

# CONCEFT: CONCENTRATION OF FREQUENCY AND TIME VIA A MULTI-TAPERED SYNCHROSQUEEZING TRANSFORM

INGRID DAUBECHIES, YI (GRACE) WANG, AND HAU-TIENG WU

## ESM-1. INTRODUCTION

These are the *Electronic Supplementary Materials* for the paper CONCEFT: CONCENTRATION OF FREQUENCY AND TIME VIA A MULTI-TAPERED SYNCHROSQUEEZING TRANSFORM. They contain, in particular, precise mathematical definitions, theorem statements and proofs that complement the more general exposition in the main body of the paper, as well as details about the numerical examples and additional examples. For the convenience of the reader, the organization into sections follows that of the paper; for instance, material in Section ESM-3 complements Section 3 in the main paper.

## ESM-2. THE CONCEFT ALGORITHM: SEVERAL REMARKS

(1) As described in Section 2 in the main paper, the SST-steps in ConceFT involve the computation of a partial derivative,  $\partial_b W_f^{(\psi)}(a, b)$  with respect to the localization parameter  $b$  of  $W_f^{(\psi)}(a, b)$ . In practice, one has  $W_f^{(\psi)}(a, b)$  only for discrete (as opposed to continuous) values of  $a$  and  $b$ , and partial differentiation would be approximated by a differentiating scheme. This can cause stability issues when  $f$  is noisy. Using the definition of  $W_f^{(\psi)}(a, b)$  as the inner product of  $f$  with  $|a|^{-1/2}\psi\left(\frac{\cdot-b}{a}\right)$ , one can compute  $\partial_b W_f(a, b)$  via the wavelet transform of  $f$  with respect to the wavelet  $\psi'$  using  $\partial_b W_f^{(\psi)}(a, b) = -W_f^{(\psi')}(a, b)$ , typically this makes the computation more stable than simple numerical differencing.

(2) The ConceFT algorithm consists in taking the average of many nonlinear SST estimates of the tvPS, each of which results from a wavelet transform with respect to a randomly picked reference wavelet; for each individual transform the corresponding reassignment is computed and carried out to find that individual SST. An alternative to the individual SSTs would be to define one “master” reassignment rule, as follows. From the collection of  $W_f^{(\psi_j)}$ ,  $j = 1, \dots, J$ , we could estimate  $\Omega^{(\Gamma)}(a, b)$  as the value of  $\xi$  for which the vector

$$\mathbf{w}(a, b) = [W_f^{(1)}(a, b), \partial_b W_f^{(1)}, \dots, W_f^{(J)}(a, b), \partial_b W_f^{(J)}(a, b)] \in \mathbb{C}^{2J}$$

is most “aligned” with the vector  $u(a, b, \xi)$

$$\mathbf{u}(a, b, \xi) = [W_{e^{i2\pi\xi t}}^{(1)}(a, b), \partial_b W_{e^{i2\pi\xi t}}^{(1)}, \dots, W_{e^{i2\pi\xi t}}^{(J)}(a, b), \partial_b W_{e^{i2\pi\xi t}}^{(J)}(a, b)] \in \mathbb{C}^{2J}.$$

In other words, the reassignment rule would become

$$\Omega(a, b) := \operatorname{argmax}_{\xi} \frac{|\langle \mathbf{w}(a, b), \mathbf{u}(a, b, \xi) \rangle|}{\|\mathbf{w}(a, b)\| \|\mathbf{u}(a, b, \xi)\|}.$$

Although numerical experiments have shown this to be an interesting approach as well, we shall not pursue this in this paper.

(3) In most of our examples and figures, we concentrate on visualizing the location in the TF plane of the curves characterizing the different IMT components of the signals considered. However, we can also

use the tvPS constructed by ConceFT to estimate the different amplitudes, as follows. Each ConceFT tvPS is the average of many SSTs constructed in such a way that the integral (sum, in practice) over  $\xi$ , on an interval around  $\phi'_l(t)$ , approximates  $A_l(t) \cos(2\pi\phi_l(t))$  (see [4, 2]). It follows that one can use the ConceFT representation to first identify the  $\phi'_l(t)$  for all  $t$  (which can be done more stably with ConceFT, for large noise, than with simple SST of MTSST), and then integrate  $\tilde{S}_Y(t, \xi)$  with respect to  $\xi$  in an appropriate interval around  $\phi'_l(t)$ , to recover  $A_l(t)$ .

### ESM-3. THEORETICAL RESULTS: MATHEMATICAL STATEMENTS AND PROOFS

The following is the mathematically precise definition of an *intrinsic-mode type* (IMT) function:

**Definition S.3.1.** Given  $\epsilon, c_1$  and  $c_2$  satisfying  $0 < \epsilon \ll 1, 0 < c_1 \leq c_2 < \infty$ , a function  $F(t)$  is said to be of type  $\mathcal{A}_\epsilon^{c_1, c_2}$  if it can be written as

$$F(t) = A(t) \cos(2\pi\varphi(t)),$$

where

$$(S.1) \quad \begin{cases} A \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \quad \varphi \in C^2(\mathbb{R}), \\ \inf_{t \in \mathbb{R}} A(t) > c_1, \quad \inf_{t \in \mathbb{R}} \varphi'(t) > c_1, \\ \sup_{t \in \mathbb{R}} A(t) \leq c_2, \quad \sup_{t \in \mathbb{R}} \varphi'(t) \leq c_2, \\ |A'(t)| \leq \epsilon \varphi'_\ell(t), \quad |\varphi''(t)| \leq \epsilon \varphi'(t) \quad \text{for all } t \in \mathbb{R}, \end{cases}$$

To model the oscillatory functions with different oscillatory modes, we also consider superpositions of IMT functions:

**Definition S.3.2.** Given  $\epsilon, c_1, c_2$  and  $d$  satisfying  $0 < \epsilon \ll 1, 0 < c_1 \leq c_2 < \infty, 0 < d < 1$ , a function  $G(t)$  is said to be of type  $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$  if it can be written as

$$(S.2) \quad G(t) = \sum_{\ell=1}^L F_\ell(t) \sum_{\ell=1}^L A_\ell(t) \cos(2\pi\varphi_\ell(t)),$$

where each  $F_\ell = A_\ell(\cdot) \cos(2\pi\varphi_\ell(\cdot))$  is of type  $\mathcal{A}_\epsilon^{c_1, c_2}$  and

$$(S.3) \quad \varphi'_{\ell+1}(t) - \varphi'_\ell(t) > d(\varphi'_{\ell+1}(t) + \varphi'_\ell(t))$$

for all  $\ell = 1, \dots, L-1$ .

Finally, we also consider the additive white Gaussian noise. Denote  $\mathcal{S}$  to be the Schwartz function space. Our model for the observed signal  $Y(t)$  is thus

$$(S.4) \quad Y(t) = \sum_{\ell=1}^L F_\ell(t) + \sigma \Phi(t) = \sum_{\ell=1}^L A_\ell(t) \cos(2\pi\varphi_\ell(t)) + \sigma \Phi(t),$$

where  $G = \sum_{\ell=1}^L F_\ell$  is of type  $\mathcal{A}_{\epsilon, d}^{c_1, c_2}$ ,  $\Phi$  is a Gaussian white noise so that the standard deviation of  $\Psi(\psi)$  is 1 for all  $\psi \in \mathcal{S}$  with norm 1, and  $\sigma > 0$  is the noise level;  $Y$  is a generalized random process, since by definition  $\sum_{\ell=1}^K A_\ell(t) \cos(2\pi\varphi_\ell(t))$  is a tempered distribution.

The  $J$  reference wavelets  $\psi_1, \dots, \psi_J$  are orthonormal, that is,  $\int \psi_i(x) \overline{\psi_j(x)} dx = \delta_{i,j}$ , where  $\delta_{i,j}$  is the Kronecker delta. For simplicity we assume that the  $\psi_j$  all have fast decay, that their Fourier transforms  $\widehat{\psi_j}$  are real functions with compact support, and  $\text{supp} \widehat{\psi_j} \subset [1 - \Delta_j, 1 + \Delta_j]$ , where  $0 < \Delta_j < 1$ .

We shall consider the continuous wavelet transforms of  $Y$  with respect to the  $\psi_j$ , and apply synchrosqueezing to them. For the  $F_\ell$ -components of  $Y$  we refer the reader to the detailed analysis in [4, 2]. In particular, we introduce the sets  $Z_\ell^{(j)}(b) = \left[ \frac{1-\Delta_j}{\varphi'_\ell(b)}, \frac{1+\Delta_j}{\varphi'_\ell(b)} \right]$ . If each  $\Delta_j$  satisfies  $\Delta_j \leq \frac{d}{1+d}$  for  $j = 1, \dots, J$  (which we shall assume for the remainder of this discussion), then one finds that, by the conditions on  $G$ , the sets  $Z_\ell^{(j)}(b)$  are disjoint. Moreover, the CWT  $W_G^{(\psi_j)}(a, b)$  is small except for those pairs  $(a, b)$  where  $a \in Z_\ell^{(j)}(b)$  for some  $\ell$ , and in that case  $\frac{-i\partial_b W_G^{(\psi_j)}(a, b)}{2\pi W_G^{(\psi_j)}(a, b)}$  is close to  $\varphi'_\ell(b)$ . (See Theorem 3.3 in [4].)

It will be convenient to use  $\underline{\Delta} := \min_{j=1}^J(\Delta_j)$ ,  $\bar{\Delta} := \max_{j=1}^J(\Delta_j)$ , and  $Z_\ell(b) = \left[ \frac{1-\underline{\Delta}}{\varphi'_\ell(b)}, \frac{1+\bar{\Delta}}{\varphi'_\ell(b)} \right]$ . Clearly  $Z_\ell(b) = \cap_{j=1}^J Z_\ell^{(j)}(b)$ .

As shown by the analysis in [4, 2], we have

$$W_G^{(\psi_j)}(a, b) = \begin{cases} e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a, b) + \epsilon_j(a, b) & \text{when } a \in Z_\ell^{(j)}(b) \text{ for some } \ell = 1, \dots, L \\ \epsilon_j(a, b) & \text{otherwise,} \end{cases}$$

where

$$(S.5) \quad Q_{j,\ell}(a, b) = A_\ell(b) \sqrt{a} \widehat{\psi_j}(a \varphi'_\ell(b)) \in \mathbb{R}$$

and  $\epsilon_j(a, b)$  is of order  $\bar{\epsilon} = \epsilon^{1/3}$  for all  $j = 1, \dots, J$ .

Adding also the noise, we have thus

$$W_Y^{(\psi_j)}(a, b) = \sum_{\ell=1}^L e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}}(b) + \epsilon_j(a, b) + \sigma \Phi(\psi_j^{(a,b)}),$$

where  $\psi_j^{(a,b)}(t) := \frac{1}{\sqrt{a}} \psi_j\left(\frac{t-b}{a}\right)$  and  $\chi_{Z_\ell^{(j)}}(b)$  is the indicator function of the set  $Z_\ell^{(j)}(b)$ .

