A New Approach to Fourier Synthesis With Application to Neural Encoding and Speech Classification

Steven Kay, *Fellow, IEEE*

***Abstract*—We describe a novel means of representing signals by a Fourier decomposition consisting of complex sinusoids with unit amplitudes and zero phases. The only information necessary to reconstruct the signal from its Fourier components consists of the "place" information, which specifies the sinusoidal frequencies to include in the synthesis. This set of frequencies results in a nonuniform distribution of sinusoidal frequency components. As such, the approach provides a means of representing a signal by a set of zeros and ones, indicating an off-on condition for each frequency component. It is conjectured that this might help explain the mechanism of auditory and visual neural encoding of acoustic and visual stimuli, respectively. As an immediate application of the theory, a classification experiment is conducted which indicates that the proposed neural encoding is more robust to noise than traditional approaches.**

***Index Terms*— Discrete Fourier transforms, multiple speech classification, speech coding.**

I. Introduction

ALTHOUGH the mechanisms of human hearing have been studied for many years [1], we still do not understand how the ear works in totality. The transduction of a sound wave into a mechanical excitation in the cochlea is fairly well understood, but how the information is actually encoded and interpreted by the brain is a mystery. The usual explanation, that the cochlea and its frequency sensitive hair cells act as a bank of narrowband filters, appears to be the accepted model. Yet, there remain many questions as to why a supposedly linear bank of filters acts in nonlinear ways as described in [11]. For example, it is known from physiological experiments that as the sound stimulus increases in amplitude, "the fiber tuning curves broaden, activity will spread to fibers between the peaks" [11]. This is tantamount to saying that the filter bank consists of a set of filters whose bandwidths change with the amplitude of the input. Clearly, this is not possible if the filters are linear and time invariant. Even more problematic is that the ultimate output of these filters is a neural firing response, either on or off, i.e., binary, that is transmitted via the auditory nerve to the brain. One wonders whether the filter bank model followed by some decision device is the appropriate model or just a convenient one spawned by appealing to well-known signal models and linear systems theory. The current state of knowledge in this area is summarized in [2] with other papers listed in [3]. It appears, however, that many of the model inconsistencies have still not been resolved.

In this paper we propose a new approach to the synthesis of a deterministic signal using Fourier frequency components. For random signals it has been shown that an analogous Fourier synthesis can be implemented using random frequency sinusoids [4]. It is worthwhile to contrast the proposed approach with some of the many available synthesis methods. Typical ones are wavelets, time-frequency representations, sine-cosine transforms, etc [9]. All these methods rely on synthesizing signals using a *weighted linear* combination of *fixed basis functions*. The set of basis functions depends upon the transform chosen. In our approach we use a dense set of basis functions, where each basis function is either included in the sum or not. No weighting is used and consequently, the "transform" is nonlinear. A wavelet transform, for example, maps a signal to a set of weighting coefficients. The sum of two signals produces the sum of weighting coefficients. The approach to be described, however, produces a set of zeros and ones as "weights", with the sum of two signals producing a different distribution of zeros and ones.

II. The New Synthesis Approach

For simplicity we will consider the synthesis of a discrete-time deterministic signal. However, the theory to be presented allows for an easy extension to continuous time signals as well as for multidimensional signals. Considering a discrete-time signal denoted by $x[n]$ for $-\infty < n < \infty$, we have the usual Fourier synthesis integral

$$x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)\exp(j2\pi fn)df \tag{1}$$

where $X(f)$ is the discrete-time Fourier transform given by $X(f) = \sum_{n=-\infty}^{\infty} x[n]\exp(-j2\pi fn)$ for $-1/2 \le f \le 1/2$.

What is obvious from (1) and what every signal processing practitioner knows, is that a Fourier synthesis involves *summing together complex sinusoids with given amplitudes and phases*. However, what is not so obvious is that $x[n]$ can be synthesized with complex sinusoids with *unit amplitudes and zero phases*. To do so will require a *nonuniform distribution of frequency components*, which entails the use of a Fourier-Stieltjes integral (see [8] for a introduction to Stieltjes integration and [7] for Fourier-Stieltjes integration). To make the subsequent discussion more intuitive consider the Riemann integration of the

Manuscript received May 03, 2010; revised July 14, 2010; accepted July 18, 2010. Date of publication July 26, 2010; date of current version August 16, 2010. This work was supported by the Air Force Office of Scientific Research under the grant FA9550-09-1-0064. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kenneth E. Barner.

