GROUP DELAY MANIPULATION FOR CREATIVE SOUND TRANSFORMATION WITH THE GIANT FFT

Ted Apel

Department of Games, Interactive Media, and Mobile Technology
Boise State University
Boise, ID, USA
tedapel@boisestate.edu

ABSTRACT

The Giant FFT is a single DFT spanning an entire audio file that produces a spectrum encoding the complete temporal evolution of a sound. Creative manipulations in this domain have produced compelling results, but typically smear discrete events into sustained textures by disrupting the temporal relationships between frequency bins. This paper introduces a framework for coherent spectral manipulation in the "group delay domain", where the derivative of the phase spectrum with respect to frequency makes the temporal center of gravity of spectral energy explicit at every frequency bin. By identifying spectral regions around amplitude peaks and grouping them by group delay similarity, spectral features can be displaced in time through uniform modification of their group delay. We demonstrate reordering of melodic events in synthetic and acoustic melodies, amplitude-proportional spectral displacement for creative sound design, and modulation of group delay for creating temporal copies of spectral features. These results show that methods of coherent temporal manipulation are possible in the Giant FFT domain without windowed analysis.

1. INTRODUCTION

The Giant FFT method of sound manipulation is based on performing a single FFT over the entire duration of an audio recording. The resultant complex spectrum encodes the complete temporal evolution of the sound, but in a domain that is difficult to intuitively manipulate. Prior work in this domain has produced compelling transformations of musical sound, even though manipulations typically disrupt the temporal relationships between spectral components. In particular, manipulations of the phase spectrum tend to destroy the temporal structure of the original sound, producing incoherent drone-like textures. A coherent manipulation would preserve the temporal relationships within a spectral feature while displacing it in time. We propose working in the group delay domain, where the temporal center of gravity of spectral energy is made explicit, enabling spectral features to be identified and displaced in time as uniform groups.

This paper introduces the group delay domain for creative manipulation of the Giant FFT, combining group delay based temporal identification of spectral features with coherent displacement of grouped frequency bins. We present background on Giant FFT sound transformation, describe a framework for coherent spectral manipulation in the group delay domain, and demonstrate its capabilities through three transformations: reordering of melodic events, amplitude-proportional time displacement of spectral features, and sinusoidal group delay modulation for creating temporal copies. Sound examples accompanying the transformations described in this paper are available online.[1]

2. BACKGROUND

The analysis of a very long data sequence with a single FFT has a long history in astrophysics and geophysics, where the intention is to find frequency components in a noisy signal [1]. In 1999 Hammer and Sundt applied this idea of computing a single FFT to musical audio [2]. Their "Giant FFT" technique calculates a single (amplitude and phase) spectrum that can be manipulated before reconstruction with the inverse FFT. Their Mammut program presented several sound transformations that operate on the amplitude or phase spectrum and produce novel sonic results. In general, these transformations can be characterized by a reduction in coherence between spectral components, leading to discrete sonic events being smeared into droning textures. In addition, some transformations impose nonlinear spectral sweeps on the sound. The Giant FFT technique has remained an alternative creative audio processing method [3, 4, 5], but its tendency to smooth auditory events into spectral drones has limited its broader application.

Recent work has begun to quantify how manipulations in the Giant FFT domain could produce more coherent sonic outcomes. Välimäki et al. [6] set the phase spectrum of the Giant FFT to zero and studied the magnitude spectrum's contribution to the resynthesized sound. They found that this manipulation produced a sound that is symmetric in time and maintains some rhythmic qualities of the original. Their introduction of zero-padding to avoid temporal aliasing removes an entire category of artifacts present in earlier Giant FFT systems, opening the domain to more principled investigation of spectral transformations.

Välimäki et al. demonstrate that the amplitude spectrum alone can retain some of the rhythmic and temporal structure of the original sound. Their work isolates the contribution of the magnitude spectrum by setting phase to zero. Here we take a complementary approach, transforming the phase spectrum into its derivative, the group delay spectrum, where temporal structure becomes more explicit and manipulable.

