Swarm Modulation: An algorithm for real-time spectral transformation

Ryan Janzen and Steve Mann
Department of Electrical and Computer Engineering
University of Toronto

***Abstract*—A novel class of modulation is introduced, for frequency transformation and spectral synthesis in real-time. Swarm modulation has potential applications to enhance human hearing with extended frequency ranges, in medical diagnostics for electrocardiogram (ECG), electroencephalogram (EEG) and other medical signals, for RADAR analysis, for user-interface sonification, and for sound synthesis or non-synthetic sound transformation. Swarm modulation is a new way to transform signals, and is demonstrated for transforming subsonic and ultrasonic sound into the audible range of human hearing.**

Swarm modulation is based on the principle of phase-incoherent frequency-roaming oscillation. Features in the frequency-time plane are reconstructed via a time-varying process, controllable with instantaneous zero-latency reaction time to new information. Swarm modulation allows prioritization of salient output spectral features for efficient processing, and overcomes cyclic beating patterns when Fourier and wavelet-based methods are applied in a stationary manner.

Swarm modulation can flexibly re-map sound when a user expressively touches physical matter creating vibration. By detecting subsonic, sonic and ultrasonic vibrations, we can add to materials a rich acoustic user-feedback that can be adjusted to sound like a bell, xylophone, dull piece of wood, or a variety of other objects, in real-time. By dynamically controlling the output sound spectrum depending on the input spectrum, simultaneously with a continuous and low-latency temporal response, the system imitates the physicality of touching a real object.

Applied in control panels and expressive control surfaces, swarm modulation can create realistic sonic feedback, for human head-up operation of controls in critical applications.

I. Introduction

Swarm modulation will be mathematically defined in Section IV. First, we explain why there is a fundamental gap in real-time acoustic spectral transformations.

A. Why are Frequency Transformations desired?

The human senses operate over various spectral bands: Human hearing is typically sensitive from $\sim$20Hz to 20kHz, while vision can sense light from approximately 390 to 750 nm in wavelength [9]. Transforming our abilities, computerized eyeglasses have been built to shift infrared light to the red spectrum, all visible light to green, and ultraviolet light to blue, to gain real-time vision beyond what is possible for a human alone to see [10]. Figure 2 illustrates.

In this work we attempt to do the same for hearing. This poses a greater challenge in real-time—the durations of wave cycles are much closer to the minimum time increments we can perceive, as compared to visible light, and thus there are complexities beyond simply "copying and mapping" sensor readings as with from pixels from video cameras.

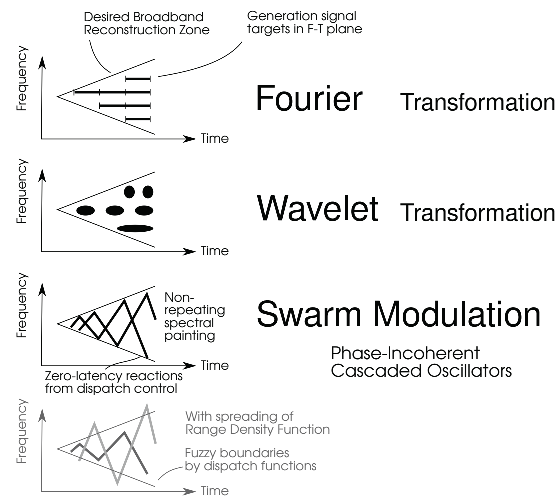

Fig. 1: Comparison of methods for tiling the frequency-time plane, with causal real-time reconstruction of a time-varying requested spectrum (shaped as '$<$' in this example). While Prolate Spheroidal Wave Functions [1–5] attempt to approach the limitation in precision of Heisenberg-Gabor uncertainty [6–8], Swarm Modulation overcomes coherent beating while enabling real-time flexible spectral mapping with zero-control-latency, for highly-realistic sound mapping.