To simplify further notation, we shall use boldface for  $J$ -dimensional “vector” quantities; for instance we denote  $\boldsymbol{\psi} := [\psi_1, \psi_2, \dots, \psi_J]^\top \in \oplus^J \mathcal{S}$  and  $\boldsymbol{\psi}^{[\mathbf{r}]} := \mathbf{r}^\top \boldsymbol{\psi}$ , where  $\mathbf{r} \in S^{J-1} = \{\mathbf{v} \in \mathbb{R}^J ; \|\mathbf{v}\|^2 = \sum_{j=1}^J v_j^2 = 1\}$ . Clearly  $\boldsymbol{\psi}^{[\mathbf{r}]}$  is also a Schwartz function, with  $\text{supp}(\widehat{\boldsymbol{\psi}^{[\mathbf{r}]}}) \subset [1-\bar{\Delta}, 1+\bar{\Delta}]$ . We similarly introduce  $\boldsymbol{\epsilon}(a, b) := [\epsilon_1(a, b), \dots, \epsilon_J(a, b)]^\top$  (a vector with norm of order  $\bar{\epsilon}$ ),  $\boldsymbol{Q}_\ell(a, b) := [Q_{1,\ell}(a, b), \dots, Q_{J,\ell}(a, b)]^\top$ . Note that all the entries of the vectors  $\boldsymbol{Q}_\ell(a, b)$  are real; this will be important for our estimates below.  $\boldsymbol{\Phi}(a, b) := [\Phi(\psi_1^{(a,b)}), \dots, \Phi(\psi_J^{(a,b)})]^\top$ .  $\boldsymbol{\Phi}(a, b)$  is a complex Gaussian random vector [6], of which the following Lemma gives some basic properties:

**Lemma S.3.3.** For all  $a > 0$  and  $b \in \mathbb{R}$ ,  $\boldsymbol{\Phi}(a, b)$  is a complex Gaussian random vector with mean  $[0, \dots, 0]^\top \in \mathbb{R}^J$ , for which the covariance matrix and the relation matrix both equal  $I_{J \times J}$ . Thus, for all  $\mathbf{v} \in \mathbb{R}^J$ ,  $\mathbf{v}^\top \boldsymbol{\Phi}(a, b)$  is a complex Gaussian random variable with mean 0 and variance  $\|\mathbf{v}\|^2$ .

*Proof.* Fix  $a > 0$  and  $b \in \mathbb{R}$ . Since  $\Phi$  is a Gaussian white noise and  $\psi$  is a complex Schwartz function, it follows that for  $j = 1, \dots, J$ ,  $\Phi(\psi_j^{(a,b)})$  is a complex Gaussian random variable [7]. By definition, its mean is 0 and its variance is

$$(S.6) \quad \text{Var}(\Phi(\psi_j^{(a,b)})) = \mathbb{E}|\Phi(\psi_j^{(a,b)})|^2 = \int \left| \widehat{\psi_j^{(a,0)}}(\xi) \right|^2 d\xi = \|\widehat{\psi_j}\|_{L^2(\mathbb{R})}^2 = 1.$$

It is clear that the variance of  $\Phi(\psi_j^{(a,b)})$  is independent of the scale  $a$ . Since  $\psi_i$  and  $\psi_j$  are orthogonal if  $i \neq j$ , a similar calculation shows that  $\text{Cov}(\Phi(\psi_i^{(a,b)}), \Phi(\psi_j^{(a,b)})) = \delta_{i,j}$ . Since we assume that  $\widehat{\psi_j}$  is real for all  $j = 1, \dots, J$ , the relation matrix of  $\boldsymbol{\Phi}(a, b)$  equals the covariance matrix, and is thus  $I_{J \times J}$  as well. It then easily follows that  $\mathbf{v}^\top \boldsymbol{\Phi}(a, b)$  is a complex Gaussian random variable with mean 0 and variance  $\|\mathbf{v}\|^2$ .  $\square$

Because the CWT is (anti)linear in the wavelet with respect to which it is computed, the CWT of  $Y$  with respect to  $\psi^{[r]}$  is given by

$$(S.7) \quad W_Y^{(\psi^{[r]})}(a, b) = \sum_{\ell=1}^L \sum_{j=1}^J r_j e^{i2\pi\varphi_j(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}}(a, b) + \mathbf{r}^\top [\boldsymbol{\epsilon}(a, b) + \sigma \boldsymbol{\Phi}(a, b)].$$

The analysis in [4, 2] also shows that

$$\partial_b W_G^{(\psi_j)}(a, b) = \begin{cases} i2\pi (\varphi'_\ell(b) e^{i2\pi\varphi_\ell(b)} Q_{j,\ell}(a, b) + \tilde{\epsilon}_j(a, b)) & \text{when } a \in Z_\ell^{(j)}(b) \text{ for some } \ell \\ i2\pi \tilde{\epsilon}_j(a, b) & \text{otherwise,} \end{cases}$$

where  $\tilde{\epsilon}_j(a, b)$  is of order  $\bar{\epsilon}$  for all  $j = 1, \dots, J$ . We thus obtain

$$-i\partial_b W_Y^{(\psi^{[r]})}(a, b) = 2\pi \sum_{\ell=1}^L \sum_{j=1}^R r_j \varphi'_j(b) e^{i2\pi\varphi_j(b)} Q_{j,\ell}(a, b) \chi_{Z_\ell^{(j)}}(a, b) 2\pi \mathbf{r}^\top [\tilde{\epsilon}(a, b) + \sigma \tilde{\boldsymbol{\Phi}}(a, b)],$$

where  $\tilde{\epsilon}(a, b) = [\tilde{\epsilon}_1(a, b), \dots, \tilde{\epsilon}_J(a, b)]^\top$ , and  $\tilde{\boldsymbol{\Phi}}(a, b) = (2\pi)^{-1}[\Phi(i(\psi_1^{(a,b)})'), \dots, \Phi(i(\psi_J^{(a,b)})')]^\top$ . Here  $\tilde{\epsilon}(a, b)$  is again a  $J$ -dim random vector with norm of order  $\bar{\epsilon}$ . The following lemma gives some basic properties of the complex random vector  $\tilde{\boldsymbol{\Phi}}(a, b)$ :

**Lemma S.3.4.** For all  $a > 0$  and  $b \in \mathbb{R}$ ,  $\tilde{\boldsymbol{\Phi}}(a, b)$  is a complex Gaussian random vector with mean  $[0, \dots, 0]^\top \in \mathbb{R}^J$ , for which the covariance matrix and the relation matrix both equal  $\text{diag}[\|\widehat{\psi}_1'\|^2, \dots, \|\widehat{\psi}_J'\|^2]/(2\pi a)^2 \in \mathbb{R}^{J \times J}$ . Thus, for all  $\mathbf{v} \in \mathbb{R}^p$ ,  $\mathbf{v}^\top \tilde{\boldsymbol{\Phi}}(a, b)$  is a complex Gaussian random variable with mean 0 and variance  $\sum_{j=1}^J \mathbf{v}_j^2 \|\widehat{\psi}_j'\|^2 / (2\pi a)^2$ .

*Proof.* The proof is the same as that of Lemma S.3.3, except for the following slight difference:

$$\text{Var}(\Phi(i(\psi_j^{(a,b)})')) = \mathbb{E}|\Phi((\psi_j^{(a,b)})')|^2 = \|(\psi_j^{(a,b)})'\|_{L^2(\mathbb{R})}^2 = \|\psi_j'\|_{L^2(\mathbb{R})}^2/a^2.$$

□

As a result, when  $a \in Z_\ell(b)$ , the reassignment rule,  $\omega_Y^{(\psi^{[r]})}(a, b)$ , becomes

$$\begin{aligned} \omega_Y^{(\psi^{[r]})}(a, b) &= \frac{-i\partial_b W_G^{(\psi^{[r]})}(a, b) + 2\pi \mathbf{r}^\top \sigma \tilde{\boldsymbol{\Phi}}(a, b)}{2\pi (W_G^{(\psi^{[r]})}(a, b) + \sigma \mathbf{r}^\top \boldsymbol{\Phi}(a, b))} \\ &= \frac{\mathbf{r}^\top [\varphi'_\ell(b) e^{i2\pi\varphi_\ell(b)} Q_\ell(a, b) + \tilde{\epsilon}(a, b) + \sigma \tilde{\boldsymbol{\Phi}}(a, b)]}{\mathbf{r}^\top [e^{i2\pi\varphi_\ell(b)} Q_\ell(a, b) + \boldsymbol{\epsilon}(a, b) + \sigma \boldsymbol{\Phi}(a, b)]}, \end{aligned}$$

It follows from Lemma S.3.3 and Lemma S.3.4 that  $\omega_Y^{(\psi^{[r]})}(a, b)$  is a ratio random variable of two independent complex Gaussian random variables with non-zero means.

Note that we are implicitly assuming here that the denominator in the fraction for  $\omega_Y^{(\psi^{[r]})}(a, b)$  is not too small (see Section 2 in the main paper). In what follows, we shall make this explicit: we shall always assume that

$$2\pi |W_Y^{(\psi^{[r]})}(a, b)| \left( = \left| \mathbf{r}^\top [e^{i2\pi\varphi_\ell(b)} Q_\ell(a, b) + \boldsymbol{\epsilon}(a, b) + \sigma \boldsymbol{\Phi}(a, b)] \right| \text{ if } a \in Z_\ell(b) \right)$$

exceeds the value  $2\kappa$ , where the value of  $\kappa$  can be set (according to the signal characteristics and noise level). At the same time, we shall assume that  $\bar{\epsilon}$  and  $\sigma$  are sufficiently small that

$$\mathbb{E}(\|\boldsymbol{\epsilon}(a, b) + \sigma \boldsymbol{\Phi}(a, b)\|^2) \leq \kappa^2.$$

(This means that the threshold for the reassignment rule must be set in accordance with the rate of change of the amplitudes and the instantaneous frequencies of the individual constituent components in the clean signal, as well as with the level of the noise – both eminently reasonable restrictions.) We shall see below how these restrictions will come into play.

Let us first prove some technical Lemmas.

**Lemma S.3.5.** Fix  $J \in \mathbb{N}$  and  $\kappa > 0$ . Denote  $S_\kappa := \{\mathbf{r} \in S^{J-1}; \mathbf{r}^\top \mathbf{v} > \kappa\}$ . For  $\mathbf{u} \in \mathbb{C}^J$  and  $\mathbf{v} \in \mathbb{R}^J$ , we have

$$(S.8) \quad \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top \mathbf{u}}{\mathbf{r}^\top \mathbf{v}} d\mathbf{r} = \mathfrak{p}_\mathbf{v} \mathbf{u} := \mathfrak{p}_\mathbf{v} \Re \mathbf{u} + i \mathfrak{p}_\mathbf{v} \Im \mathbf{u},$$

where  $\Re \mathbf{u}$  is the real part of  $\mathbf{u}$ ,  $\Im \mathbf{u}$  is the imaginary part of  $\mathbf{u}$  and  $\mathfrak{p}_\mathbf{v}(\mathbf{w})$  is the component of the vector  $\mathbf{w}$  along the direction of  $\mathbf{v}$ ,  $\mathfrak{p}_\mathbf{v}(\mathbf{w}) := \mathbf{v}^\top \mathbf{w} / \|\mathbf{v}\|$ . Furthermore

$$(S.9) \quad \frac{1}{|S_\kappa|} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top \mathbf{u}}{\mathbf{r}^\top \mathbf{v}} \right|^2 d\mathbf{r} = |\mathfrak{p}_\mathbf{v} \mathbf{u}|^2 + c \frac{\|\mathcal{P}_\mathbf{v}^\perp \mathbf{u}\|_2^2}{J-1},$$

where  $\mathcal{P}_\mathbf{v}^\perp$  is the projection operator onto the subspace perpendicular to  $\mathbf{v}$  and

$$c = \frac{2\Gamma((J-1)/2)}{\sqrt{\pi}\Gamma(J/2)} \int_\kappa^1 \frac{(1-x^2)^{(J-1)/2}}{x^2} dx \approx \frac{2\sqrt{2}}{\sqrt{\pi J} \kappa}.$$