The author is with Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881 USA (e-mail: kay@ele.uri.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2010.2060721

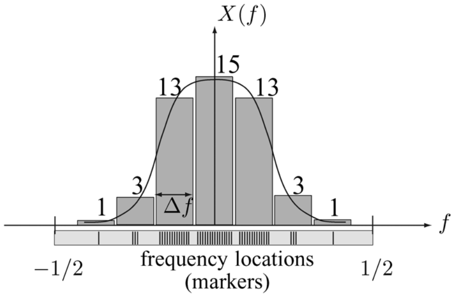

Fig. 1. Alternative viewpoint of Riemann integration.

function shown in Fig. 1. We may view this as the calculation of $x[0]$, which is

$$x[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)df$$

and where $X(f)$ is real and positive. Referring to Fig. 1 we see that the value of the integral is obtained by summing up the areas of the approximating rectangles. The width of each rectangle is Δf so the approximate value is $1 \cdot \Delta f + 3 \cdot \Delta f + 13 \cdot \Delta f + 15 \cdot \Delta f + 13 \cdot \Delta f + 3 \cdot \Delta f + 1 \cdot \Delta f = 49\Delta f$ and clearly as $\Delta f \to 0$, we will obtain the true value of $x[0]$ as a limit. Alternatively, we can think of a very fine grid of frequencies over the interval $-1/2 \le f \le 1/2$ and compute the integral by placing "markers" with each marker having the same value of Δf at the frequency locations shown. In the absence of a marker, we can think of that sinusoidal frequency component as not appearing or equivalently of having an amplitude of zero. By using either the customary approach to Riemann integration as a limit of the areas of rectangles or as the limit of a set of nonuniformly spaced markers, we obtain the same result.

In the case of Fourier synthesis we will sum a set of unit amplitude and zero phase sinusoids whose frequencies are nonuniformly spaced, with Fig. 1 serving as an example. The fact that the neighboring frequencies corresponding to a given rectangle are slightly different for each marker will disappear in the limit as $\Delta f \to 0$ or as the frequencies become more densely spaced. In summary, a signal may be constructed and therefore represented by a set of complex sinusoids, whose amplitudes are one, phases are zero, and *whose frequencies are chosen so that when summed over each small Δf bin will yield the amplitude of the Fourier transform*. Since the Fourier transform amplitude $X(f)$ is in general not constant, neither is the density of the frequencies.

III. The Mathematical Development: Fourier Transform Real and Positive

We restrict the Fourier transform to be real and positive. The extension to more general Fourier transforms will be described in Section IV. Hence, we have that $X(f) > 0$, where $X(f)$ is real. First we define the *integrated Fourier transform*, which may be thought of as a cumulative distribution function (CDF), although $X(f)$ need not integrate to one over $-1/2 \le f \le 1/2$, as

$$F(f) = \int_{-1/2}^{f} X(\xi)d\xi \qquad -1/2 \le f \le 1/2. \tag{2}$$

Because $X(f)$ is real and positive, it is easily shown that $F(f)$ must be a *strictly* monotonically increasing function. That is to say, for $f_2 > f_1$, $F(f_2) > F(f_1)$. Each distinct value of $F(f)$ is associated with a distinct value of f. As a result, the inverse function $F^{-1}(\cdot)$ exists and if $u = F(f)$, then

$$f = F^{-1}(u) \qquad 0 = F(-1/2) \le u \le F(1/2). \tag{3}$$

Next, since $F(f)$ is absolutely continuous, it is well known that the derivative of $F(f)$ is $F'(f) = X(f)$ [8] and this allows us to write (1) as

$$x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f)\exp(j2\pi fn)df \tag{4}$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(j2\pi fn)dF(f) \tag{5}$$

since $dF(f) = F'(f)df = X(f)df$. The latter integral is known as a Stieltjes integral or more specifically, because of the integrand being $\exp(j2\pi fn)$, as a *Fourier-Stieltjes integral*. It is known to exist for all $F(f)$ of bounded variation and an integrand that is continuous [8]. Since $F(f)$ is monotonically increasing, it will be of bounded variation (assuming $F(1/2)$ is finite). This will be sufficient for our present purposes but note that much more general forms of $F(f)$ can be accommodated, including jumps that can model pure sinusoidal components.