The problem is especially acute when attempting to "hear" user-interfaces, where we wish to create tactile input devices that are more expressive than a simple on/off button or switch, and which produce a rich acoustic user-feedback response, beyond a simple "beep" sound whenever a button is pushed. Previous work has been able to *sonify* [11, 12] user-interfaces based on triggering sound samples [13–15] — which destroys most of the input information. Sensing the original acoustic sound allows the device to conform more closely to physical reality [10, 16–19]. The result allows a user to touch highly sensitive surfaces in order to expressively control a computer system, beyond binary on/off keys or switches [15, 19–23].

However, previous frequency transformations that are based on naturally-occurring acoustic sound from human touch are afflicted with beating [21, 22], spectral dead zones in the input, and other artifacts which create a low-quality sonic response to touch.

We therefore wish to create a new, flexible algorithm for frequency transformation. A key need exists in natural user-interfaces [10], to transform subsonic and ultrasonic vibrations into expressive sound in the audible range. We desire flexibility in mapping spectral ranges, and desire an audio rendering that is responsive to physical acoustics (as argued by Kapralos [16] and Parker & Heerema [17]).

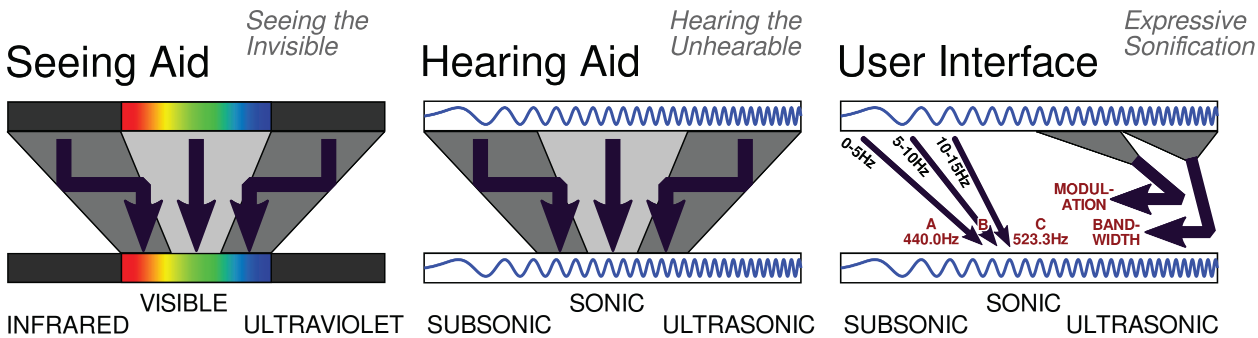

Fig. 2: Proposed uses of frequency shifting for sound. Previous researchers have shifted light wavelengths to build seeing-aids. We propose to tackle the more difficult challenge of shifting sound frequency ranges, to enable rapid and accurate user-interface feedback, and to hear beyond the range of human hearing, while fundamentally allowing a low-latency real-time response.

The result, for example, could be used to create custom DIY-style musical instruments, by simply setting up one acoustic pickup and one frequency shifter for each note of a scale. In a simple case, an array of identical blocks of wood could be individually affixed with geophones, so they can each affect a different output tone when struck [18].

We will identify fundamental limitations in existing methods for expressive interface sonification, and present a new algorithm that overcomes their flaws.

Swarm modulation can be used to create more expressive natural user-interfaces [10] and physiphones [18], where users directly touch physical matter (solids, liquids, and gases). Subsonic, sonic, and ultrasonic content from human touch could then be sensed directly by geophones, hydrophones or microphones, to capture a significant amount of information from the original interaction.

By making it now possible to immediately hear the response of that physical matter, with reduced beating artifacts, low latency, with a response much more expressive than binary keys, and with rich spectral information previously unavailable to the human ears, we can permit a tighter user-feedback loop, thus increasing control accuracy, and deepening the sense of engagement between a user and a device he/she is touching.

B. Previous transformation methods for sonification

Previous user-interfaces have used primitive methods to transform frequencies, leading to poorly-responsive sound. A basic method to add sound to a user-interface has been to trigger an audio sample (*e.g.* using MIDI and synthesis[1] whenever user interaction is detected [13, 14, 27]. Unfortunately, this triggering method is non-expressive: It eliminates much of the rich information from continuous-time, continuous-amplitude user-interfaces, or from sound recordings.

