Sinewave Representations of Nonmodality

Nicolas Malyska, Thomas F. Quatieri, and Robert Dunn

MIT Lincoln Laboratory
244 Wood Street, Lexington, MA 02420

[nmalyska, quatieri, rbd]@ll.mit.edu

ABSTRACT

Regions of nonmodal phonation, exhibiting deviations from uniform glottal-pulse periods and amplitudes, occur often and convey information about speaker- and linguistic-dependent factors. Such waveforms pose challenges for speech modeling, analysis/synthesis, and processing. In this paper, we investigate the representation of nonmodal pulse trains as a sum of harmonically-related sinewaves with time-varying amplitudes, phases, and frequencies. We show that a sinewave representation of any impulsive signal is not unique and also the converse, i.e., frame-based measurements of the underlying sinewave representation can yield different impulse trains. Finally, we argue how this ambiguity may explain addition, deletion, and movement of pulses in sinewave synthesis and a specific illustrative example of time-scale modification of a nonmodal case of diplothemia.

Index Terms: sinewave analysis, sinewave synthesis, nonmodal phonation

1. INTRODUCTION

Nonmodal speech is characterized by nonuniform glottal pulses that can occur during phonation due to irregular vibration of the vocal cords. Nonmodality in normal speech is important not only because of its potential dialect, language, and speaker dependence [1] but also because it taxes speech excitation modeling, specifically the addition, deletion, and movement of pulses. An example of such a phenomenon is shown in Figure 1 using standard sinewave analysis/synthesis [3].

![Figure 1: Generation of an unwanted pulse in sinewave analysis/synthesis: (left) original; (right) synthesis.](image)

It is surprising that the sinewave synthesis of nonmodal impulses can result in effects that are more similar to modifications of the original than to distortion such as pulse smearing. Ultimately, we seek then to understand how a pulse can be added, deleted, or moved in synthesis relative to its position in the source signal. Section 2 of this paper provides conditions on the fundamental sinewave phase to maintain the impulsiveness of harmonic signals. Section 3 then focuses on the problem of uniqueness and shows that different frequency and phase trajectories can result in the same impulse train, for both modal and nonmodal cases. The converse problem, where measurements of the underlying sinewave representation can yield different impulse trains, is also introduced. Examples are given illustrating both the issue of uniqueness and its converse. In Section 4, we discuss extensions to real speech including analysis/synthesis, bandlimiting, and vocal-tract filtering, and illustrate our theory with the application of time-scale modification of speech. Section 5 gives conclusions and needed future studies.

2. CONDITION FOR IMPULSIVENESS

This section presents conditions that relate a series of continuous-time impulses to a sinusoidal model consisting of harmonically-related sinusoidal components.

2.1. Definitions

The class of signals that we consider in this paper involves harmonically-related instantaneous frequencies. At a given time, t, each harmonic, denoted by the index k, has the instantaneous frequency,

\[ \dot{\theta}_k(t) = \theta_k(t) \]

where \( \dot{\theta}_k(t) \) is termed the instantaneous fundamental frequency. Using this definition, we can find the instantaneous phase of each harmonic component by integrating (1), yielding

\[ \theta_k(t) = \int_{-\infty}^{t} \dot{\theta}_k(\tau)d\tau = k \int_{-\infty}^{t} \dot{\theta}_k(\tau)d\tau \]

This work was sponsored by the United States Air Force Research Laboratory under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.
The output signal, \( s(t) \), created using the sum of an infinite number of such components, each having constant amplitude \( C_k \), is

\[
s(t) = \sum_{k=1}^{\infty} C_k \cos(\theta_k(t)) = \sum_{k=1}^{\infty} C_k \cos(k\theta(t)). \tag{3}
\]

If we further restrict all \( C_k \) to be a constant, \( C \), we obtain

\[
s(t) = C \sum_{k=1}^{\infty} \cos(k\theta(t)). \tag{4}
\]

For a constant fundamental frequency, \( \dot{\theta}_a(t) = f_o \), it is clear that (4) gives a periodic series of impulses. With time-varying fundamental frequency, on the other hand, it is not clear if (4) yields a series of impulses and, if so, where the impulses lie in time. In the following section, we show that indeed (4) is always an impulse train (generally nonmodal) and we will provide a condition for where in time the impulses occur.