*Proof.* Without loss of generality, we can assume that  $\mathbf{u} \in \mathbb{R}^p$  and  $\|\mathbf{v}\| = 1$ . We can find  $\mathcal{R} \in SO(J)$  so that  $\mathcal{R}\mathbf{v} = \mathbf{e}_1$ , where  $\mathbf{e}_1 := [1, 0, \dots, 0]^\top \in S^{J-1}$ . Under this change of variable, we write

$$\mathcal{R}\mathbf{u} = [d_1, d_2, \dots, d_J]^\top \in \mathbb{R}^J,$$

where

$$d_1 = \mathbf{u}^\top \mathbf{v} = \mathfrak{p}_\mathbf{v}(\mathbf{u}).$$

As a result, we have

$$\begin{aligned} \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top \mathbf{u}}{\mathbf{r}^\top \mathbf{v}} d\mathbf{r} &= \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{(\mathcal{R}\mathbf{r})^\top \mathcal{R}\mathbf{u}}{(\mathcal{R}\mathbf{r})^\top \mathcal{R}\mathbf{v}} d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\mathbf{r}^\top [d_1, d_2, \dots, d_J]^\top}{\mathbf{r}^\top \mathbf{e}_1} d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\sum_{j=1}^J r_j d_j}{r_1} d\mathbf{r}, \end{aligned}$$

which becomes  $d_1 = \mathbf{p}_v \mathbf{u}$  since, for  $j \neq 1$ ,  $\frac{r_j}{r_1}$  is an odd function of  $r_j$ , and the domain of integration is invariant under sign reversal of  $r_j$ . For the second part, note that by the same change of variable, we have

$$\begin{aligned} & \frac{1}{|S_\kappa|} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top \mathbf{u}}{\mathbf{r}^\top \mathbf{v}} \right|^2 d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\left( \sum_{j=1}^J r_j d_j \right)^2}{r_1^2} d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\sum_{j=1}^J r_j^2 d_j^2 + 2 \sum_{i \neq j} r_i r_j d_i d_j}{r_1^2} d\mathbf{r} \\ &= d_1^2 + \sum_{j=2}^J d_j^2 \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{r_j^2}{r_1^2} d\mathbf{r}, \end{aligned}$$

where the last equality holds because  $\frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\sum_{i \neq j} r_i r_j d_i d_j}{r_1^2} d\mathbf{r} = 0$  since in each term  $\frac{r_i r_j}{r_1^2}$  with  $i \neq j$ , there is at least one index different from 1, so that this term changes sign when it is mirrored with respect to that index, while the domain of integration is invariant under this mirroring operation. For the other terms, note that when  $j = 2, \dots, p$ , symmetry arguments imply that

$$\begin{aligned} & \frac{1}{|S_\kappa|} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{r_j^2}{r_1^2} d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \frac{1}{J-1} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{\sum_{l=2}^p r_l^2}{r_1^2} d\mathbf{r} \\ &= \frac{1}{|S_\kappa|} \frac{1}{J-1} \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{1 - r_1^2}{r_1^2} d\mathbf{r}. \end{aligned}$$

To evaluate the last term, we use spherical coordinates in  $J$  dimensions. Rewrite  $\mathbf{r} = [r_1, \dots, r_J]^\top \in S^{J-1}$  as

$$\begin{cases} r_1 = \cos \varphi_1 \\ r_2 = \sin \varphi_1 \cos \varphi_2 \\ \vdots \\ r_{J-1} = \sin \varphi_1 \dots \sin \varphi_{J-2} \cos \varphi_{J-1} \\ r_J = \sin \varphi_1 \dots \sin \varphi_{J-2} \sin \varphi_{J-1}, \end{cases}$$

where  $\varphi_1, \dots, \varphi_{J-2} \in [0, \pi]$  and  $\varphi_{J-1} \in [0, 2\pi)$ . In this coordinate system, the volume form becomes  $d\mathbf{r} = (\sin^{J-2} \varphi_1)(\sin^{J-3} \varphi_2) \dots (\sin \varphi_{J-2}) d\varphi_{J-1} d\varphi_{J-2} \dots d\varphi_1$ ; we obtain thus

$$\begin{aligned} & \int_{\{\mathbf{r} \in S^{J-1}; r_1 > \kappa\}} \frac{1}{r_1^2} d\mathbf{r} \\ &= 2 \int_{I_\kappa} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} \frac{\sin^{J-2} \varphi_1 \sin^{J-3} \varphi_2 \dots \sin \varphi_{J-2}}{\cos^2 \varphi_1} d\varphi_{J-1} d\varphi_{J-2} \dots d\varphi_1 \\ &= 2|S^{J-2}| \int_{I_\kappa} \frac{\sin^{J-2} \varphi_1}{\cos^2 \varphi_1} d\varphi_1, \end{aligned}$$

where  $I_\kappa = \{\varphi_1 \in [0, \pi/2]; \cos \varphi_1 > \kappa\}$ . Similarly, we have

$$\int_{\{r \in S^{J-1}; r_1 > \kappa\}} dr = 2|S^{J-2}| \int_{I_\kappa} \sin^{J-2} \varphi_1 d\varphi_1.$$

By putting the above together, we have finished the claim since

$$\frac{1}{|S_\kappa|} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top \mathbf{u}}{\mathbf{r}^\top \mathbf{v}} \right|^2 dr = d_1^2 + c \frac{\sum_{j=2}^J d_j^2}{J-1} = |\mathbf{p}_v \mathbf{u}|^2 + c \frac{\|\mathcal{P}_v^\perp \mathbf{u}\|_2^2}{J-1},$$

where

$$(S.10) \quad c = \frac{2|S^{J-2}|}{|S_\kappa|} \int_{I_\kappa} \sin^{J-2} \varphi_1 \left( \frac{1}{\cos^2 \varphi_1} - 1 \right) d\varphi_1 = \frac{2\Gamma((J-1)/2)}{\sqrt{\pi}\Gamma(J/2)} \int_{I_\kappa} \frac{\sin^J \varphi}{\cos^2 \varphi} d\varphi.$$

Notice that the Gamma function ratio  $\frac{\Gamma((J-1)/2)}{\Gamma(J/2)}$  can be asymptotically approximated by  $(J/2)^{-1/2}$  as  $J \rightarrow \infty$  and that

$$\int_{I_\kappa} \frac{\sin^J \varphi}{\cos^2 \varphi} d\varphi = \int_\kappa^1 \frac{(1-u^2)^{(J-1)/2}}{u^2} du = \int_\kappa^1 \frac{1}{u^2} (1+O(u^2)) du = \frac{1}{\kappa} + O(1).$$

It follows that  $c$  is approximately

$$(S.11) \quad c \approx \frac{2\sqrt{2}}{\sqrt{\pi J \kappa}}.$$

□

We are now ready to study the statistical behavior of  $\omega_Y^{(\psi^{[\mathbf{r}]})}(a, b)$  as the unit vector  $\mathbf{r}$  is picked randomly, uniformly in  $\tilde{S}_\kappa^{(\ell)} = \{\mathbf{r} \in S^{J-1}; \mathbf{r}^\top (e^{-i2\pi\varphi_\ell(b)} \mathbf{Q}_\ell(a, b) + \boldsymbol{\epsilon}(a, b) + \sigma\Phi(a, b)) > 2\kappa\} \subset S^{J-1}$ . In the next proposition, we keep  $\ell, b$  and  $a$  fixed, on the understanding that  $a \in Z_\ell(b)$ . To ease up on notation, we shall suppress  $(a, b)$  and  $\ell$  in the notation, and use  $\omega_Y^{(\psi^{[\mathbf{r}]})}, \mathbf{Q}, \varphi(b), \tilde{\boldsymbol{\epsilon}}, \tilde{\Phi}, \tilde{S}_\kappa$ , etc, to denote  $\omega_Y^{(\psi^{[\mathbf{r}]})}(a, b), \mathbf{Q}_\ell(a, b), \varphi_\ell(b), \tilde{\boldsymbol{\epsilon}}(a, b), \tilde{\Phi}(a, b), \tilde{S}_\kappa^{(\ell)}$ , etc.

**Proposition S.3.6.** Fix a realization of  $\Phi, \kappa > 0, b \in \mathbb{R}$  and  $a \in Z_\ell(b)$ . Assume that  $\mathbf{r}$  is sampled uniformly from  $\tilde{S}_\kappa = \{\mathbf{r} \in S^{J-1}; \mathbf{r}^\top (e^{-i2\pi\varphi(b)} \mathbf{Q} + \boldsymbol{\epsilon} + \sigma\Phi) > 2\kappa\} \subset S^{J-1}$ . When  $\|\boldsymbol{\epsilon} + \sigma\Phi\|_2^2 < \kappa$ , we have

$$(S.12) \quad \mathbb{E}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} = \varphi'(b) + e^{-i2\pi\varphi(b)} \mathbf{p}_Q (\tilde{\boldsymbol{\epsilon}} + \sigma\tilde{\Phi} - \varphi'(b)[\boldsymbol{\epsilon} + \sigma\Phi]) + E_1,$$

where  $\mathbb{E}_{\mathbf{r}}$  is the expectation of  $\omega_Y^{(\psi^{[\mathbf{r}]})}(a, b)$  as  $\mathbf{r}$  is sampled randomly and uniformly from  $\tilde{S}_\kappa$ , and  $E_1$  is bounded by

$$(S.13) \quad |E_1| \leq \frac{1}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q (\tilde{\boldsymbol{\epsilon}} + \sigma\tilde{\Phi} - \varphi'(b)[\boldsymbol{\epsilon} + \sigma\Phi])|^2 + c \frac{\|\tilde{\boldsymbol{\epsilon}} + \sigma\tilde{\Phi} - \varphi'(b)[\boldsymbol{\epsilon} + \sigma\Phi]\|^2}{J-1} \right)^{1/2}.$$

Furthermore we have

$$\text{Var}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} \leq \frac{5}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q (\tilde{\boldsymbol{\epsilon}} + \sigma\tilde{\Phi} - \varphi'(b)[\boldsymbol{\epsilon} + \sigma\Phi])|^2 + c \frac{\|\tilde{\boldsymbol{\epsilon}} + \sigma\tilde{\Phi} - \varphi'(b)[\boldsymbol{\epsilon} + \sigma\Phi]\|^2}{J-1} \right).$$

where  $\text{Var}_{\mathbf{r}}$  is the variance of  $\omega_Y^{(\psi^{[\mathbf{r}]})}$  over  $\tilde{S}_\kappa$ .

Before the proof, we have the following remark about the Proposition.