Preset-waveform heterodyning (*e.g.* gain-control) is also limited in expressive sound quality, despite having a continuous response. Sensor signals are multiplied by a waveform of a chosen frequency, to "hear" vibrations [20, 27, 28]. This is used with *duringtouch* and *FLUIDI* [27], an extension of MIDI, to listen to subsonic vibrations using commercial sound synthesizers. Despite a continuously-variable sound response to the acoustic input medium, only slowly-varying DC control inputs can be detected [21].

Another simple previous method has been to use bandpass filters [18, 21, 22, 29, 30]. Touching a set of tactile surfaces, for example, can excite a series of tones, each distinguishable and separate, using a bank of bandpass filters (a "filterbank"). This method, though, is prone to unpleasant beating sounds and is non-functional if the input lacks the desired frequency.

Finally, the input signal can be converted to the frequency domain or wavelet domain, and remapped before converting back to the time domain [31–33]. However, this T/IT method is subject to repeating patterns in reconstruction (including beating), and it cannot instantly respond to sudden events in the input. For example, a wavelet is rendered in full even if a last-minute change is demanded according to the input. Fourier transforms have limited precision in the time domain, conflicting with spectral precision [6]. The response is "chopped up" in frequency or time.

Limitations of the Preset-Waveform-Heterodyning method	**H,GC**
• The input's spectrum is merely transposed to different locations, without the option of frequency-dependent mapping, stretching, or contraction. For example, the range from 50-100kHz cannot be mapped/compressed to 100-200Hz for hearing. • The input's amplitude can only linearly control the output amplitude, and not other parameters. • In the common special case, the **gain-control method** of controlling the gain of a preset waveform, only low-frequency control signals can be captured from a sensor, (*e.g.* a simple sensor input from 0 Hz DC up to about 10 Hz) to avoid extended convolution of the preset waveform. Pertinent higher-frequency information must be dropped. See Fig. 4.	
Limitations of the Bandpass Filtering method	**BPF**
• Only works if the input happens to contain the desired output frequency. For user-interfaces the response to touch is unreliable. • Delay: slow response time of the filter; • Prone to erratic output envelope, even when the input envelope is roughly constant, due to phase mismatch and constructive/destructive interference (*i.e.* "sounds bad")	
Limitations of the FT/IFT, WT/IWT methods	**T/IT**
• Latency of input events, due to discrete windows or wavelets • Processing power required to transform full 2D F-T/W-T plane • Poor precision (time/frequency-domain tradeoff)	

[1]There is no need here to compare synthesis methods such as ring modulation or FM synthesis, which merely generate artificial sounds, but do not themselves suggest how to decompose and remap arbitrary spectral regions of an input signal, to different arbitrary spectral regions.

