting to find out how to enhance the amount of smoothing for ODAT in future work.

It is rewarding to investigate how well a nonlinear nonlocal Schrödinger equation can model the ear's nonlinear responses. Ear's nonlinearities are nonlinear and nonlocal in nature, and the physiological models are dispersive, nonlinear, and nonlocal [9, 4, 3, 16, 17].

4. Concluding Remarks

Orthogonal discrete auditory transforms (ODAT) are introduced based on nonlocal spatially discrete Schrödinger equations. Dispersive smoothing, mass conservation, and robustness of the Schrödinger equation allows one to inject auditory knowledge in the transform while preserving orthogonality. Numerical tests on two tone and speech

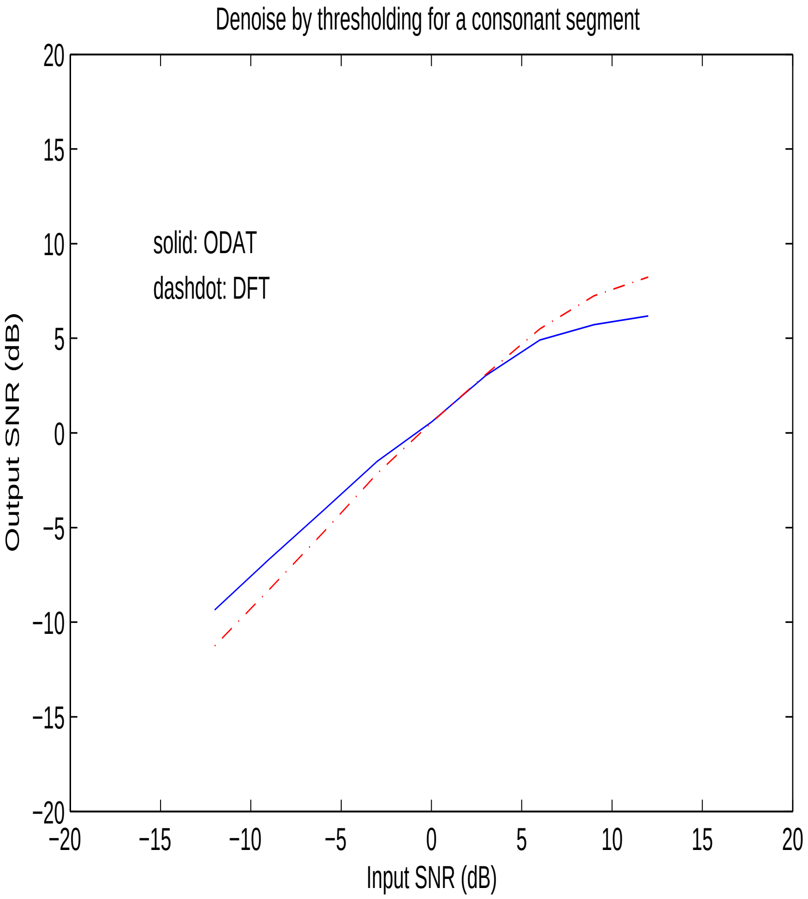

Fig. 3.4. *Comparison of ODAT (solid) and DFT (dashdot) denoising by spectral thresholding for a vowel segment. Spectral spreading property of ODAT helps to increase signal content when the noise level is relatively high, e.g. input SNR below zero decible (dB).*

segments demonstrate the spectral spreading property of ODAT and advantage in denoising. Future work will explore efficient ways to enhance spectral spreading for ODATs and more complex signal processing applications.

Acknowledgements. This work was supported in part by NSF grant ITR-0219004, and a Fellowship from the John Simon Guggenheim Memorial Foundation. We thank Prof. G. Papanicolaou and Prof. H-T Yau for helpful conversations. Results in the paper were presented at the IPAM workshop on "Mathematics of the Ear and Sound Signal Processing", UCLA, Jan 31 – Feb 2, 2005. We thank many workshop speakers and participants for their comments and interest.