*Remark.* In the statement of this proposition, we encounter several times the expression  $|\mathbf{p}_Q \mathbf{V}|$  (using the shorthand notation  $\mathbf{V} = \tilde{\epsilon} + \sigma \tilde{\Phi} - \varphi'(b)[\epsilon + \sigma \Phi]$ ), which can be bounded by  $\|\mathbf{V}\|$ . In practice, however, the term  $|\mathbf{p}_Q \mathbf{V}|$  will likely be significantly smaller than its norm, with high probability if  $J$  is large. Indeed, the vector  $\mathbf{Q}$  is fixed, while the vector  $\mathbf{V}$  is a random vector in  $J$  dimensions, depending on the random realization of the noise function  $\Phi$ , which is much more likely than not to lie in a region near the equator, perpendicular to  $\mathbf{Q}$ , since this region contributes the lion share of the sphere “area” (really a  $J - 1$ -dimensional volume), increasingly so as  $J$  increases. Denoting  $I_\gamma := \{\varphi_1 \in [0, \pi/2] ; 0 \leq \cos(\varphi_1) < \gamma\}$ , we have indeed

$$\begin{aligned} & |\{\mathbf{r} \in S^{J-1} ; 0 \leq r_1 < \gamma\}| \\ &= \int_{I_\gamma} \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} (\sin^{J-2} \varphi_1) (\sin^{J-3} \varphi_2) \dots (\sin \varphi_{J-2}) d\varphi_{J-1} d\varphi_{J-2} \dots d\varphi_1 \\ &= |S^{J-2}| \int_{I_\gamma} \sin^{J-2} \varphi_1 d\varphi_1. \end{aligned}$$

Consequently,

$$\frac{|\{\mathbf{r} \in S^{J-1} ; r_1 < \gamma\}|}{|\{\mathbf{r} \in S^{J-1} ; 0 \leq r_1 \leq 1\}|} = \frac{\int_{I_\gamma} \sin^{J-2} \varphi d\varphi}{\int_0^{\pi/2} \sin^{J-2} \varphi d\varphi} = 1 - \frac{\int_\gamma^1 (1-u^2)^{(J-3)/2} du}{\int_0^1 (1-u^2)^{(J-3)/2} du},$$

which approaches 1 as  $J$  increases to  $\infty$ .

*Proof.* (of the Proposition.) To simplify the notation in the computation, we set  $\mathbf{A} := \varphi'(b)e^{i2\pi\varphi(b)}\mathbf{Q}$ ,  $\mathbf{a} := \tilde{\epsilon} + \sigma \tilde{\Phi}$ ,  $\mathbf{B} := e^{i2\pi\varphi(b)}\mathbf{Q} = \mathbf{A}/\varphi'(b)$ ,  $\mathbf{b} := \epsilon + \sigma \Phi$ .

By the assumption that  $\mathbf{r}$  is sampled uniformly from  $S_\kappa$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} &= \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top (\mathbf{A} + \mathbf{a})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} d\mathbf{r} = \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top (\varphi'(b)\mathbf{B} + \mathbf{a})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} d\mathbf{r} \\ &= \frac{\varphi'(b)}{|S_\kappa|} \int_{S_\kappa} \left[ 1 + \frac{\mathbf{r}^\top \left( \frac{\mathbf{a}}{\varphi'(b)} - \mathbf{b} \right)}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} \right] d\mathbf{r} \\ (S.14) \quad &= \varphi'(b) + \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} d\mathbf{r}. \end{aligned}$$

We next use the identity

$$\frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} = \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top \mathbf{B}} - \frac{\mathbf{r}^\top \mathbf{b}}{\mathbf{r}^\top \mathbf{B}} \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})},$$

combining it with Lemma S.3.5 (since  $\mathbf{Q} \in \mathbb{R}^p$ ), to obtain

$$(S.15) \quad \mathbb{E}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} = \varphi'_k(b) + e^{-i2\pi\varphi(b)} \mathbf{p}_Q \left( \tilde{\epsilon} + \sigma \tilde{\Phi} - \varphi'(b)[\epsilon + \sigma \Phi] \right) + E_1,$$

where

$$(S.16) \quad E_1 := - \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top \mathbf{b}}{\mathbf{r}^\top \mathbf{B}} \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} d\mathbf{r}.$$

Note that by the assumptions that  $\|\epsilon + \sigma \Phi\|_2 \leq \kappa$  and  $|\mathbf{r}^\top (e^{i2\pi\varphi(b)}\mathbf{Q} + \epsilon + \sigma \Phi)| > 2\kappa$ , we have

$$\frac{|\mathbf{r}^\top \mathbf{b}|}{|\mathbf{r}^\top (\mathbf{B} + \mathbf{b})|} < \frac{1}{2},$$

so that

$$|E_1| \leq \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{|\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})|}{2|\mathbf{r}^\top \mathbf{Q}|} d\mathbf{r}.$$

Next, we apply the Cauchy-Schwarz inequality to this integral, together with Lemma S.3.5, which leads to

$$\begin{aligned} |E_1| &\leq \frac{1}{|S_\kappa|} \left[ \int_{S_\kappa} d\mathbf{r} \right]^{1/2} \left[ \int_{S_\kappa} \frac{|\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})|^2}{2|\mathbf{r}^\top \mathbf{Q}|^2} \right]^{1/2} \\ &\leq \frac{1}{2} \left( |\mathbf{p}_Q(\mathbf{a} - \varphi'(b)\mathbf{b})|^2 + c \frac{\|\mathcal{P}_Q^\perp(\mathbf{a} - \varphi'(b)\mathbf{b})\|^2}{J-1} \right)^{1/2} \\ &= \frac{1}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q(\mathbf{a} - \varphi'(b)\mathbf{b})|^2 + c \frac{\|\mathbf{a} - \varphi'(b)\mathbf{b}\|^2}{J-1} \right)^{1/2} \\ &= \frac{1}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q(\tilde{\epsilon} + \sigma\tilde{\Phi} - \varphi'(b)[\epsilon + \sigma\Phi])|^2 + c \frac{\|\tilde{\epsilon} + \sigma\tilde{\Phi} - \varphi'(b)[\epsilon + \sigma\Phi]\|^2}{J-1} \right)^{1/2}. \end{aligned}$$

The variance can be evaluated in the same manner. Noting that for a random variable  $X$ ,  $\text{Var}X = \text{Var}(X - c)$  for any constant  $c$ , we have

$$\begin{aligned} \text{Var}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} &= \text{Var}_{\mathbf{r}} [\omega_Y^{(\psi^{[\mathbf{r}]})} - \varphi'(b)] \\ &= \frac{1}{|S_\kappa|} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} \right|^2 d\mathbf{r} - \left| \frac{1}{|S_\kappa|} \int_{S_\kappa} \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} d\mathbf{r} \right|^2. \\ (S.17) \quad &\leq \frac{1}{|S_\kappa|} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} \right|^2 d\mathbf{r} \end{aligned}$$

This last expression (S.17) can be bounded by

$$\begin{aligned} &\frac{2}{|S_\kappa|} \int_{S_\kappa} \left( \left| \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top \mathbf{B}} \right|^2 + \left| \frac{\mathbf{r}^\top \mathbf{b}}{\mathbf{r}^\top \mathbf{B}} \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top (\mathbf{B} + \mathbf{b})} \right|^2 \right) d\mathbf{r} \\ &\leq \frac{2}{|S_\kappa|} \frac{5}{4} \int_{S_\kappa} \left| \frac{\mathbf{r}^\top (\mathbf{a} - \varphi'(b)\mathbf{b})}{\mathbf{r}^\top \mathbf{B}} \right|^2 d\mathbf{r}. \end{aligned}$$

We have encountered this exact same integral before, and bounded it by invoking Lemma S.3.5. We thus obtain

$$\begin{aligned} \text{Var}_{\mathbf{r}} \omega_Y^{(\psi^{[\mathbf{r}]})} &\leq \frac{5}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q(\mathbf{a} - \varphi'(b)\mathbf{b})|^2 + c \frac{\|\mathbf{a} - \varphi'(b)\mathbf{b}\|^2}{J-1} \right) \\ &= \frac{5}{2} \left( \left[ 1 - \frac{c}{J-1} \right] |\mathbf{p}_Q(\tilde{\epsilon} + \sigma\tilde{\Phi} - \varphi'(b)[\epsilon + \sigma\Phi])|^2 + c \frac{\|\tilde{\epsilon} + \sigma\tilde{\Phi} - \varphi'(b)[\epsilon + \sigma\Phi]\|^2}{J-1} \right). \end{aligned}$$

□

This concludes this section concerning the details for the technical estimates in section 3 of the main paper.

## ESM-4. NUMERICAL RESULTS

As described in the main paper, we consider both CWT and STFT-based ConeCFT representations. In both cases, the orthogonal family of reference functions (wavelets for the CWT, windows for the STFT) are the eigenfunctions, up to a certain order, of a time-frequency localization operator that is particularly well suited to the CWT or STFT framework [5, 8, 3, 12]. Figure S.1 below shows the shape and size of TF domains of this type. In both cases, the shapes correspond to a two-parameter family, and the localization operators behave approximately like projection operators. More precisely, once the parameters  $\Lambda$  determining the shape are picked, there is a natural family of (commuting) operators  $T^{(\Lambda, R)}$  and an orthonormal family of functions  $\psi_j^{(\Lambda)}$  such that

$$T^{(\Lambda, R)} \psi_j^{(\Lambda)} = E_j^{(\Lambda, R)} \psi_j^{(\Lambda)},$$

where the eigenvalues  $E_j^{(\Lambda, R)}$ , all between 0 and 1, constitute a strictly decreasing sequence, tending to 0 as  $j$  tends to  $\infty$ ; for fixed  $\Lambda$  and  $j$ , each  $E_j^{(\Lambda, R)}$  increases with  $R$ , tending to 1 as  $R$  (which indicates the size of the region characterized by  $\Lambda$ ) tends to  $\infty$ . The eigenfunctions themselves (which do *not* depend on  $R$ ) are scaled and possibly chirped Hermite functions for the STFT case, and Morse functions in the CWT case.

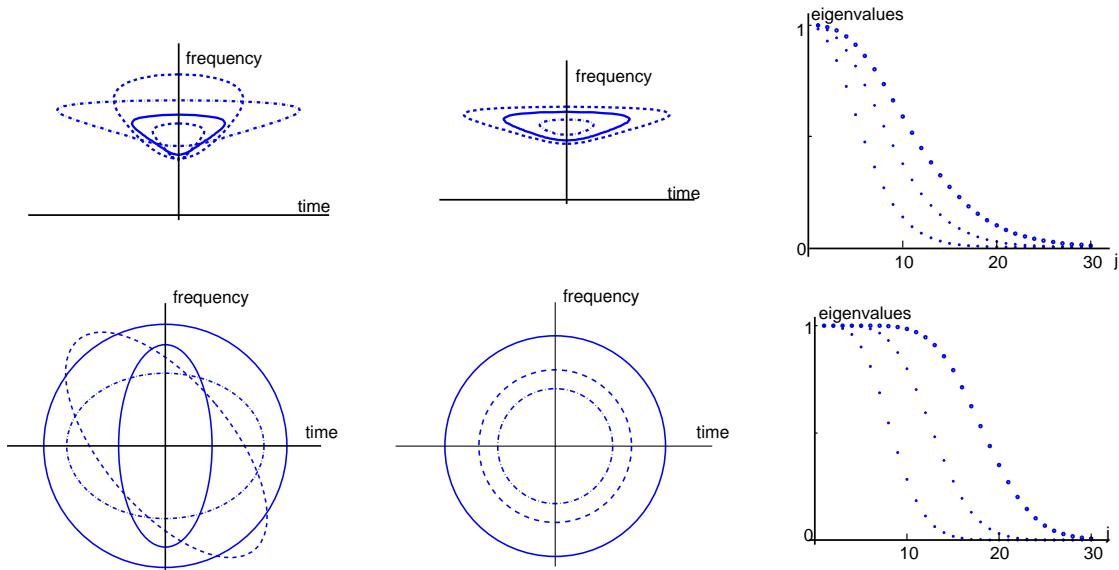


FIGURE S.1. Localization domains in the TF plane for the reference windows or CWT reference wavelets: Top: CWT, bottom: STFT. From left to right: different shapes of the TF domain, corresponding to different parameter choices  $\Lambda$ ; different sizes of one domain shape, corresponding (for one fixed  $\Lambda$ ) to different  $R$ ; eigenvalues  $E_j^{(\Lambda, R)}$ , for different  $R$ .