[1] https://vud.org/giant-fft-group-delay/

3. METHOD

The phase spectrum of the Giant FFT encodes much of the time evolution of the analyzed sound. Principled manipulation of this spectrum could produce sounds whose spectral components maintain their temporal structure while being displaced in time. The fundamental difficulty with manipulating the phase spectrum is that individual phase values do not correspond to meaningful perceptual quantities, especially for large FFT lengths. Changing a phase value does not produce a meaningful manipulation of a sonic event when the sound is synthesized back in the time domain. Phase values, even when unwrapped, do not directly reveal the temporal nature of spectral energy. We seek a form of the phase spectrum whose values can be read directly as time. A sound event at time t contributes a phase of $-2\pi ft$ at each frequency f, so the rate of phase change across frequency is proportional to the event's position in time. The group delay, the negative derivative of the phase spectrum with respect to frequency, reads this slope at every frequency bin, producing values in units of time. Components arising from the same event share one slope, and therefore share a group delay value. The group delay allows us to interpret its values as the temporal center of gravity of spectral energy at each frequency bin [7], that is, the amplitude-weighted mean temporal position of the energy at that frequency. The Giant FFT provides very high frequency resolution across the entire spectrum while the group delay provides the temporal center of gravity of energy at each frequency bin. We refer to this combined representation of magnitude and group delay as the group delay domain. In this domain, spectral features can be identified by their relative temporal center of gravity and manipulated according to when they occur.

Although manipulating values of the group delay spectrum is more intuitive than manipulating the phase spectrum directly, we still cannot change individual values without producing incoherent results. A sonic event affects many frequency bins simultaneously. If these bins are shifted by different amounts, the phase relationships between them are disrupted, smearing the event in time rather than displacing it. For the event to move cleanly, all bins belonging to it must be shifted by a uniform amount. We address this by identifying regions in the frequency spectrum where coherent displacement matters most. In pitched sounds, these are the regions around peaks in the amplitude spectrum. Each identified region is treated as indivisible in group delay shape. With this, we will see that the region can be displaced forward or backward in time by preserving the relative group delay values across its bins.

The importance of phase in spectral representations has long been recognized [8]. The group delay spectrum has been used as an analysis tool for understanding the temporal center of gravity of spectral components [9, 10]. Canfield-Dafilou and Abel [11] have used group delay trajectories for creative allpass filter design, but not in the Giant FFT context. Until now, group delay has not been explored as a basis for coherent spectral transformations in the Giant FFT domain. Figure 1 shows the processing pipeline of the entire analysis-synthesis system.

3.1. Analysis

Our method begins with a Giant FFT over the entire input sound, regardless of its duration. The method is most appropriate for sounds with pitched content, as it uses amplitude peaks in the spectrum as a grouping mechanism. The input sound is first padded with silence sized to the intended displacements, before the signal for backward shifts and after it for forward shifts. The padded signal is then zero-padded to the next power of two, preventing circular aliasing in the output. The FFT of the zero-padded input yields a magnitude spectrum and a phase spectrum. Together these constitute the Giant FFT representation.