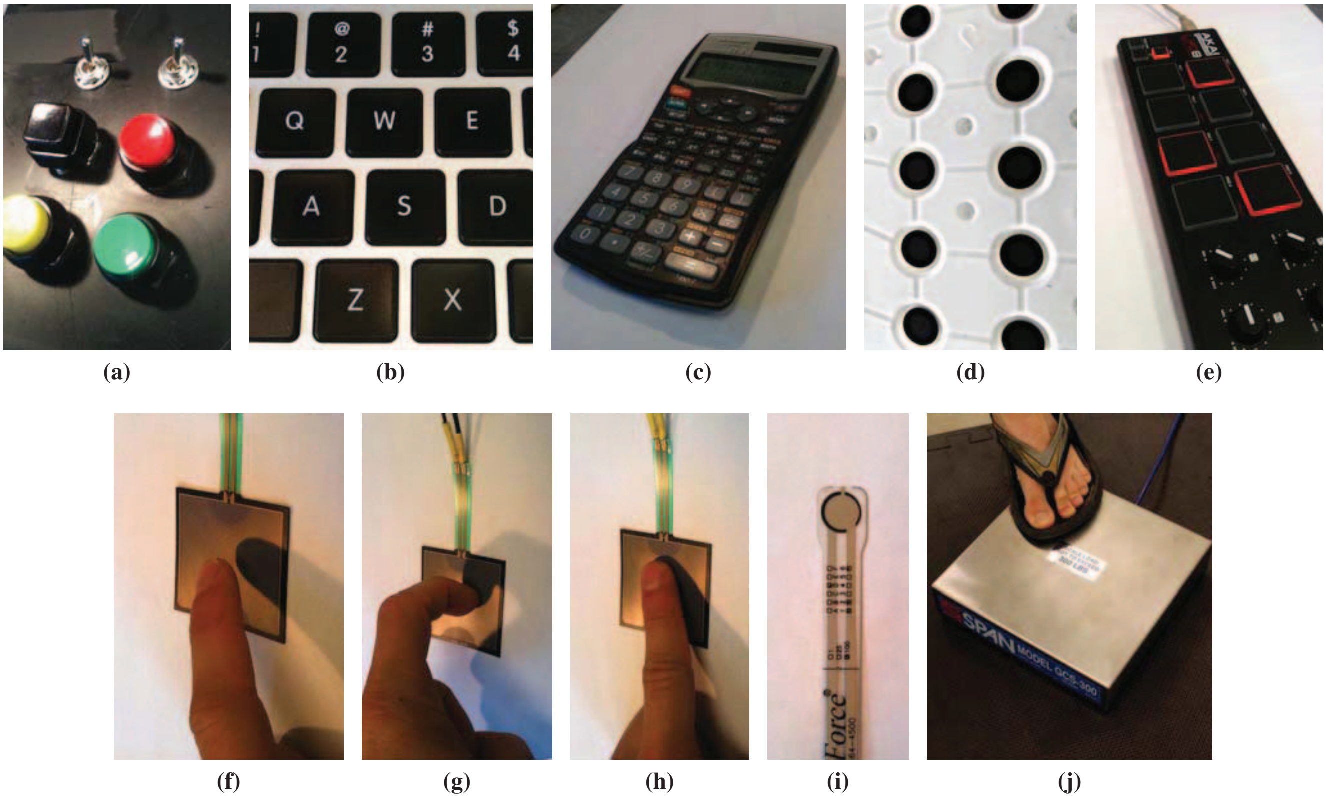

Fig. 3: Gaming, office, industrial, and mission-critical applications require user-interfaces which give a precise, low-latency response and which allow head-up operation. For head-up operation, the user must experience tactile feedback, auditory feedback, or other user-feedback, so they know to what degree they are pushing, tapping or rubbing the key. User-interface sensors with binary *v.s.* continuous response are compared here. (a) Binary buttons and switches. Only 1 bit of information can be expressed at any one time. (b) A computer keyboard also uses binary keys. We could imagine upgrading a computer interface to have analog keys, so a user could perhaps beat the typing speed record of binary keys, by being able to express continuous-time information with many fingers simultaneously. (c-d) While a calculator uses binary keypresses, the keys themselves are actually capable of sensing an analog continuous response, due to using pressure-sensitive carbon pads. (e) An analog response is actually used in a music MIDI controller. Unfortunately these keys can only trigger discrete-time events. (f-j) We desired a continuous-time, continuous-amplitude user interface, to maximize a user's control. To create this, we can employ analog sensors and sample their outputs as continuous-time analog signals, rather than triggered binary/analog events. The goal of this paper is to create acoustic user-feedback within the range of human hearing, from expressive user-interaction signals. Subsonic signals occur from tapping (f), scratching (g), rubbing (h), or stepping (j) [24–26]. Can these expressive subsonic signals be accurately and quickly frequency-shifted into the audible range? Hence, the need for swarm modulation.

This paper describes a more sophisticated frequency shifter which aims to have a fast-responding, flexible, customizable real-time response.

II. Expressive wideband low-latency frequency-time-mapping: requirements

The earlier methods were extremely limited by only accessing limited spectral or temporal information from the original sound. Achieving better expressivity and acoustic realism depends on having flexibility in mapping signals, as sensed by acoustic pickups (geophones, hydrophones, and microphones, for vibrations in solid, liquid and gas, respectively).