It seems natural to pick these special orthonormal families, since each family provides, in some sense (made precise in [5, 8, 3, 12]) the “best” localization, simultaneously, by different orthonormal functions, for

one shared time-frequency domain. (A similar reason underlies the choice, in standard multi-taper methods for spectral estimation, of the prolate spheroidal wave functions for the taper functions [10, 9, 1].) However, the method does not depend on these particular choices, and it is not only conceivable, but indeed likely, that for particular applications, other choices may be more suitable and give better results.

**ESM-4a Data simulation.** Figure S.2 below shows the graph of (the restriction to  $[15, 40]$ ) of another signal  $s^* \in \mathcal{C}$ . This signal is used in the main paper to illustrate the action of ConceFT on a signal from  $\mathcal{C}$  that has played no role in calibrating the ConceFT parameters (unlike  $s$ ).

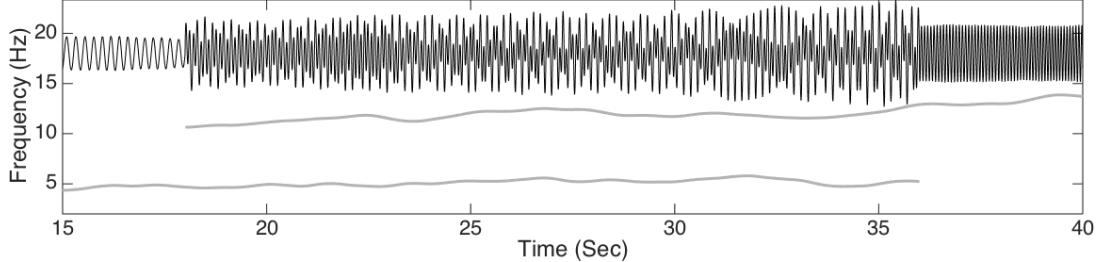


FIGURE S.2. Another signal  $s^*$  (in black) in  $\mathcal{C}$ , and the corresponding instantaneous frequencies (in gray) of the two components, restricted to the time interval  $[15, 40]$ .

Figure S.3 plots a realization of  $Y^*(t) = s^*(t) + \sigma\xi(t)$  for each of the three noise processes (Gaussian, ARMA(1,1) and Poisson), restricted to the subinterval  $[15, 40]$ .

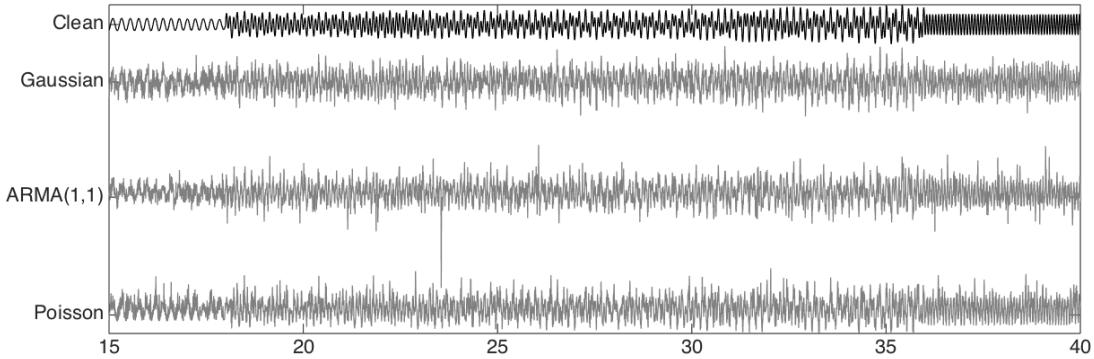


FIGURE S.3. The restrictions to  $[15, 40]$  of the noisy signal  $Y^* = s^* + \sigma\xi$  (2nd row to bottom), where  $s^*$  is the clean signal from the previous figure (plotted again in the top row) and where the added noise is Gaussian, ARMA, or Poisson noise (in order, from); in each case  $\sigma$  is picked so that the noisy signal has 0 dB SNR (signal to noise ratio). The four plots are at the same vertical scale.

**ESM-4b Performance evaluation.** To assess the performance of ConceFT, we must compare the time-varying Power Spectrum (tvPS)  $\tilde{P}_Y$ , as estimated via ConceFT, with the *ideal time-varying power spectrum*

(itvPS) of the clean simulated signal  $s$ , defined (in a natural interpretation of its construction procedure) as

$$\mathbb{P}_s(t, \omega) := \sum_{k=1}^2 A_k^2(t) \delta_{\varphi'_k(t)(\omega)}.$$

Viewing both the itvPS and the tvPS as distributions on the TF-plane, we want to assess, in particular, whether the regions in the TF plane where they each concentrate, coincide or lie close to each other. The *Optimal Transport* (OT) distance (also called the Earth Mover distance) is a distance that is designed to do this: given two probability measures on the same set, their OT-distance gives the amount of “work” needed to “deform” one into the other. A bit more precisely, it computes the total (integral/sum of the product) *mass*  $\times$  *distance traveled* for the transformation (i.e. the transportation plan that minimizes this quantity) that maps one to the other. Because the principle of ConceFT is to “reassign” content in the TF plane, keeping the time-variable fixed (see Section 2), we also compute the OT-distance for each individual  $t$  (keeping  $t$  fixed), and then take the average over all  $t \in [0, T]$ . This has a fortuitous advantage, in that it reduces the OT-distance computations to 1-dimensional problems, for which there exists a computational short-cut: the standard definition for the OT-distance between probability distributions  $\mu$  and  $\nu$  on a metric space  $(S, d)$  involves an optimization over  $\mathcal{P}(\mu, \nu)$ , the set of all probability measures on  $S \times S$  that have  $\mu$  and  $\nu$  as marginals,

$$d_{\text{OT}}(\mu, \nu) := \inf_{\rho \in \mathcal{P}(\mu, \nu)} \int d(x, y) \, d\rho(x, y),$$

which can be computationally quite expensive. In the one-dimensional case (i.e. when  $S \subset \mathbb{R}$ , and  $d$  is the canonical Euclidean distance,  $d(x, y) = |x - y|$ ), however, it turns out (see e.g. section 2.2 in [11]) that, defining  $f_\mu(x) = \int_{-\infty}^x d\mu$  (analogously for  $f_\nu$ ), we have

$$d_{\text{OT}}(\mu, \nu) = \int_S |f_\mu(x) - f_\nu(x)| \, dx.$$

The OT-distance is defined for *probability* distributions, and it is by no means guaranteed that the positive functions  $\tilde{\mathbb{P}}_Y(t, \cdot)$  and  $\mathbb{P}_s(t, \cdot)$  have integral 1 for all  $t$ ; for this reason, we normalize them before computing their OT-distance. We may also want to capture (and penalize in the distance metric) possible differences in the total weights of  $\tilde{\mathbb{P}}_Y(t, \cdot)$  and  $\mathbb{P}_s(t, \cdot)$ ; we can introduce a term for this as well. More precisely, assuming that the frequency domain over which  $\tilde{\mathbb{P}}_Y$  and  $\mathbb{P}_s$  range is  $[0, \Omega]$ , and assuming also that  $\int_0^T \int_0^\Omega \tilde{\mathbb{P}}_Y(t, \omega) \, d\omega \, dt = \int_0^T \int_0^\Omega \mathbb{P}_s(t, \omega) \, d\omega \, dt$  (which can be achieved by multiplying  $\tilde{\mathbb{P}}_Y$  with a constant, if necessary), we define

$$\begin{aligned} \tilde{p}_Y(t, \omega) &= \int_0^\omega \tilde{\mathbb{P}}_Y(t, \xi) \, d\xi & p_s(t, \omega) &= \int_0^\omega \mathbb{P}_s(t, \xi) \, d\xi \\ \tilde{\rho}_Y(t, \omega) &= \tilde{p}_Y(t, \omega) / \tilde{p}_Y(t, \Omega) & \rho_s(t, \omega) &= p_s(t, \omega) / p_s(t, \Omega) \\ D_\alpha(\tilde{\mathbb{P}}_Y, \mathbb{P}_s) &= \alpha \frac{1}{T} \int_0^T \frac{|\tilde{p}_Y(t, \Omega) - p_s(t, \Omega)|}{\tilde{p}_Y(t, \Omega) + p_s(t, \Omega)} \, dt + (1 - \alpha) \frac{1}{T} \int_0^T \int_0^\Omega |\tilde{\rho}_Y(t, \omega) - \rho_s(t, \omega)| \, d\omega \, dt \end{aligned}$$

In practice, we picked  $\alpha = 0$  in our evaluations, since the corresponding pure OT distance already gave us a reasonable way to quantify how well a tvPS reflected “its” itvPS, consistent with our (subjective) appraisals. In concrete computations, the integrals are approximated by sums of the corresponding discretized quantities.

**ESM-4c Parameter Selection.** In this subsection, we report the details of our exploration of the parameter space, leading to our choice of  $\beta = 30$ ,  $\gamma = 9$  and  $J = 2$  as the optimal one for the CWT-based ConceFT algorithm, when applied to the signal class  $\mathcal{C}$ .

We applied ConceFT to the noisy signals, with  $\gamma = 3, 4, \dots, 10$  (8 choices);  $\beta = 20, 30, \dots, 70$  (6 choices) and  $J = 1, 2, 3, 4$  (4 choices). All 192 possible combinations of these options are investigated. For each example and each parameter setting, we applied the ConceFT algorithm 10 times, each time with 10 random projections; the average of the OT distances over these 10 attempts was then computed.

Figure S.4 visualizes the results by means of a “heat map”. In this figure, the  $x$ -axis is the selection of  $\gamma$  and  $\beta$ , the  $y$ -axis is  $J$  and the color at each entry represents the averaged OT distance for the corresponding choice of the parameters  $\gamma, \beta$  and  $J$ ; the lighter the color, the smaller the OT distance and hence the better the performance. The Figure shows the averaged OT distance of the ConceFT result for all choices of parameters, for one signal  $s^\#$  in  $\mathcal{C}$ , and three types of noise, giving three heat maps in total. The  $x$ -coordinate in each heat map cycles through the 6 values of  $\beta$  before it moves on to a new value of  $\gamma$ ; Table S.1 below gives the value of  $x$  for each pair of Morse parameters considered.

	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 8$	$\gamma = 9$	$\gamma = 10$
$\beta = 20$	1	7	13	19	25	31	37	43
$\beta = 30$	2	8	14	20	26	32	38	44
$\beta = 40$	3	9	15	21	27	33	38	45
$\beta = 50$	4	10	16	22	28	34	39	46
$\beta = 60$	5	11	17	23	29	35	40	47
$\beta = 70$	6	12	18	24	30	36	41	48

TABLE S.1. The numbers on the  $x$  axis and their corresponding Morse parameters.

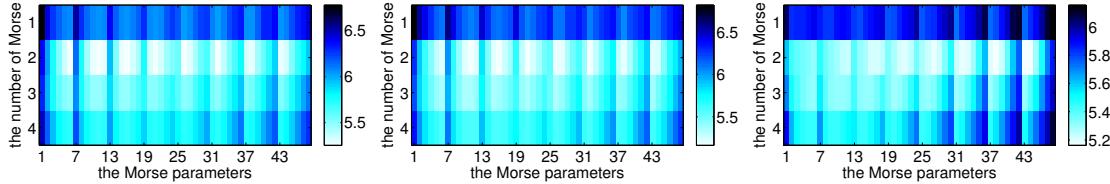


FIGURE S.4. Exploring the parameter space: heat maps visualizing the OT distance between the itvPS for a clean signal and the ConceFT-based tvPS of noisy versions, with SNR of 0 dB, for a two-component signal  $s^\#$  in  $\mathcal{C}$ , and for three different types of additive noise: Gaussian (left), ARMA(1,1) (middle) and Poisson (right). Each heat map shows the results for the 192 different parameter combinations described in the text. The color of each box represents the OT distance: lighter colors indicate better performance.