### 2.2. Condition

We now show that for any phase \( \theta_a(t) \neq 2\pi m \), where \( m \) is an integer, our time function (4) is guaranteed to be zero. We will thus show that (4) always yields a train of impulses, with \( m \) being the positive or negative index for the \( m \)th impulse in the series. That is, our objective is to show:

\[
s(t) = C \sum_{k=1}^{\infty} \cos(k\theta_a(t)) = 0 \quad \text{for} \quad \theta_a(t) \neq 2\pi m
\]

where \( m \) is any integer.

This can be considered a condition of impulsiveness.

Equation (4) can be interpreted as a mapping of each possible pair of values \( (\theta_a(t), C) \) at a given time \( t \) to an output signal \( s(t) \). In other words,

\[
s(t) = f(\theta_a(t), C).
\]

The periodic impulse train, \( s(t) = C \delta(t - m f_o) \), where \( m \) is any integer, can be written as a Fourier series:

\[
s(t) = C \sum_{k=1}^{\infty} \cos(k2\pi f_o t)
\]

where \( f_o \) is the fundamental frequency.

This specific function captures the relationship between any \( s(t) \), \( \theta_a(t) \), and \( C \) because \( 2\pi f_o t \) covers all possible values of phase. We see that \( s(t) \) is only nonzero when the phase is equal to a multiple of \( 2\pi \). Therefore, the function described by (6) may be rewritten

\[
s(t) = f(\theta_a(t), C) = \left\{ \begin{array}{ll}
0 & \text{for } \theta_a(t) \neq 2\pi m \text{ where } m \in \mathbb{Z} \\
C \delta(t) & \text{for } \theta_a(t) = 2\pi m \text{ where } m \in \mathbb{Z}
\end{array} \right.. \tag{8}
\]

Another way to think about this function is that for \( \theta_a(t) \) not equal to a multiple of \( 2\pi \), one can always find a corresponding phase in (7) that gives a zero sum.

### 3. Uniqueness

#### 3.1. Finding the Instantaneous Frequency from an Observed Impulse Train

Given the above result, we now strive to find a set of functions \( \dot{\theta}_a(t) \) that yield the phase \( \theta_a(t) = 2\pi m \) where \( m \in \mathbb{Z} \) at the times of an observed set of impulses (and only at these points). Let us write the function for \( s(t) \) in terms of the instantaneous frequency by combining (2) and (4):

\[
s(t) = C \sum_{t=1}^{\infty} \cos(k \int_{\tau}^{t} \dot{\theta}_a(t) \, dt).
\]

Find that for \( t > t_o \) we have reached

\[
s(t) = C \sum_{t=1}^{\infty} \cos(k \int_{\tau}^{t} \dot{\theta}_a(t) \, dt) \quad \text{for } t > t_o.
\]

In order to simplify our expressions and avoid the situation of having negative frequencies, we constrain \( \dot{\theta}_a(t) \) to be positive for all \( t \). When \( s(t) \) becomes nonzero again, we have reached the next impulse, denoted \( b \), and the following equation holds:

\[
\int_{t_o}^{t} \dot{\theta}_a(t) \, dt = \frac{2\pi}{f_o} \quad \text{for } t > t_o.
\]

If the instantaneous frequency is set to be the constant \( 2\pi f_o \), then solving for the time of impulse \( b \), we obtain

\[
2\pi f_o (t_b - t_o) = 2\pi.
\]

\[
t_b - t_o = \frac{1}{f_o}.
\]

This is how we conventionally define the period, \( T_o \).

We may also allow \( \dot{\theta}_a(t) \) to vary with time, exhibiting frequency modulation. If we choose \( t_o \) apriori and solve for all the possible \( \dot{\theta}_a(t) \) functions that yield that time for impulse \( b \), we see that there are an infinite number and that they are constrained only by:

\[
\int_{t_o}^{t_b} \dot{\theta}_a(t) \, dt = 2\pi \quad \text{where } \dot{\theta}_a(t) > 0 \quad \text{for } t_o < t < t_b.
\]

One can model the time-varying \( \dot{\theta}_a(t) \) between impulses \( a \) and \( b \) in a number of ways. One typical way might be to assume that it is the sum of a constant component with period \( (t_b - t_o) \) and a time-varying component:

\[
\dot{\theta}_a(t) = 2\pi/(t_b - t_o) + \dot{\theta}_m(t) \quad \text{for } t_o < t < t_b.
\]

If this is the case, then we can solve for \( \dot{\theta}_m(t) \) resulting in:

\[
\int_{t_o}^{t} \dot{\theta}_m(t) \, dt = 0
\]

where \( \dot{\theta}_m(t) > -2\pi/(t_b - t_o) \) for \( t_o < t < t_b \).
The lower limit put on the value of $\bar{\theta}_{tm}(t)$ ensures that $\bar{\theta}_i(t)$ remains positive for all $t$ when the frequency-modulated part is summed with the constant-frequency component.