In this work, it was desired to flexibly map sounds across arbitrary ranges of frequencies (unlike the GC and BPF methods), cross-dependent on fast-updating temporal information. As well, we desired a low-latency response to the original sound, and a continuous, non-discretized responsiveness (*i.e.* continuous in time and amplitude.) Beyond the range of human hearing, we wished to enhance user-interface feedback by making subsonic and ultrasonic sounds audible. For example, we wanted to be sensitive to subsonic sounds such as a foot stomping on the ground (including frequencies all the way down to DC, 0Hz, which is a result of the person's static body weight). We also wished to shift ultrasonic sounds into the audible range. (For these subsonic and ultrasonic sounds we used scientific data acquisition hardware rather than simple audio inputs on a computer, or other analog-to-digital converters specifically designed for audio, which are typically bandpass filtered near the limits of human hearing.)

Flexibility of mapping, in this work, was achieved in part by parameterizing the input spectrum and temporal features, and transforming the set of parameters. A very simple parameterization is illustrated in Fig. 5, with analysis and reconstruction. A space of possible cross-dependent frequency-temporal transformations is thus made available. For the analysis stage, we initially chose a set of parameters based on

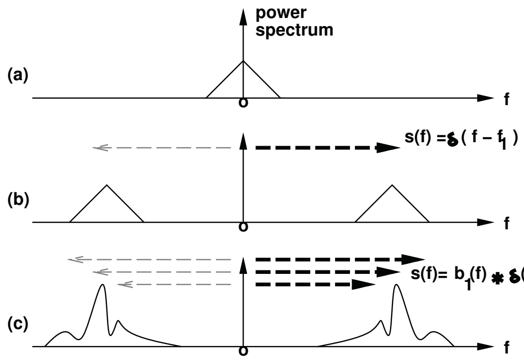

Fig. 4: A simple form of frequency-shifting (gain control) which can only handle slowly-varying baseband control inputs. (a) Input signal must be limited to low frequencies near 0 Hz, which reduces the expressivity of a user-interface. (b) Frequency-shifted output, in the most simple form, through the use of a sine wave oscillator. In the frequency domain, this is equivalent to convolving the input signal with a Dirac delta measure at frequency f_1 (and additionally convolved with a separate Dirac delta measure at $-f_1$); (c) The gain-control method simply multiplies each input signal with a desired output waveform, $b_n(f)$, at center frequency f_n.

total input spectral power within logarithmically-distributed frequency bands, as well as temporal envelope. Reconstruction, however, requires a new innovation in real-time rendering.

III. Reconstructing a Tuned Note with Nonzero Bandwidth, While Avoiding Beats

If we were willing to simply frequency-transform by having all input information control one output spectral peak, shaped like a Dirac delta measure, the reconstruction step would be trivial. In this case it is only necessary to use a single sinusoidal oscillator at the given tuned frequency, and control its amplitude. Sustained constantly for long durations, the spectrum approaches a precise Dirac delta measure shape.

It is more complex to transform one wide spectral range over to a different wide spectral range, with an arbitrary transformed shape, and zero time delay for envelope features. Let us consider the reconstruction of a wide spectral range.

It was desired to avoid using several fixed oscillators to mimic a broad spectral peak. The problem with this method is that two or more fixed oscillators produce beats. For example, two oscillators produce a beat frequency equal to the difference between their respective frequencies, which can be jarring to listen to, being mismatched to the input envelope.

The same problem can occur when rendering a Fourier spectrum in the time domain. An inverse Fourier transform is, in its very nature, a set of oscillators. The oscillators are positioned at fixed equal intervals of frequency. Another limitation of standard Fourier transforms is the frame-based segmentation of time, leading to coarse precision and latency for a real-time system. It was desired to generate a tuned note with nonzero bandwidth, with a fast update rate, continuous controllability, and without cyclic beating from fixed oscillators.