When we computed similar heat maps for other randomly picked signals in  $\mathcal{C}$ , the results were virtually identical. Our exploration showed that the combination  $\beta = 30, \gamma = 9, J = 2$  lead to the best performance; we thus chose these values for the remainder of the paper.

**ESM-4e. ConceFT results for noisy signals.** Figure S.5 is the analog of Figure 8 in the main paper, for  $s$ , the signal used to calibrate the parameter  $N$  for the ConceFT algorithm rather than the “new” signal  $s^*$ .

For each type of noise (Gaussian, ARMA(1,1) or Poisson) and each SNR considered (i.e.  $x$  dB, where  $x \in \{-7, -6, \dots, 6, 7\}$ ), 20 independent realizations of the noise process are considered; for each of the resulting noisy signals the ConceFT analysis and the OT-distance of the resulting tvPS to the itvPS of the clean signal are computed; the mean and the standard deviation for each are shown in Figure ESM-4.6.

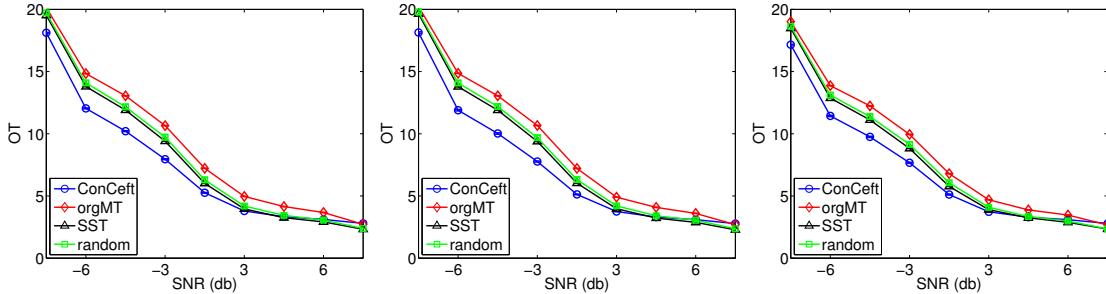


FIGURE S.5. OT distance of CWT-based ConCeFT results against signal to noise ratio (SNR) for the signal  $s(t)$ , and comparison with standard SST (with respect to the lowest-order Morse wavelet in black, and a random combination of the first two Morse wavelets in green) and standard multi-taper SST, both also CWT-based (see text). Noise type (left to right): Gaussian, ARMA(1,1), and Poisson. The ConCeFT result is the mean OT-distance for 20 independent ConCeFT computations; the standard deviation is smaller than the size of the marker.

Figure S.5 also compares the ConCeFT results with those of simple SST (using either the first Morse wavelet with parameters  $\beta = 30$ ,  $\gamma = 9$  as reference wavelet, or *one* random linear combination of the two first Morse wavelets) and of multi-taper SST (denoted as orgMT), using the same  $\psi_j$  as ConCeFT. For each of these alternate methods, we likewise computed the mean OT-distance of the tvPS to the itvPS for 20 noise realizations. (Note that the results are very similar to those in Figure 8 in the main paper, for  $s^*$ .)

**ESM-4.f ConCeFT with STFT.** As a complement to Figure 2 of the main paper, which shows STFT-based ConCeFT results and compares them with other SST-based algorithms (simple STFT-based SST with a Gaussian window, or multi-taper SST), we show below the results of STFT-based ConCeFT for the same signals  $s$  and  $s^*$  for which the main paper showed, in Figure 7, CWT-based ConCeFT results. Before we do this, we give some more details about how these STFT-based ConCeFT results are obtained.

We explain in the main paper that SST can be defined starting from a STFT just as well as from a CWT. The whole ConCeFT analysis, theoretical as well as numerical, can be carried out equally well using such STFT-SST representations. At the start of Section ESM-4, above, we explained the rationale for choosing Morse functions for the  $\psi_j$ ; this rationale leads similarly to the choice of Hermite functions as the natural basis windows  $h_j$  for STFT-based ConCeFT. As in the CWT case, the choice of the family is completely fixed by the values of two parameters; in the STFT case these correspond to the eccentricity of the elliptic localization domain in the TF-plane and the tilt of the major axis of this ellipse with the time-axis (see Figure S.1 bottom-left). Since there is no a priori reason to expect that chirping the Hermite functions in any direction (which tilting the elliptic domain would lead to) gives any advantage, we left this parameter out of consideration. The remaining parameter then corresponds to scaling the Hermite functions.

In analogy with the analysis in subsection EMS-4c, we thus explored the OT-distance of the STFT-based ConCeFT tvPS of signals in  $\mathcal{C}$  to their itvPS, for different rescalings and different numbers  $J$  of Hermite functions. Minimizing this OT-distance led us to picking Hermite functions for which the underlying Gaussian function was scaled so that the bandwidth  $h = 5/16$ ; that is, the Gaussian function is  $\frac{4}{\sqrt{5\sqrt{\pi}}} e^{-128t^2/25}$  (when measured in samples, since the sampling rate is 160Hz, this corresponds to an effective width of 600 samples, or 3.75 sec, for the window functions  $h_j$ ); the optimal number  $J$  of functions was 4. We then kept

these parameter choices for our further experiments. The number  $N$  of randomly picked linear combinations of the window functions was taken to be 20, as in the CWT case.

Once all the parameters are fixed, we can use the calibrated STFT-based ConceFT approach to study noisy versions of signals in  $\mathcal{C}$ . To compress dynamical range of the tvPS plots, we use the same trick as for Figure 7 of the main paper: we first reduce all the tvPS to the same total “energy”, by multiplying each discretized tvPS  $\tilde{\mathbb{P}}_Y \in \mathbb{R}^{m \times n}$  with an appropriate constant so that the “mean energy” of all entries equals the same number for all subfigures; that is, for some  $\theta > 0$  so that  $\frac{1}{nm} \sum_{k=1}^m \sum_{l=1}^n (\tilde{\mathbb{P}}_Y)_{k,l} = \theta$ . We take  $\theta = 5$ , as in the main paper, for Figures 7 and 9. Then we plot  $\mathbf{R} \in \mathbb{R}^{m \times n}$  rather than  $\tilde{\mathbb{P}}_Y \in \mathbb{R}^{m \times n}$  itself, where  $\mathbf{R}_{k,l} := \log(1 + \min\{(\tilde{\mathbb{P}}_Y)_{k,l}, q\})$ ,  $k = 1, \dots, m$ ,  $l = 1, \dots, n$  and  $q$  is the same cut-off as used in the main paper in Figures 7 and 9 (ensuring that the gray-scale value plots are all comparable).

Figure S.6 shows the results for STFT-based SST, with a Gaussian window with the bandwidth  $h = 5/16$ , on the (discretized) signal  $Y(t_k) = s(t_k) + \sigma \xi_k$ , where the  $\xi_k$  are i.i.d. realizations of a Gaussian noise process, and  $\sigma$  is chosen so that the SNR equals 0 dB; next, it shows the tvPS obtained with multi-taper SST, using the first 6 Hermite functions as window tapers (note that the Gaussian used for the simple SST is included among these; it is the lowest-order Hermite function); finally it shows the result of ConceFT, averaging SST results using 20 random linear combinations of those same 6 Hermite functions. In all three cases the OT-distance to the itvPS is given as well. Figure S.7 is entirely similar, but now for the signal  $s^*$  rather than  $s$ .

For Figures 7 and 9 in the main paper, and Figures S.6 and S.7, we used the MATLAB command `imagesc` to generate the plots of the different tvPS  $\tilde{\mathbb{P}}_Y$ , in the form `imagesc(log [1 + P_Y], [0 log(1+q)])`, where  $q > 0$ ; the two-entry array `[0 log(1 + q)]` in this expression ensures that the gray scale value at point  $(t, \xi)$  in the plot is linearly proportional to the value of  $\log [1 + \max(\tilde{\mathbb{P}}_Y(t, \xi), q)]$ , with white standing for 0 and black for  $q$ . The same value of  $q$  is used for all the tvPS plots. Plotting  $\log(1 + \tilde{\mathbb{P}}_Y)$  rather than  $\tilde{\mathbb{P}}_Y$  itself makes it possible to display a wider dynamical range; fixing the full gray-scale range to cover exactly `[0, log(1 + q)]` in each plot ensures that the figures present a fair visual comparison of the different tvPS. The maximum  $q$  also functions as a saturation cut-off: all values of  $\tilde{\mathbb{P}}_Y(t, \xi)$  exceeding  $q$  are rendered as black in the plots, regardless of the excess. The numerical value of  $q$  was picked so that the saturation cut-off is active on only an exceedingly low number of outliers.

Note that we have systematically normalized the  $\tilde{\mathbb{P}}_Y$  to have the same mean, which we picked to be 5 here. This value is not completely arbitrary: we picked it so that the dynamical-range compressing function  $\log(1 + \tilde{\mathbb{P}}_Y)$  makes the noise artifacts clearly visible. Different values are possible, and the choice depends on the applications; different types of signals and different desired visualization characteristics, typically correspond to different choices for  $\theta$ .

The other issue is the determination of a “natural” or “good” value for the cut-off  $q$ . For the purposes of this paper, where we wanted to give a visualization of the goodness-of-fit to the itvPS of a signal  $s$  of the tvPS for different noisy versions  $Y$  (obtained by adding to  $s$  different types of noise, possibly also of different strength), and compare these for different analysis methods and different noise types/strengths, it is most natural to fix a uniform value of  $q$  (after uniform normalization of the  $\tilde{\mathbb{P}}_Y$ ). When ConceFT is used in practice, however, we expect to use one particular analysis method, to have at hand only one realization of the (unknown) noise process, and (of course) not to have a ground truth with which to compare. An important role of  $q$ , for the rescaling used by `imagesc` for the visual display, is to downplay an otherwise exaggerated impact from outliers. To determine  $q$ , based only on  $\tilde{\mathbb{P}}_Y$  itself, it would thus be natural to choose it as some fixed percentile of the distribution of values of  $\tilde{\mathbb{P}}_Y$ .

Figures S.8 through S.10 show the *same*  $\tilde{\mathbb{P}}_Y$  as also plotted in Figure 7 in the main paper (for CWT) and Figures S.6 and S.7 (for STFT), but with a plotting scheme that corresponds more to the realistic signal analysis situation, where we only have the signal at hand. More precisely, for Figures S.8 through S.10,

we do not normalize the tvPS to have the same total ‘‘energy’’, and we determine the cut-off  $q$  for each  $\tilde{P}_Y$  individually; we plot  $\mathbf{R} \in \mathbb{R}^{m \times n}$ , which is defined as  $\mathbf{R}_{k,l} := \log(1 + \min\{(\tilde{P}_Y)_{k,l}, q\})$ ,  $k = 1, \dots, m$ ,  $l = 1, \dots, n$ , where for each plot, the value  $q$  is given by the 99.8% percentile of only the  $\tilde{P}_Y$  of the transform/data for that individual plot itself. The difference between the two cases is striking, especially for the STFT figures. This suggests a more thorough exploration would be useful of how to optimally pick  $q$  depending on noise and signal structure and on analysis method chosen; this is beyond the scope of this paper, however.