Now in (5) we use a change of variables given by $u = F(f)$. Since the inverse of F exists, we have that $f = F^{-1}(u)$ and therefore $x[n] = \int_{F(-1/2)}^{F(1/2)} \exp[j2\pi F^{-1}(u)n]du$ and since $F(-1/2) = 0$, we have our final result

$$x[n] = \int_{0}^{F(1/2)} \exp\left[j2\pi F^{-1}(u)n\right] du. \tag{6}$$

We see that the synthesis of $x[n]$ can be effected by "summing" together unit amplitude, zero phase complex sinusoids with frequencies $f = F^{-1}(u)$ for $0 \le u \le F(1/2)$. If the integral is discretized with uniform u spacing, then the sinusoidal frequencies are *nonuniformly distributed* as will be illustrated next. Finally, the assumption that $F(1/2)$ is finite is easily translated into the practical constraint $F(1/2) = \int_{-1/2}^{1/2} X(\xi)d\xi = x[0] < \infty$, which of course will be satisfied.

IV. An Example

Assume that $X(f) = 12f^2$, noting that it is real, positive, and even. It is easily shown that the corresponding signal is given by

$$x[n] = \begin{cases} 1 & n = 0 \\ \frac{6}{n^2\pi^2}(-1)^n & n \neq 0. \end{cases} \tag{7}$$

To obtain the alternative representation we first find the integrated Fourier transform as

$$F(f) = \int_{-1/2}^{f} 12\xi^2 d\xi = 4\left(f^3 + \frac{1}{8}\right).$$

Setting this equal to u to yield $u = F(f)$, the inverse transformation becomes

$$f = F^{-1}(u) = \left(\frac{u}{4} - \frac{1}{8}\right)^{1/3} \qquad 0 \le u \le F(1/2) = 1.$$

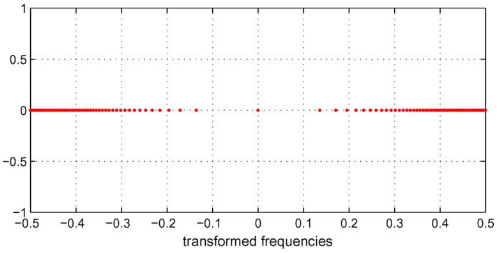

Fig. 2. Discretized frequency components given by (11) for alternative synthesis expression using $N_f = 100$.

The synthesis integral becomes from (6)

$$x[n]=\int_0^1 \exp\left[j2\pi\left(\frac{u}{4}-\frac{1}{8}\right)^{1/3} n\right]du=\int_{-\frac{1}{2}}^{\frac{1}{2}} \cos\left[2\pi\left(\frac{v}{4}\right)^{1/3} n\right]dv.$$

where the last step is due to the odd symmetry of the sine function.

In summary, the usual synthesis expression is replaced by either

$$x[n] = \int_0^1 \cos\left[2\pi\left(\frac{u}{4}-\frac{1}{8}\right)^{1/3} n\right] du \qquad (8)$$

or

$$x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos\left[2\pi\left(\frac{v}{4}\right)^{1/3} n\right] dv \qquad (9)$$

where it is seen that the sinusoids have unit amplitudes, zero phases, and transformed frequencies given by $f = (v/4)^{1/3}$ for $-1/2 \leq v \leq 1/2$.

Next to show that this produces the same $x[n]$ signal, we first discretize (8) by choosing $\Delta u = 1/N_f$, where N_f is the number of frequencies. This yields the Riemann sum approximation of

$$x[n] \approx \frac{1}{N_f}\sum_{k=0}^{N_f} \cos(2\pi f_k n) \qquad (10)$$

where the nonuniform frequencies are

$$f_k = \left(\frac{k/N_f}{4} - \frac{1}{8}\right)^{1/3} \qquad k = 0, 1, \ldots, N_f \qquad (11)$$

and span the interval [−1/2, 1/2]. For $N_f = 100$ these frequencies are shown by dots in Fig. 2. As expected, there are very few frequencies near $f = 0$ since the Fourier transform $X(f) = 12f^2$ is small there but increases rapidly as f increases. The behavior is just the opposite of that shown in Fig. 1 due to the type of $X(f)$ used.

Finally, we synthesize the signal using (10) and (11) for $N_f = 100$. The results are shown in Fig. 3. The top plot is the true signal given by (7) and the bottom plot displays the synthesized signal.