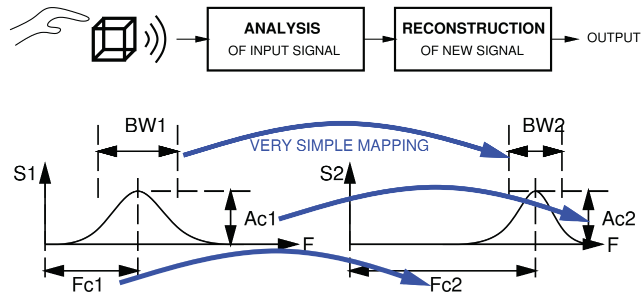

Fig. 5: (top) Transforming tactile-physiphonic sound, in analysis and reconstruction steps. (bottom) Parameterized spectral transformation: a very simple case. The output sound can be a pure tone with a well-tuned frequency, or can have a wide bandwidth. Cross-effects between the parameters can be added, as well. A further innovation of *stratiresponsive processing* will be added in Section VI, strategically taking advantage of time- and frequency-domain information.

IV. Swarm Modulation: Dynamic Phase-Incoherent Modulator Array

To solve the reconstruction-stage problem, a set of carrier waves were phase-incoherent frequency-modulated, to occupy a frequency bandwidth surrounding the tuned output frequency.

First, a set of phase-incoherent modulator waveforms were generated (stage-I). Each stage-I oscillator was given a different frequency, each of which dynamically responded to the input waveform. The stage-I oscillators were used to frequency-modulate stage-II oscillators to cause the stage-II oscillators to roam across a "territory" of spectrum. By using phase incoherence, the diverse set of stage-I modulators reduced the presence of periodic beating in the final waveform.

This can be likened to airplanes asserting national territory; each airplane only covers one point, but they patrol a fixed large area by continuously flying in distributed patterns. In this analogy, position is likened to frequency.

A more direct acoustic analogy is a string section playing in unison, where you constantly update each string player on how much vibrato to play with (both depth of vibrato and rate of vibrato), so that the group of all string players collectively occupies whatever bandwidth you desire, or a more general arbitrary spectral function that can change over time.

To make this system especially robust against beat artifacts and cyclic repetition, we constrained the stage-I modulator frequencies to consistently maintain relationships to each other that avoid cyclic patterns.

Prime numbers can be chosen to somewhat reduce cyclic repetition. But to truly prevent cyclic behaviour, it is necessary to use irrational numbers, *e.g.* sets of transcendental numbers based off of π, e, $\sqrt{\ }$, *etc.* For our purposes we used M incoherent modulator frequencies:

$$f_m^{[I]} = f_1^{[I]} \cdot e^{\delta(t)\frac{m-1}{M}} \tag{1}$$

with $m = 1 \ldots M$ modulator index, δ as stage-I frequency diversity, and $f_1^{[I]}$ as the seed frequency.

The swarm-modulated output thus consists of two stages of modulation, compounded across multiple two-stage oscillator cascades:

$$y(t)=\sum_{m=1}^{M}a_m(t)\cdot\cos\left(2\pi\int_0^t\left[f_c(\tau)+f_\Delta(\tau)\cos(2\pi f_1^{[I]}e^{\delta\frac{m-1}{M}}\tau)\right]d\tau\right) \tag{2}$$

where $a_m(t)$, $f_c(t)$, $f_\Delta(t)$, are the fast-responding control signals for amplitude, centre frequency, and bandwidth.

This scheme can be improved in how it fills the spectrum. Rather than directly modulating stage-II oscillators according to sine waves, which creates a biased emphasis on the outer extrema of spectrum bandwidth coverage, we also used an RDF [34] expansion transformation of the stage-I oscillators to compensate for extra time spent at outer extreme frequencies:

$$\begin{aligned}y(t)=&\sum_{m=1}^{M}a_m(t)\cdot\cos\\&\left(2\pi\int_0^t\left[f_c(\tau)+f_\Delta(\tau)\beta_m w^{[I]}\left(\int_0^\tau f_m^{[I]}(\gamma)d\gamma\right)\right]d\tau\right)\end{aligned} \tag{3}$$

where $w^{[I]}(t)$ is a stage-I waveform (such as $\tan(t)$) chosen such that RDF$\{w^{[I]}(t)\}$ is a rectangular, Gaussian, or other function as desired [34], and β_m is a bandwidth factor for each oscillator m, to permit terraced layering of spectra. Thus far, swarm modulation reconstructs spectral bands of various shapes and sizes, based on a center frequency, $f_c(t)$.