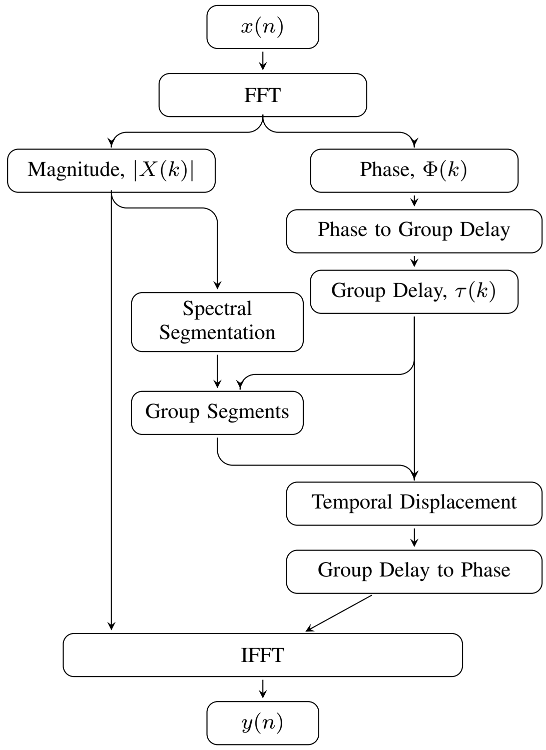

Figure 1: Processing pipeline. The input signal is transformed into the group delay domain, where spectral segments are identified and manipulated before resynthesis.

We transform this representation into the group delay domain by computing the group delay spectrum from the phase spectrum. The group delay can be computed directly from the complex spectrum without requiring phase unwrapping [12]. However, we compute it by unwrapping the phase and taking the difference between adjacent phase values, as this method is exactly invertible through cumulative summation, ensuring lossless analysis-synthesis:

$$\tau(k) = -\left(\Phi(k+1) - \Phi(k)\right) \qquad (1)$$

where $\tau(k)$ is the group delay in radians per bin, $\Phi(k)$ is the unwrapped phase spectrum, and k is the frequency bin index. The analysis uses a real-input FFT, with the negative frequencies implied by conjugate symmetry. This first-order finite difference differs from $-d\Phi/d\omega$ only by a constant factor, the bin spacing, and is used because it integrates back to phase exactly. Figure 2 shows the group delay spectrum of two synthetic tones separated by one second with added noise. Here $\tau(k)$ values are converted from radians per bin to seconds by dividing by the bin spacing in angular frequency, $2\pi f_s/N$ radians per second, where f_s is the sample rate and N is the zero-padded FFT length. As can be seen, group delay values cluster around the temporal positions of the two tones. This clustering is the phenomenon we will exploit when grouping spectral segments. Our method treats a cluster of similar group delay values as bins belonging to the same spectral event.

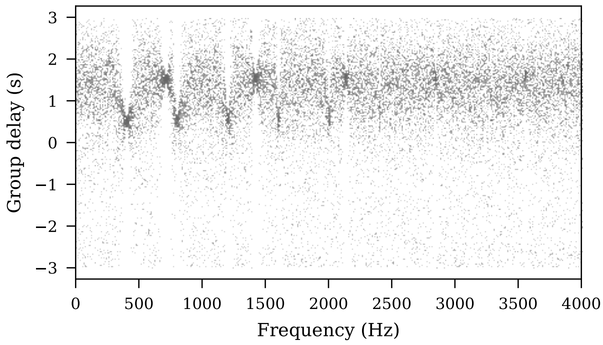

Figure 2: Group delay spectrum of two synthetic tones separated by one second. Values cluster around the temporal center of gravity of each tone.

3.2. Coherent Spectral Segmentation

The next step is to segment the group delay spectrum into regions of high spectral energy that share similar group delay values. The concept of spectral peaks having regions of influence was introduced in the context of phase vocoder phase locking [13]. We apply a similar concept in the Giant FFT group delay domain. The amplitude spectrum is smoothed with a Gaussian kernel of a specified width in bins. Contiguous regions where the smoothed amplitude exceeds a threshold relative to the global peak are identified as spectral peaks. Each peak is then extended to a specified segment width, limited by the midpoint between neighboring peaks, capturing the spectral energy surrounding the peak. Figure 3 shows the segmentation for the same synthetic signal. Shaded regions indicate identified segments around spectral peaks.

Next, the temporal position of each segment is computed as the magnitude-weighted mean of its group delay values in seconds. Segments whose mean group delay values fall within a specified tolerance in seconds are grouped together, forming a single group that spans multiple spectral peaks. This grouping mechanism identifies areas of the spectrum that are both part of an amplitude peak and share similar group delay values.