We have assumed that the Fourier transform was real and positive. This allowed us to integrate it to yield a strictly monotonically increasing function. More generally, though, the Fourier transform has real and imaginary parts, each of which may not be positive. Integrating these parts then will not produce a monotonically increasing function that is necessary for the existence of an invertible distribution function. A standard approach in this situation is to decompose each real function into its positive and negative parts. This will allow us to extend the previous results to the more general Fourier transform case. Mathematically, we have $X(f) = X^+(f) - X^-(f)$, where $X^+(f) \geq 0$ and $X^-(f) > 0$. Note that this decomposition is easily accomplished using $X^+(f) = (|X(f)| + X(f))/2$ and $X^-(f) = (|X(f)| - X(f))/2$. Next, break up $X(f)$ into its real part $X_R(f)$ and imaginary part $X_I(f)$ and then break up each of these into its positive and negative parts, i.e., $X_R^+(f), X_R^-(f), X_I^+(f), X_I^-(f)$. The previous results can now be applied.

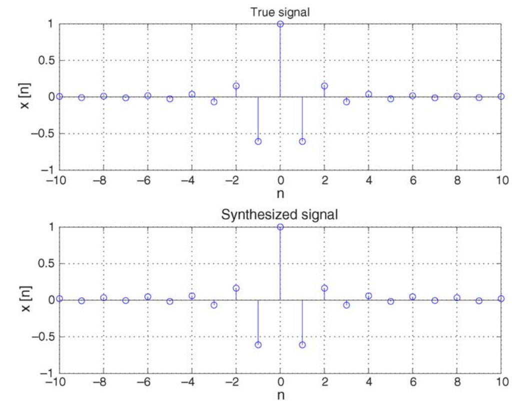

Fig. 3. True signal and synthesized signal using $N_f = 100$ frequency components with unit amplitudes and zero phases.

V. Possible Explanation for Neural Encoding

The output of the cochlea in response to an acoustic waveform at the input to the ear is a sequence of neuron firings that are transmitted to the brain. These neuron outputs, which are either on or off, can be thought of as the "markers" shown in Fig. 1. This is the so called "place" information [12]. Only the cochlea hair cells that are tuned to the frequencies in the waveform produce a firing rate above the quiescent value. Hence, our model would seem to be able to explain, at least to some degree, the cochlear mechanism. For instance, the nonlinear response described earlier as a "spreading" of the fiber firings might be due to the fact that in our model each hair cell can only contribute a unit of amplitude, and hence for different waveform amplitudes it may be necessary for neighboring hair cells to fire; thus, adding to the overall amplitude in frequency.

VI. Classification Based on Neural Encoding

We next give an application of the proposed neural encoding. A computer simulation is presented to compare several traditional methods of speech classification for two voiced speech-like signals. As such the goal is to compare the probability of correct classification. To indicate the robustness of the neural approach we add varying amounts of Laplacian noise to the signals. The three methods to be compared are the asymptotic maximum likelihood (ML) method, the cepstral distance metric, and a method based on neural encoding. The classifiers are all based on the measured power spectrum in accordance with the ear being relatively insensitive to phase. This allows the theory previously described to be applied to a synthesis of the autocorrelation in terms of its positive Fourier transform, the power spectral density. Specifically, we consider the two signals whose power spectral densities are shown in Fig. 4. In accordance with the ability of the ear to understand sounds of varying levels,

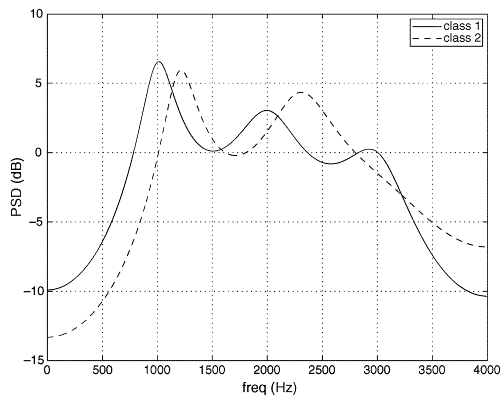

Fig. 4. Power spectral densities of the two signals used to model two different voiced speech sounds.

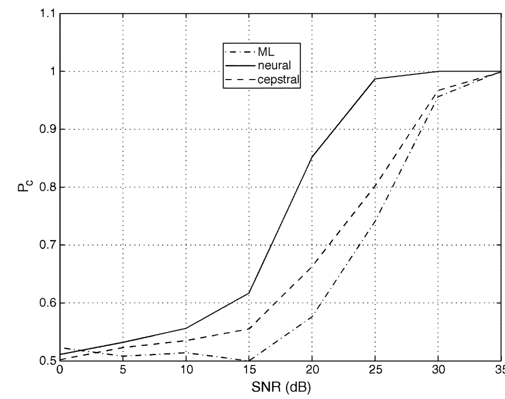

Fig. 5. Probability of correct classification of the three classifiers versus SNR.

we assume that the signal power is unknown in formulating the classifiers. The signal spectra are both autoregressive (AR) spectra with 6 poles and are meant to model typical voiced speech sounds. However, the AR model is not used as prior information in any of the classifiers, but only to generate the data. The classification methods are briefly described next.