Alternatively, using a more general definition of swarm modulation, the stage-II oscillators can be intelligently controlled in order to fill the spectrum with arbitrary dynamic curvature, in soft-edged gradations from weak to strong intensity. This can be thought of as "painting" the frequency-time plane with a soft-edged paintbrush, with zero-latency instantaneous reaction times, rather than a hard-edged paintbrush.

V. General definition

Swarm modulation, when generalized, consists of a dispatching modulation process bank:

$$y(t)=s\{D,A\}(t)=\sum_{m=1}^{M}a_m(t)\cdot\exp\left(\mathbf{i}\int_0^t d_m(\tau)d\tau\right) \tag{4}$$

with a *dispatching matrix* and an *amplitude matrix*

$$D(t)=\begin{bmatrix}d_1(t)\\d_2(t)\\\vdots\\d_M(t)\end{bmatrix},\quad A(t)=\begin{bmatrix}a_1(t)\\a_2(t)\\\vdots\\a_M(t)\end{bmatrix} \tag{5}$$

for which the following conditions are satisfied:

- Orthogonality of stage-I dispatching:
$$\left\langle\, d_j\,(t)\,,\ d_k(t)\,\right\rangle_t=0\quad\forall j,k\in[1,M] \tag{6}$$
- Orthogonality of stage-II oscillation, in the limit:
$$\frac{1}{t}\left\langle\exp\left(\mathbf{i}\int_0^t d_j(\tau)d\tau\right),\ \exp\left(\mathbf{i}\int_0^t d_k(\tau)d\tau\right)\right\rangle_t\to 0 \tag{7}$$
as $t\to\infty$, for all $j,k\in[1,M]$.
- The duration over which any two frequencies have a rational quotient must be instantaneous. That is, any two frequencies may only instantaneously have a common divisor, whenever they "cross" each other.
$$\sup_{t\in\mathbb{R}}\left\{\Delta t\in\mathbb{R}^+:\frac{d_j(\tau)}{d_k(\tau)}=z\in\mathbb{Z}\quad\forall\tau\in[t,t{+}\Delta t]\right\}=0 \tag{8}$$
unless $a_j(t)=0$ or $a_k(t)=0$, for all $j,k\in[1,M]$.
- Common divisors between any set of three frequencies never occur:
$$\prod_{\xi}a_\xi(t)\cdot\text{GCD}\{d_i(t),d_j(t),d_k(t)\}=0 \tag{9}$$
for all $t>0$, $i,j,k\in[1,M]$ with the allowed exceptions being the initial condition ($t=0$) and whenever $a_\xi(t)=0$ for $\xi\in\{i,j,k\}$. Note that GCD refers to the greatest-common-divisor (extended to yield zero for an absence of a common divisor, *e.g.* for irrational frequencies). When pure tones are summed, their overall cyclic frequency is equal to the GCD of their constituent frequencies; therefore, a GCD of zero implies an infinitely-long period without cyclic behavior. Cyclic patterns are thus removed.

The oscillators are thus incoherent and aperiodic in the manner specified by this definition. Using sets of irrational numbers—"seeds" specified earlier—as factors, the dispatch frequencies $d_m(t)$ can furthermore be intelligently controlled.

The dispatching matrix is typically controlled such that a continuously-varying desired spectrum is spanned by time-varying elements $d_m(t)$.

VI. Stratiresponsive processing: Stratified flow rates of temporal and frequency information

The class of transformations enabled by this work allow fast-responding and slow-responding signal analysis to simultaneously control reconstruction. The transfer characteristic thus consists of a combination of fast responding components and slow responding components.

For example, a sudden input event can cause reconstruction to start immediately, with zero latency, beginning with an approximate spread spectrum, which gradually (over a fraction of a second) converges to a more precisely defined spectrum when this information becomes known.

Swarm modulation allows this immediate, real-time reconstruction, since it can dynamically change its dispatch parameters as spectral information becomes known. This can be accomplished by controlling swarm parameters according to frequent update information from time-domain analysis of the input signal (*e.g.* envelope, range density [34]) and slow update information from frequency-domain analysis of the input signal (*e.g.* periodogram).

We thus have stratified flows of fast-responding information and slow-responding information from the input to output, communicated at different rates, with (conversely) different amounts of information per update.