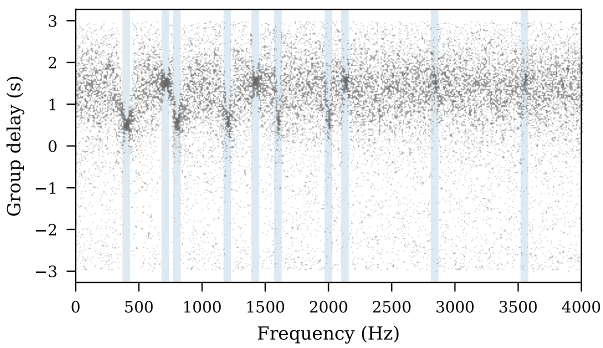

Figure 3: Group delay spectrum with identified segments (shaded) around amplitude peaks.

After analysis there are two user-adjustable parameters. One is the width of the segments around spectral peaks, adjustable from narrow to spanning the entire spectrum. The examples of Section 4 use segment widths of 8000 to 12000 bins. For sounds with sharp onsets, segments wider than the spectral peak are needed to capture the transient energy that extends across a broad spectral region around each peak. When a wide segment is displaced, this transient energy moves as a unit, preserving some of the attack character of the original sound. Two, the energy in bins not identified within any segment can optionally be zeroed to remove low-energy or noisy spectral components. With this analysis complete we can now transform the group delay spectrum to create novel audio effects. In each transformation, with the exception of the zeroing mentioned above, the amplitude spectrum is left completely unaltered.

3.3. Group Delay Manipulation

The particular shape of the group delay values in a band encodes the temporal structure of that band. To preserve this temporal structure (note onsets and decays, for example) the shape must be maintained. By adding a constant to a segment's group delay values, the segment's energy is displaced in time:

$$\tau_{\text{new}}(k) = \tau(k) + C \tag{2}$$

for all bins k within the segment, where C is the displacement amount in radians per bin. Adding a constant to the group delay corresponds to a linear phase shift across frequency, which produces a time displacement in the resynthesized signal. This operation is the basis of all the effects presented here. Uniform displacement of group delay segments allows spectral peak energy to be displaced independently of other spectral energy.

3.4. Synthesis

Synthesis begins with a cumulative summation of the modified group delay spectrum to create a new phase spectrum:

$$\Phi_{\text{new}}(k) = \Phi(0) - \sum_{j=0}^{k-1} \tau_{\text{new}}(j) \tag{3}$$

where $\Phi(0)$ is the first value of the unwrapped phase spectrum, retained from the analysis. This summation is the exact inverse of the difference in Eq. (1). This phase spectrum is combined with the magnitude spectrum and the inverse FFT is performed to produce the output audio. With no transformations, the analysis-synthesis process, including the group delay conversion, recovers the original signal to floating-point precision.

4. TRANSFORMATIONS AND RESULTS

With the analysis system in place we can explore transformations of the segmented group delay spectrum. The following examples were computed in Python using NumPy, SciPy, and librosa. In this section we present example transformations that show the capabilities and limitations of this technique. All of the transformations are based on modifying the group delay of entire segments.