## REFERENCES

- [1] B. Babadi and E. N. Brown. A review of multitaper spectral analysis. *IEEE Trans. Biomed. Eng.*, 61(5):1555–1564, 2014.
- [2] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu. Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *J. Roy. Stat. Soc. B*, 76:651–682, 2014.
- [3] I. Daubechies. Time-frequency localization operators: a geometric phase space approach. *IEEE Trans. Inform. Theory*, 34:605–612, 1988.
- [4] I. Daubechies, J. Lu, and H.-T. Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.*, 30:243–261, 2011.
- [5] I. Daubechies and T. Paul. Time-frequency localization operators: a geometric phase space approach II. The use of dilations. *Inverse Problems*, 4(3):661–680, 1988.
- [6] R. Gallager. Circularly-Symmetric Gaussian random vectors. January 2008. <http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf>.
- [7] I. Gel'fand and N. Ya. Vilenkin. *Generalized function theory Vol 4*. Academic Press, 1964.
- [8] S. C. Olhede and A. T. Walden. Generalized Morse Wavelets. *IEEE Trans. Signal Process.*, 50(11):2661–2671, 2002.
- [9] D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge University Press, 1993.
- [10] D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70:1055–1096, 1982.
- [11] C. Villaniic. *Topics in Optimal Transportation*. Graduate Studies in Mathematics, American Mathematical Society, 2003.
- [12] J. Xiao and P. Flandrin. Multitaper Time-Frequency Reassignment for Nonstationary Spectrum Estimation and Chirp Enhancement. *IEEE Trans. Signal Process.*, 55:2851–2860, 2007.

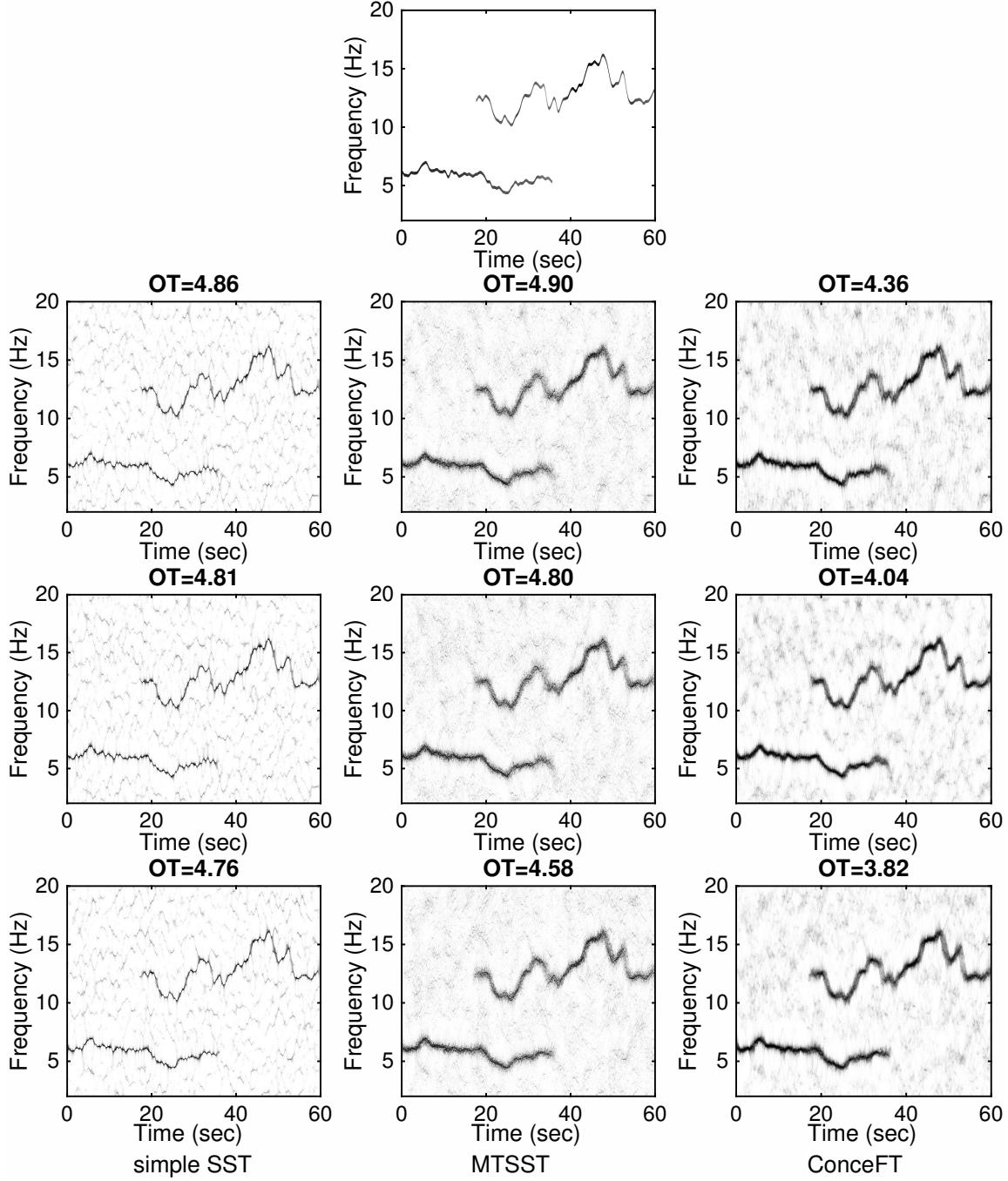


FIGURE S.6. STFT-based results for noisy versions of the same signal  $s$  as in Figures 3, 4 and the top half of Figure 7 in the main paper; the noisy versions considered are also the same realizations as in Figure 4 in the main paper. Top: tvPS of the clean signal  $s$ ; next three columns: the tvPS of  $Y$  with different noises. Each column corresponds to one algorithm; each row to one noise type. Noise types: from top to bottom, in order: Gaussian, ARMA(1, 1) and Poisson noise, in each case with SNR of 0 dB. Different approaches: from left to right, in order: SST using a Gaussian window with the bandwidth  $h = 5/16$ ; STFT-based multi-taper-SST (using the top 6 Hermite functions: the same Gaussian again, and the next 5 Hermite functions); STFT-based ConceFT (using 20 random combinations of the top 4 Hermite functions).

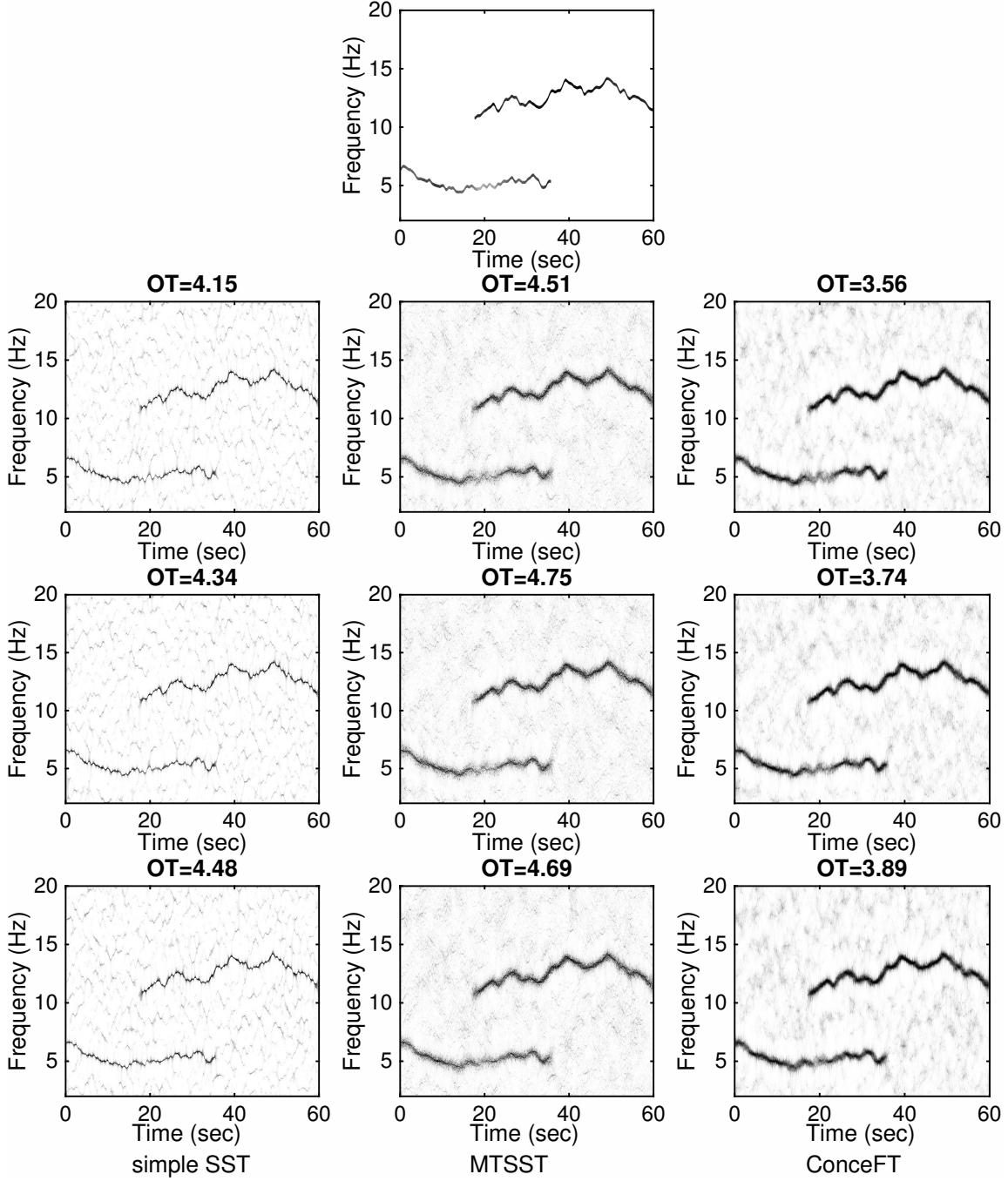
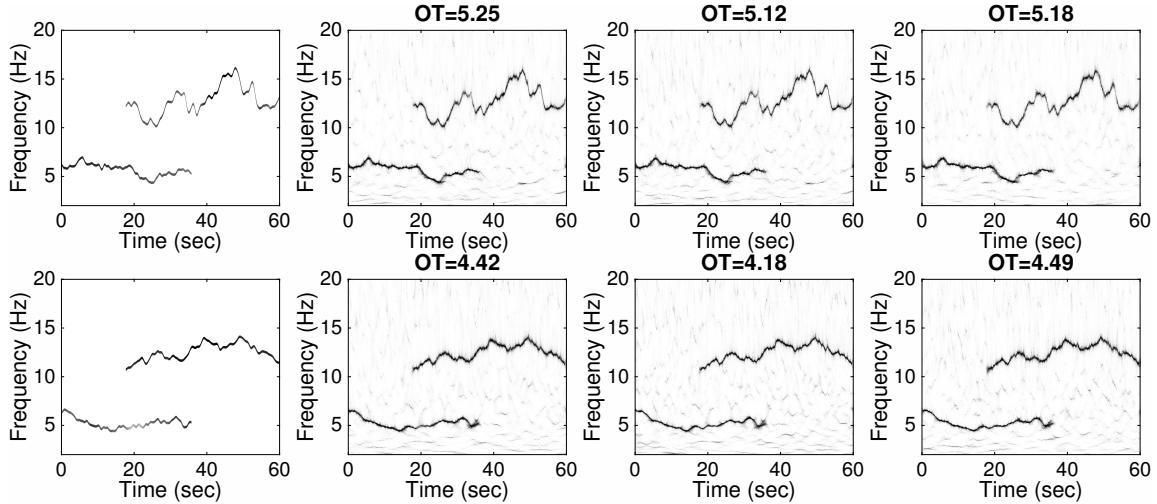
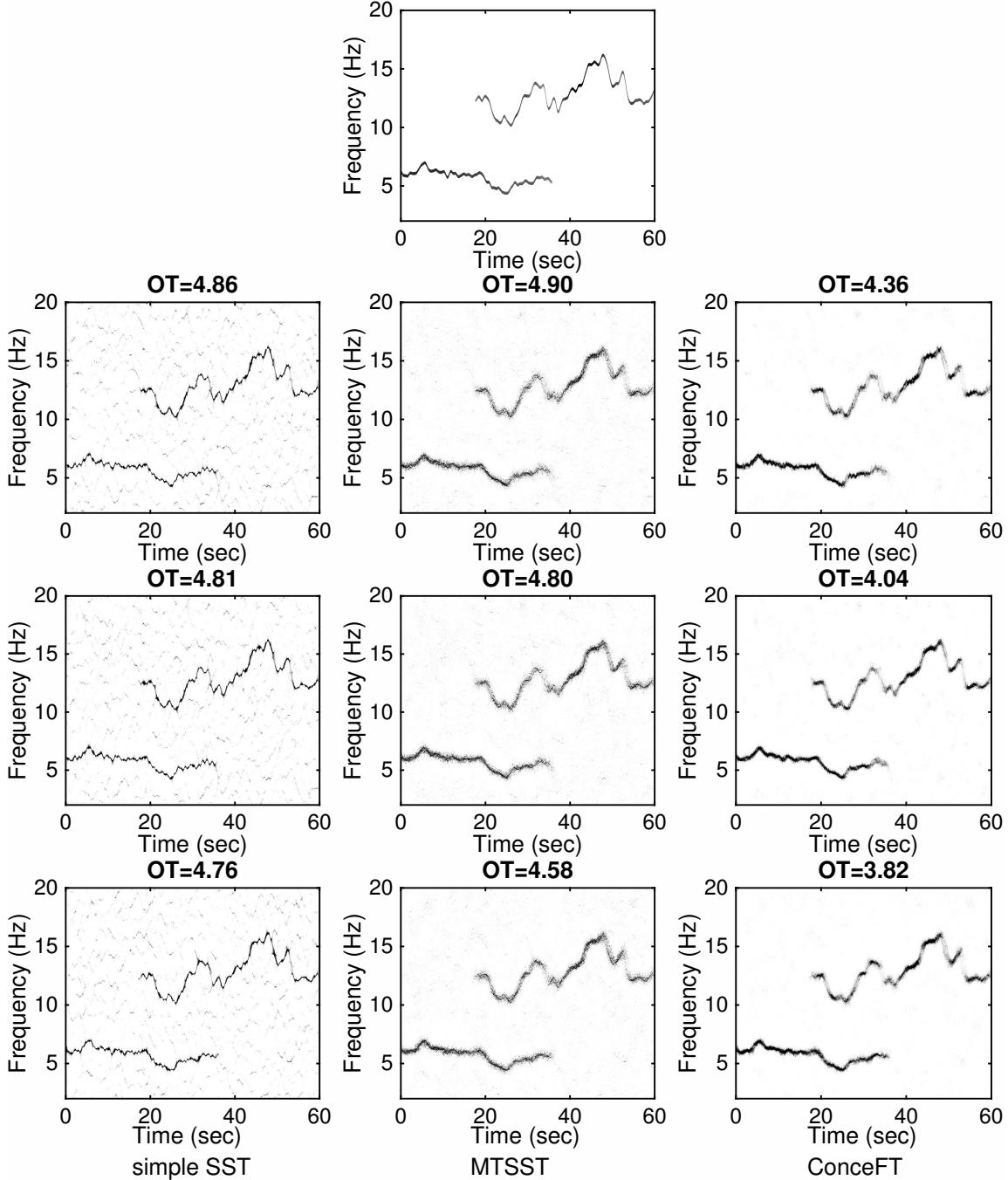