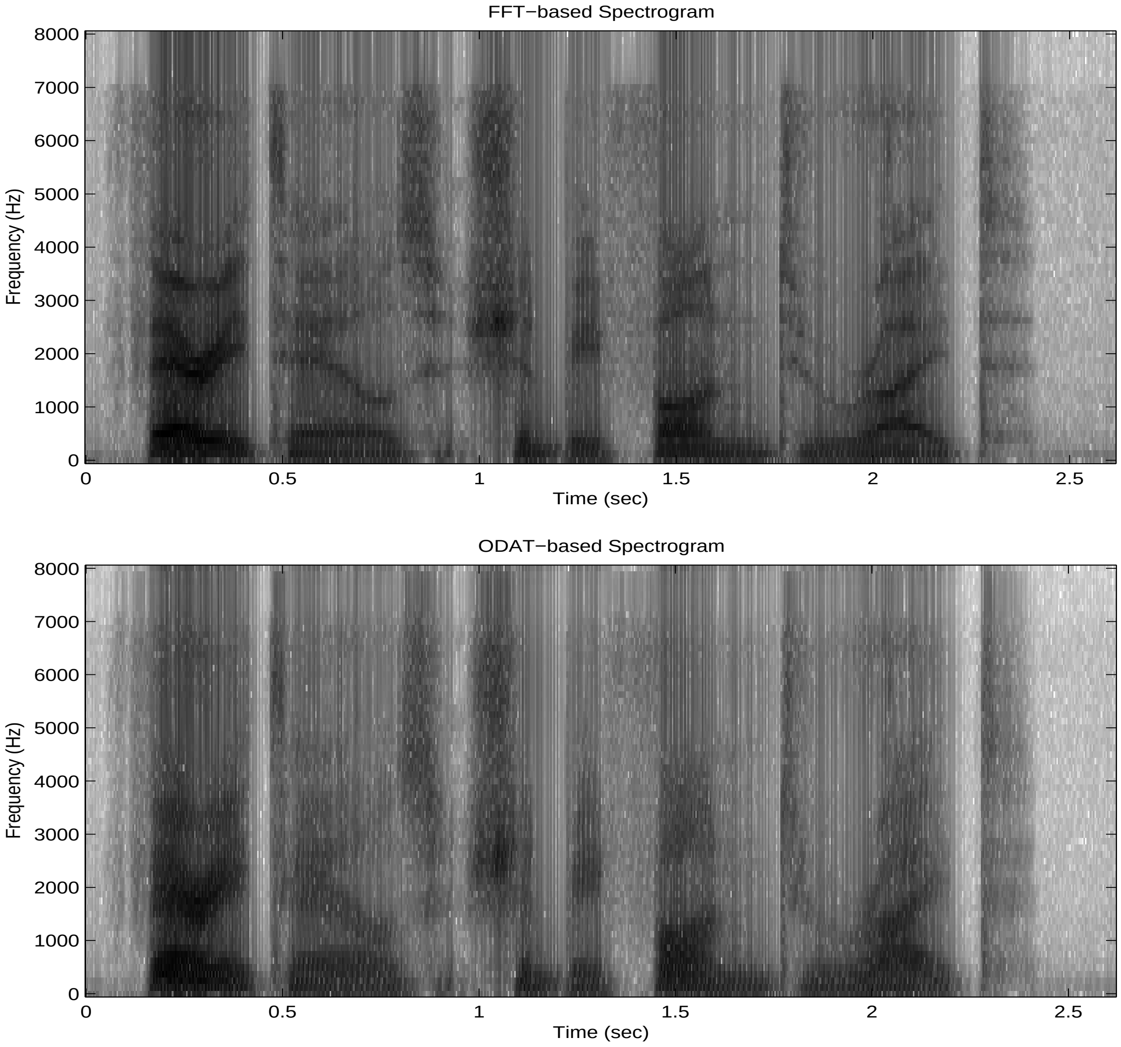

Fig. 3.5. *The FFT (top) and ODAT (bottom) spectrograms of the speech utterance "fairy tales should be fun to write".*

REFERENCES

[1] P. Brémaud, *Mathematical Principles of Signal Processing: Fourier and Wavelet Analysis*, Springer-Verlag, 2002.

[2] I. Debauchies, *Ten Lectures on Wavelets*, CMS-NSF Regional Conference in Applied Mathematics, SIAM, Philadelphia, 1992.

[3] E. de Boer and A. L. Nuttall, *Properties of amplifying elements in the cochlea*, in "Biophysics of the Cochlea: From Molecules to Models", A. W. Gummer, ed. Proc. Internat. Symp., Titisee, Germany, 2002.

[4] L. Deng, *Processing of acoustic signals in a cochlear model incorporating laterally suppressive elements*, Neural Networks, 5, 1, 19-34, 1992.

[5] A. Elgart, L. Erdös, B. Schlein and H-T Yau, *Nonlinear Hartree equation as the mean field*

limit of weakly coupled fermions, J. Math. Pures Appl., 83, 1241-1273, 2004.

[6] J. Ginibre, *A Remark on some papers by N. Hayashi and T. Ozawa*, J. Func Analysis, 85, 349-352, 1989.

[7] W. M. Hartmann, *Signals, Sound, and Sensation*, Springer, 251-254, 2000.

[8] N. Hayashi and T. Ozawa, *Smoothing effect for some Schrödinger equations*, J. Functional Analysis, 85, 307-348, 1989.

[9] Y. Jau, C. D. Geisler, *Results from a cochlear model utilizing longitudinal coupling*, in "Mechanics of Hearing", E. de Boer and M. Viergever eds, 169-176, 1983.

[10] A. Jensen, *Commutator methods and a smoothing property of the Schrödinger evolution group*, Math Zeitschrift, 191, 53-59, 1986.

[11] L. Kapitanski, Y. Safarov, *Dispersive smoothing for Schrödinger equations*, Math Res. Letters, 3, 77-91, 1996.

[12] K. Pohlmann, *Principles of Digital Audio*, 4th edition, McGraw-Hill Video/Audio Professional, 2000.

[13] M. R. Schroeder, B. S. Atal and J. L. Hall, *Optimizing digital speech coders by exploiting properties of the human ear*, Journal Acoust. Soc. America, 66, 6, 1647-1652, 1979.

[14] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wesley-Cambridge Press, 1997.

[15] J. Xin and Y. Qi, *An invertible discrete auditory transform*, Comm. Math. Sci, 3, 1, 47-56, 2005.

[16] J. Xin and Y. Qi, *Global well-posedness and multi-tone solutions of a class of nonlinear nonlocal cochlear models in hearing*, Nonlinearity 17, 711-728, 2004.

[17] J. Xin, Y. Qi, and L. Deng, *Time domain computation of a nonlinear nonlocal cochlear model with applications to multitone interaction in hearing*, Comm. Math. Sci., 1, 2, 211-227, 2003.

[18] E. Zwicker, H. Fastl, *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences, 22, 2, 1999.