4.1. Temporal Reordering

In the first example, a synthesized arpeggio of 4 notes with short silences between notes is analyzed. Figure 4(a) shows the amplitude spectrum and (b) shows the corresponding group delay spectrum of this sound. Colored bars indicate identified segments, with segments of the same color belonging to the same group. Horizontal lines in (b) show the mean group delay of each segment. The algorithm has identified four groups corresponding to the four notes. By adding or subtracting a constant to each of these groups, the tonal components of each note can be moved in time freely. In this example we move these groups to play the arpeggio in reverse order. Unlike reversing the time domain signal, in which the temporal evolution of each note would also be reversed, here the original amplitude envelopes are retained while the individual notes play in reverse order. Notably, the temporal positions and frequency content of the musical notes were found entirely in the frequency domain using spectral peaks and group delay similarity, without any time-domain onset detection or temporal analysis. This demonstrates that the segmentation mostly identified the spectral components of each note across frequency, and that uniform displacement preserved the amplitude envelope of each note while changing only its temporal position. Figure 5 shows the spectrogram of the original sound and the spectrogram after the notes have been reordered. Some artifacts are audible in this example. They reflect a tradeoff in segment width. Narrow segments leave low-level energy behind at its original position, while overly wide segments displace energy belonging to other events.

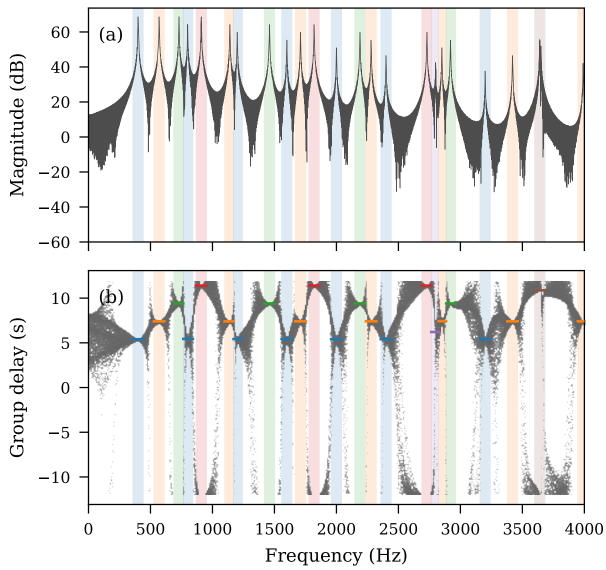

Figure 4: Synthetic 4-note arpeggio. (a) Amplitude spectrum with identified segments shown as colored bars. (b) Group delay spectrum with the same segments. Horizontal lines show the mean group delay of each segment. Segments of the same color belong to the same temporal group.

A second example of the same reversal transformation uses a recording of a glockenspiel playing a 4-note arpeggio. Figure 6 shows the amplitude spectrum with identified groups shown as colored regions. The gray background shows the full spectrum while colored regions indicate segments that have been grouped by group delay similarity. The inharmonic partials of the glockenspiel have been correctly identified as belonging to different notes in time. This grouping succeeds because the notes rarely share frequency bins, a constraint discussed in Section 5.1. In this recording of an acoustic instrument, the percussive note onsets are largely preserved because the segments are wide enough to capture the transient spectral energy around each peak. The result is a glockenspiel arpeggio played in reverse temporal order, demonstrating that the technique works on real recorded sounds with inharmonic partials. This example can be heard on the accompanying website.

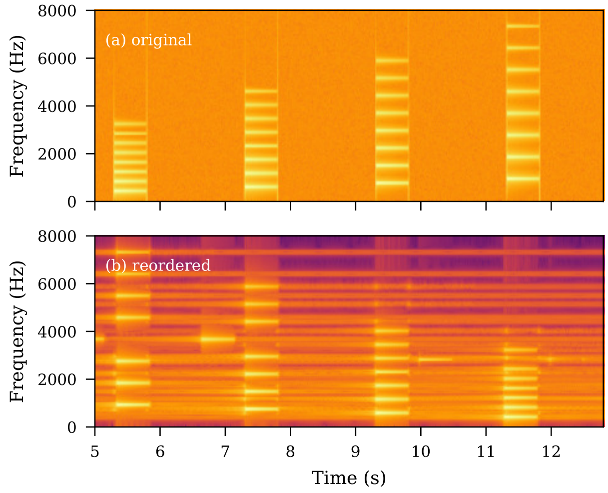

Figure 5: (a) Spectrogram of the original synthesized arpeggio. (b) Spectrogram after temporal reordering of the note groups.