3.2. Finding the Impulse Train from Samples of the Underlying Instantaneous Frequency

There is an alternative way to pose the issue of multiple valid fundamental-phase tracks for a given set of impulses. In this section, we will show that given a set of phase measurements, there exist multiple valid sets of impulses that fit those measurements. One implication of this observation is that the times of impulses may be shifted (or added or deleted) depending on the type of phase interpolation used in synthesis.

As we have shown, if the phases and frequencies are harmonically related, we get an impulse each time the phase of the fundamental crosses $2\pi$. We have shown an integral constraint for the instantaneous fundamental frequency between two impulses. We can equivalently examine the converse case where we only know the instantaneous phases at several points and wish to find the time instants of the impulses. This type of phase measurement is made in standard sinewave analysis/synthesis or modification (e.g., time-scale or pitch modification) [3] as well as more recent renditions of sinewave-based processing [4].

Depending on how we interpolate the phases between a set of known phases, we will get impulses at varying points. Fig 2 illustrates this situation, with the times of impulses using a linear phase interpolation denoted with unfilled dots and the times of impulses using a nonlinear phase interpolation denoted with filled dots. Let us initially assume that we have a set of correctly-unwrapped and monotonically-increasing impulses. This type of phase measurement is made in standard sinewave analysis/synthesis or modification (e.g., time-scale or pitch modification) [3] as well as more recent renditions of sinewave-based processing [4].

**Figure 2:** Comparison of impulse locations determined by nonlinear-phase interpolation or linear-phase interpolation.

Depending on how we interpolate the phases between a set of known phases, we will get impulses at varying points. Fig 2 illustrates this situation, with the times of impulses using a linear phase interpolation denoted with unfilled dots and the times of impulses using a nonlinear phase interpolation denoted with filled dots. Let us initially assume that we have a set of correctly-unwrapped and monotonically-increasing phase measurements and that each of the sinusoids is constrained to be harmonically related to the fundamental at each time. If the difference in $\bar{\theta}_i(t)$ exceeds $2\pi$ between two fundamental phase measurements, at least one impulse will occur between the measurements, and its location can change drastically depending on the interpolation technique used. We know from the condition derived in Section 2, that the output will be zero when $\bar{\theta}_i(t) = 2\pi m$ and nonzero only when $\bar{\theta}_i(t) = 2\pi n$.

With real signals, we cannot unwrap the phase, and the measurement at each point is only known modulo $2\pi$. This means that both the number and locations of the impulses between the phase measurements are ambiguous.

Nonmodal speech may require rapid movements in the sinusoidal representation’s instantaneous frequency. We can interpret Figure 2 as illustrating the implications of an incorrect linear interpolation. Let us assume that the nonlinear solid line indicates the true fundamental-phase trajectory used to generate the impulse train, and assume that the measured phases will be the same for both tracks. The dashed line indicates the estimated phase trajectory using a linear interpolation between the measurement points. Using these tracks, the times of the actual impulses are indicated with filled dots and the estimated impulse times using the linear phase interpolation are denoted with unfilled dots. We can see that the interpolation method could lead to significant errors in the estimated times of the impulses.

**Figure 3:** Multiple harmonically-related instantaneous frequencies (bottom panels) can yield the same modal impulse train (top panels). A constant instantaneous fundamental frequency (left) and rapidly-varying track (right) are compared.

**Figure 4:** Multiple harmonically-related instantaneous frequencies (bottom panels) can yield the same nonmodal impulse train (top panels). A piecewise-constant instantaneous fundamental frequency (left) and piecewise-linear track (right) are compared.

**Figure 5:** Different interpolations of samples of an underlying phase function yield different impulse trains. The phase function on the left is the “true” phase function. The function on the right is that created by linearly interpolating phase samples taken every 10 ms.