The asymptotic (large data record) maximum likelihood method chooses the signal which maximizes the asymptotic likelihood function [5]

$$T_{\mathrm{ML}}[i] = -\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln Q_i(f)df - \ln \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{Q_i(f)}df \quad i = 1, 2$$

where $Q_i(f)$ is the assumed known power spectrum, normalized to unity power, of the ith class (as shown in Fig. 4) and $I(f)$ is the periodogram. This statistic is derived from a random process AR model and *so does not account for the Laplacian observation noise*. Since the overall power of the signal is assumed unknown, it is estimated as part of the statistic. In fact, all the classifiers produce decisions that are scale invariant, not depending on knowledge of the signal power. The next method utilizes the cepstral measure, a common choice for speech recognition [10]. It chooses the signal that minimizes the cepstral spectral distance measure

$$d[i] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\ln \bar{I}(f) - \ln Q_i(f)\right)^2 df \quad i = 1, 2$$

where $\bar{I}(f)$ is the unity power normalized periodogram, which ideally should be close to $Q_i(f)$ in the absence of observation noise. Finally, the neural classifier chooses the class that maximizes

$$T_{\mathrm{neural}}[i] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln Q_i\left(F^{-1}(u)\right) du \approx \frac{1}{N_f} \sum_{k=0}^{N_f} \ln Q_i(f_k) \quad i = 1, 2$$

where $F(f) = \int_{-1/2}^{f} \bar{I}(\xi) d\xi$ and the f_k's are the nonuniformly spaced frequencies. The neural classifier evaluates the log spectrum in accordance with the logarithmic nature of hearing and uses as input the nonuniformly spaced frequency encoded periodogram. It should be noted that the encoding is nonlinear and so the addition of noise to the signal is reflected in a nonlinear way to produce the frequency positions.

For the two signals we choose $N = 80$ samples and use a pitch period of $P = 25$ samples. To each signal we add enough Laplacian noise [6] to achieve a given signal-to-noise ratio (SNR). The probability of correct classification is estimated by generating one of the two signals with equal probabilities, adding the Laplacian noise, and finally determining the fraction of correct decisions for 1000 independent realizations. The entire procedure is replicated for each SNR. The results are shown in Fig. 5. It is clear that of the three classifiers the proposed neural approach is the most robust with respect to SNR. This is in agreement with the remarkable ability of the ear to discern different speech sounds even in noisy environments.

VII. Conclusions

A new method of Fourier synthesis has been presented that allows the composing sinusoids to have amplitudes of ± 1 and zero phases. It is conjectured that such a representation may be inherent in neural encoding since only impulses are transmitted to the brain as a result of auditory or visual stimuli. Preliminary application to classification indicates a marked improvement using neural encoding.

References

[1] J. L. Flanagan, *Speech Analysis, Synthesis, and Perception*. New York: Springer-Verlag, 1972.

[2] B. Delgutte, "Physiological models for basic auditory precepts," in *Auditory Computation*. New York: Springer, 1996.

[3] R. R. Fay and A. N. Popper, *Springer Handbook of Auditory Research*. New York: Springer, 1996, vol. 1–7.

[4] S. Kay, "Representation and generation of NonGaussian wide sense stationary random processes with arbitrary PSDs and a class of PDFs," *IEEE Trans. Signal Process.*, pp. 3448–3458, July 2010.

[5] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[6] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[7] E. K. Haviland and A. Wintner, "On the Fourier-Stieltjes transform," *Amer. J. Math.*, pp. 1–7, 1934.

[8] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*. New York: Dover, 1970.

[9] A. Poularikis, *Transforms and Applications*, 3rd ed. Boca Raton, FL: CRC, 2010.

[10] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[11] M. B. Sachs and E. D. Young, "Effect of nonlinearities on speech encoding in the auditory nerve," *J. Acoust. Soc. Amer.*, pp. 858–875, Sept. 1980.

[12] W. M. Siebert, "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," *Proc. IEEE*, pp. 723–730, May 1970.