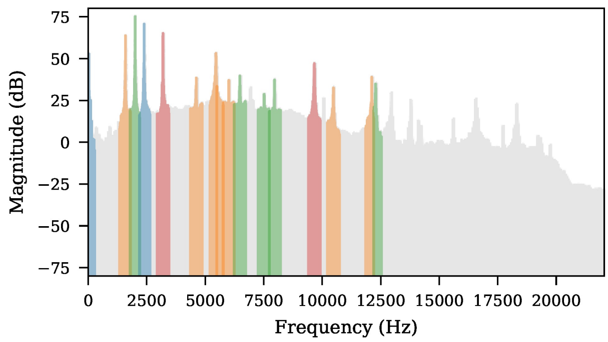

Figure 6: Amplitude spectrum of a 4-note glockenspiel arpeggio. Colored regions show segments grouped by group delay similarity. Each color represents a different note. Gray background shows the full ungrouped spectrum.

4.2. Amplitude-Proportional Displacement

In the previous examples, groups of segments were displaced by prescribed amounts. In this transformation, these displacements are set automatically by using the amplitude of each spectral peak. For example, a segment centered around a high amplitude peak is displaced more than a segment with a low amplitude peak. The result is that the spectral content of a sound unfolds over a user-specified duration, with each partial arriving at a different time proportional to its amplitude. A single note becomes a sequence of its partials spread across a longer time span, while preserving the internal temporal structure of each spectral band. Figure 7 illustrates this transformation and the importance of segmentation. Panel (a) shows the spectrogram of a single harpsichord note. Panel (b) shows the result of applying amplitude-proportional displacement to every bin independently, without segmentation and grouping. The spectral energy is smeared across time, producing the drone-like texture characteristic of incoherent Giant FFT manipulations. Panel (c) shows the same displacement applied uniformly within identified segments. Individual partials remain distinct and are cleanly displaced in time, with higher amplitude partials displaced further.

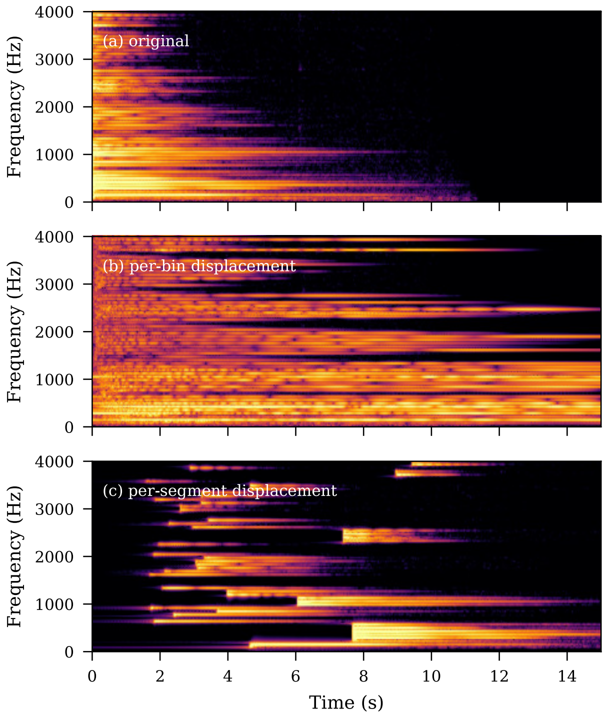

Figure 7: (a) Spectrogram of a single harpsichord note. (b) Amplitude-proportional displacement without segmentation. (c) Per-segment amplitude-proportional displacement.

4.3. Periodic Group Delay Modulation

The final transformation not only displaces spectral segments in time but also duplicates them at varying amplitudes. The combined effect can be thought of as composing with the spectral content of a sound fragment entirely in the spectral domain. By displacing and duplicating spectral content, musical textures that unfold in time can be constructed entirely from qualities of the sound itself.