3.3. Examples

An example of the ambiguity of sinusoidal tracks corresponding to a modal impulse train is shown in Figure 3. Here, a constant instantaneous frequency (left) and a rapidly-varying track (right) produce the same sequence of impulses. Figure 4 shows this ambiguity for nonmodal impulses. Identical impulse trains are generated by both a piecewise constant track (left) and a piecewise linear track (right).

Figure 5 shows a discrete-time simulation of the converse property, whereby different interpolations of known phase values yield different impulse trains. In the left panel, we see the impulse train and the instantaneous phase and frequency used to generate it. In the right panel, we see the result of linearly interpolating samples of the known phase taken every 10 ms. We can see that the impulses have moved relative to the original impulses (dashed line).
4. EXTENSIONS TO REAL SPEECH

One of our purposes is to provide insight into why sinusoidal analysis/synthesis can add, delete, and move pulses of nonmodal speech, as we saw illustrated in Figure 1. We have in this paper investigated the simplifying case of idealized impulses and have argued in Section 3.2, for harmonically-related instantaneous frequencies, how their locations can move in synthesis. With similar arguments, we can show that addition and deletion of impulses are also possible with synthesis from a sampled phase function.

4.1. Considerations

Developing a formal argument for the modification of underlying pulses in actual speech will require a generalization to the relations presented in this paper and specifically will require addressing two properties of real speech: First, standard sinusoidal analysis does not always result in frequency harmonicity or continuity such as we have described in our model. In nonmodal speech regions, sinusoidal analysis via peak-picking can yield what appear to be frequencies that are erratic spectrally and temporally [5]. An important consequence is that, although impulses are generated approximately in synthesis, the resulting frequency tracks are broken and sporadic and appear to have little meaning physiologically. The second property to address involves the condition that we have shown in Section 2 and its implications that relate to the case of continuous-time impulse trains. We do not yet fully understand how associated sinusoidal tracks change with the introduction of vocal-tract linear filtering and bandlimiting. Bovik [6], however, discusses the process of approximating the effect of a linear filter on a single sinusoidal track which may be extended to filtering effects for our harmonically-related sinusoidal model.

Despite these unknown elements, it is clear that the underlying impulse train for a speech waveform can have many sinusoidal representations. When an impulse train is filtered, the ambiguities remain. Hence, sinusoidal analysis of real speech will suffer from the same sinusoidal-model uniqueness problems of impulse trains. In future work, we must explore how these ambiguities manifest themselves.

4.2. Example Applications

An application where the effects of pulse alternation can be seen due to interpolation is sinusoidal time-scale expansion. Figure 6 shows one such example, where a diplophonic region of nonmodality, characterized by every other pulse shifted and reduced in energy, is expanded. A section of diplophonia was introduced in Figure 2. We do not yet fully understand how associated sinusoidal tracks change with the introduction of vocal-tract linear filtering and bandlimiting. Bovik [6], however, discusses the process of approximating the effect of a linear filter on a single sinusoidal track which may be extended to filtering effects for our harmonically-related sinusoidal model.

Despite these unknown elements, it is clear that the underlying impulse train for a speech waveform can have many sinusoidal representations. When an impulse train is filtered, the ambiguities remain. Hence, sinusoidal analysis of real speech will suffer from the same sinusoidal-model uniqueness problems of impulse trains. In future work, we must explore how these ambiguities manifest themselves.

5. CONCLUSIONS

In this paper, we have shown two important properties of the speech source:

(1) A nonmodal or modal impulse train does not have a unique sinusoidal representation. There are many possible sinusoidal representations of the impulsive source that yield the same output. We have shown this property using one possible sinusoidal representation which is a series of harmonically-related sinusoids.

(2) If we estimate the harmonically-related phases of an underlying impulsive source based on a discrete set of known fundamental phases, then the impulse train that results is not unique. Additionally, phase-unwrapping issues in this estimate can yield additions, deletions, and movements of impulses.

Finally, we have illustrated how our theory of impulse ambiguity is consistent with observations in a real speech application, and we have discussed requirements for extending the concepts presented in this paper to actual speech. This extension will require addressing the nonharmonicity of measured instantaneous frequency and the effect of vocal tract filtering.

6. REFERENCES