FIGURE S.7. STFT-based results for noisy versions of the same signal  $s^*$  as in Figures S.2, S.3 and the bottom half of Figure 7 in the main paper; the noisy versions considered are also the same realizations as in Figure S.3. Top: itvPS of the clean signal  $s^*$ ; next three columns: the tvPS of  $Y^*$  with different noises. Each column corresponds to one algorithm; each row to one noise type. Noise types: from top to bottom, in order: Gaussian, ARMA(1, 1) and Poisson noise, in each case with SNR of 0 dB. Different approaches: from left to right, in order: SST using a Gaussian window with the bandwidth  $h = 5/16$ ; STFT-based multi-taper-SST (using the top 6 Hermite functions; the same Gaussian again, and the next 5 Hermite functions); STFT-based ConceFT (using 20 random combinations of the top 4 Hermite functions).



**FIGURE S.8.** This figure contains CWT-based ConceFT of the same signal  $s$  as in Figure 7 in the main paper, but with a different truncation plotting strategy (see text). First row: results for the signal  $s$ ; second row: results for the signal  $s^*$ . Left to right: ideal time-varying TF power spectrum (itvPS) for the clean signal, followed by results of ConceFT with Morse wavelets after (in order) Gaussian, ARMA(1,1) or Poisson noise was added, with SNR of 0 dB. The figures are plotted with  $q$  chosen to be the 99.8% quantile of each figure, and without normalizing  $\tilde{P}_Y$ . For each of the tvPS panels, the header gives the OT distance to the corresponding itvPS, which is the same as before, since the  $\tilde{P}_Y$  are the same.



**FIGURE S.9.** STFT-based ConceFT results for noisy versions of the same signal  $s$  as in Figures S.6 and the top half of Figure 7 in the main paper; the noisy versions considered are also the same realizations as in Figures 3 and 4 in the main paper. The plotting strategy is different, however (see text). Top: itvPS of the clean signal  $s$ ; next three columns: the tvPS of  $Y$  with different noise types. Each column corresponds to one algorithm; each row to one noise type. Noise types: from top to bottom, in order: Gaussian, ARMA(1, 1) and Poisson noise, in each case with SNR of 0 dB. Different approaches: from left to right, in order: SST using a Gaussian window with the bandwidth  $h = 5/16$ ; STFT-based multi-taper-SST (using the top 6 Hermite functions: the same Gaussian again, and the next 5 Hermite functions); STFT-based ConceFT (using 20 random combinations of the top 4 Hermite functions). The figures are plotted with  $q$  chosen to be the 99.8% quantile of each figure, and without normalizing  $\tilde{P}_Y$ . For each of the tvPS panels, the header gives the OT distance to the corresponding itvPS, which is the same as before, since the  $\tilde{P}_Y$  are the same.

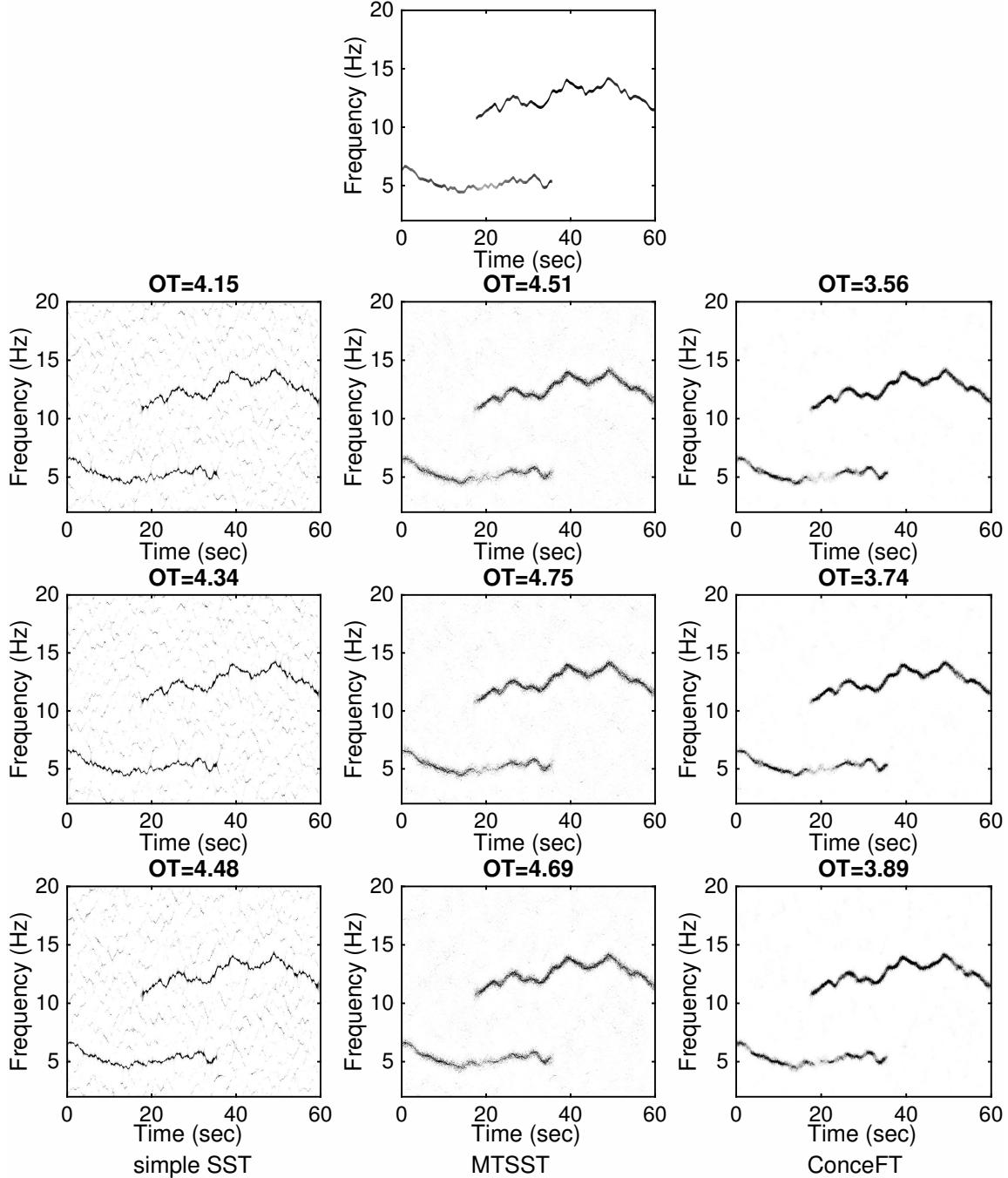


FIGURE S.10. STFT-based ConceFT results for noisy versions of the same signal  $s^*$  as in Figures S.7 and the bottom half of Figure 7 in the main paper; the noisy versions considered are also the same realizations as in Figures S.2 and S.3. The plotting strategy is different, however (see text). Top: itvPS of the clean signal  $s^*$ ; next three columns: the tvPS of  $Y^*$  with different noises. Each column corresponds to one algorithm; each row to one noise type. Noise types: from top to bottom, in order: Gaussian, ARMA(1, 1) and Poisson noise, in each case with SNR of 0 dB. Different approaches: from left to right, in order: SST using a Gaussian window with the bandwidth  $h = 5/16$ ; STFT-based multi-taper-SST (using the top 6 Hermite functions: the same Gaussian again, and the next 5 Hermite functions); STFT-based ConceFT (using 20 random combinations of the top 4 Hermite functions). The figures are plotted with  $q$  chosen to be the 99.8% quantile of each figure, and without normalizing  $\tilde{P}_Y$ . For each of the tvPS panels, the header gives the OT distance to the corresponding itvPS, which is the same as before, since the  $\tilde{P}_Y$  are the same.