To create copies of spectral segments, we modulate the group delay values of a segment across frequency to produce discrete temporal copies of that segment's energy. The number and strength of these copies are controlled by the modulation depth and period:

$$\tau_{\text{new}}(k) = \tau(k) + (M)(T)\sin\left(\frac{2\pi kTf_s}{N}\right) \tag{4}$$

where M is the scalar modulation depth, T is the repetition period in seconds, f_s is the sample rate, N is the zero-padded FFT length of Section 3.1, and τ is expressed in seconds. A constant group delay offset produces a single time shift. A sinusoidal variation across frequency produces periodically spaced copies in time. The modulation period T controls the spacing between copies, and the depth M controls how many copies appear. The factor of T preceding the sine scales the modulation so that the number of copies is controlled by M alone, independent of the period. In this method we modulate each segment with a different frequency sinusoid. Different segments can then repeat at different rates, creating polyrhythmic textures in time from a single input sound. In this example a single harpsichord note is displaced and duplicated, with lower frequency segments repeating slower and higher frequency segments repeating faster. M values between 0.4 and 2 and T values between 0.01 and 4 seconds were used. The length of the output sound is a user-defined value. Figure 8 shows the spectrogram of the output sound, extended to 20 times the original duration. Each repetition retains the envelope character of the original harpsichord note, demonstrating that the modulation creates discrete copies rather than continuous smearing.

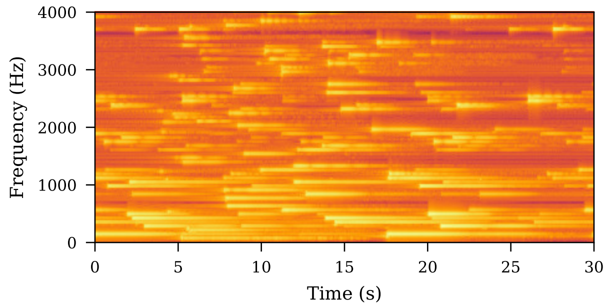

Figure 8: Spectrogram of a harpsichord note after displacement and sinusoidal group delay modulation, extended to 20 times the original duration. Different spectral segments repeat at different rates.

4.4. Comparison with Spectral Delays

All three transformations displace spectral segments in time, and in this respect they resemble spectral delays [14, 15], in which frequency bands are delayed by prescribed amounts. Indeed, the sound transformations made available by the group delay domain could be approximated by manipulations in a time-frequency representation such as the STFT. A windowed analysis, however, must commit to a window length, trading frequency resolution against temporal extent. No window can both resolve closely spaced partials and contain a partial's complete evolution in time. The Giant FFT has no such tradeoff. The temporal envelope of each partial, no matter its duration, is concentrated in the bins surrounding its spectral peak, and uniform displacement moves that envelope largely intact.

More fundamentally, the spectral delay has no temporal information about the signal itself. In the group delay domain, by contrast, the temporal center of gravity of every spectral component is an explicit quantity. A segment's center of gravity can therefore be moved to a chosen position in time. And because the positions of all segments are known, they can be redistributed or reordered in relation to one another. Transformations are specified in terms of spectral features and their positions in time, rather than frequency bands and delay amounts.

5. CONCLUSIONS

In this paper we introduced the group delay domain as a representation for creative sound manipulation with the Giant FFT. In this domain, the temporal center of gravity of spectral energy is explicit at every frequency bin, enabling spectral features to be identified and displaced in time. We devised a partitioning algorithm to identify temporally coherent groups within this spectrum. We demonstrated reordering of melodic events and creative transformations based on displacing and duplicating harmonic and inharmonic partials in time.

5.1. Limitations

The method works best when events do not share frequency bins. It is most effective on individual sound events or sequences of temporally non-overlapping events. When multiple events overlap in frequency, such as in sustained polyphonic material, the group delay of shared bins reflects the amplitude-weighted average of all contributing events rather than the position of any single event. Events that share frequency bins cannot be displaced independently. Shifting those bins moves all contributing events together, preserving their temporal relationships but preventing selective reordering of individual events. Nevertheless, the displacement and duplication transformations can still produce interesting results on sounds with dense overlapping onsets.

5.2. Future Work

The modulation of group delay segments to create temporally displaced copies has produced musically compelling sonic transformations. Future work includes exploring other periodic group delay modulation patterns to produce more complex rhythmic repetitions. The group delay domain also suggests the possibility of "temporal filtering", where bins are retained or zeroed according to their group delay values, selecting spectral energy by when it occurs rather than by its frequency. This might allow the extraction of individual events from a mixture.

The Giant FFT domain has long been intriguing but resistant to perceptually meaningful manipulation. Projecting a sound into a single complex spectrum produces a representation that is non-intuitive to manipulate creatively. However, entirely incoherent manipulations such as those in Mammut and zero-phase sound have shown that the sonic results are compelling. This work has shown that the group delay domain is a fruitful representation of the Giant FFT phase spectrum, in which temporal structure becomes visible and manipulable, leading to interesting and comprehensible transformations of sound. The group delay domain is not offered as a replacement for windowed methods, but as a different way of working with sound, one where time is not a location in the signal but a property of each spectral component.

6. ACKNOWLEDGMENTS

The author thanks the anonymous reviewers for their careful readings and very constructive suggestions.

7. REFERENCES

[1] R. R. Hocking, "Performing fourier transforms on extremely long data streams," *Computers in Physics*, vol. 3, no. 1, pp. 59–65, 1989.

[2] Ø. Hammer and T. Sundt, "Musical applications of decomposition with global support," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, Trondheim, Norway, 1999.

[3] T. Burt, "Composition folio," Ph.D. dissertation, University of York, 2012.

[4] V. Välimäki, J. Rämö, and F. Esqueda, "Creating endless sounds," in *Proceedings of the International Conference on Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, 2018, pp. 32–39.

[5] T. Apel, "Negated sound," in *Proceedings of the Speculative Sound Synthesis Symposium*, 2019, online: https://speculativesoundsynthesis.iem.sh/symposium/docs/proceedings/apel/.

[6] V. Välimäki, R. Salmi, S. Bilbao, S. J. Schlecht, and D. Zicarelli, "Zero-phase sound via giant fft," in *Proceedings of the 28th International Conference on Digital Audio Effects (DAFx25)*, Ancona, Italy, 2025.

[7] L. Cohen, *Time-Frequency Analysis*. Prentice Hall, 1995.

[8] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[9] A. Röbel, "Transient detection and preservation in the phase vocoder," in *Proceedings of the International Computer Music Conf. (ICMC)*, 2003.

[10] Z. Průša and N. Holighaus, "Phase vocoder done right," in *Proceedings European Signal Processing Conference (EUSIPCO)*, 2017.

[11] E. K. Canfield-Dafilou and J. S. Abel, "Group delay-based allpass filters for abstract sound synthesis and audio effects processing," in *Proceedings of the International Conference on Digital Audio Effects (DAFx18)*, 2018.

[12] J. O. Smith III, "Numerical computation of group delay," Online: https://ccrma.stanford.edu/~jos/fp/Numerical_Computation_Group_Delay.html, accessed 2026.

[13] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323–332, 1999.

[14] D. Kim-Boyle, "Spectral delays with frequency domain processing," in *Proceedings of the International Conference on Digital Audio Effects (DAFx04)*, Naples, Italy, 2004.

[15] A. Cont, C. Laurenzi, and M. Stroppa, "Chromax, the other side of the spectral delay between signal processing and composition," in *Proceedings of the International Conference on Digital Audio Effects (DAFx13)*, Maynooth, Ireland, 2013.