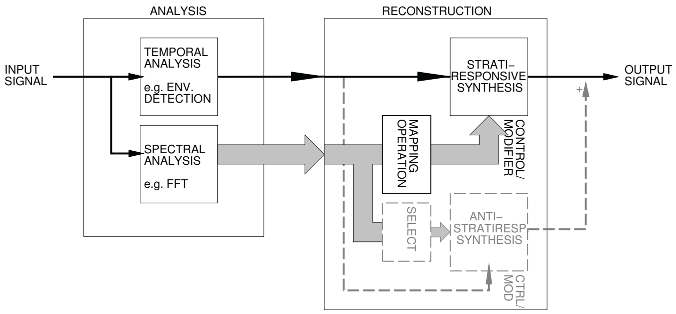

Fig. 6: Real-time spectral remapping: The new transformation uses stratified flow rates of temporal and frequency information, processed separately in terms of sparse yet quickly-updated information, and detailed, slowly-updated information. An "anti-stratiresponsive" transform would be the opposite stratification of update rates. Information flows are drawn in such a way as to suggest signals being processed from left to right, with a control/modifier signal (bottom) controlling the exact nature of that processing.

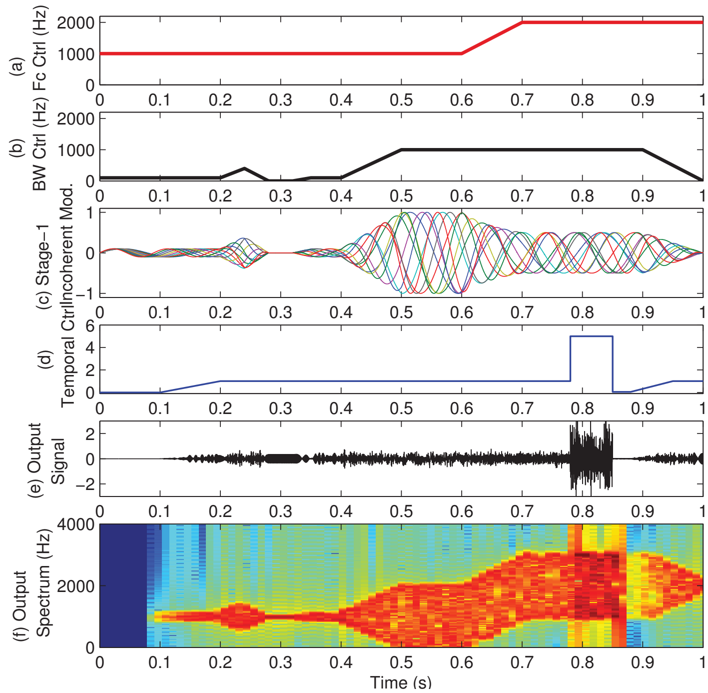

Fig. 7: Swarm Modulation: a synthesis control test. This demonstration shows three basic input parameters being controlled (center frequency (a), bandwidth (b), and fast temporal control: amplitude modulation (d)), reduced from the full space of input vectors. The final output (e,f) can be compared to the control objectives (a,b,d). Remember that this figure is *not* a test of the full stratiresponsive system, but only one element of the reconstruction stage. Swarm modulation is designed to enable stratiresponsive processing, and to be capable of transforming input signals into tuned spread-spectrum sound for use in tactile user-interfaces and hearing augmentation.

We call this a "stratiresponsive" transform, akin to stratified fluid flow, where one layer of fluid is constrained to flow slowly but has a large cross section (analogous to spectral information, containing a large amount of data per window) while another layer of fluid flows quickly according to a small available cross section (analogous to fast-updated temporal information). This analogy is illustrated by line thickness in Fig. 6.

Put another way, we can take the view that the fast-updating information is the signal being processed, and the slow-updating information controls or modifies that processing in a very detailed manner: for example, a filter processing a temporal signal, with a separate input allowing you to vary the filter coefficients in real-time.

VII. Bank of Swarm Transformations

We propagated the algorithm to create a frequency shifter-bank for an array of tactile inputs. The result is a **bank** of **banks** of **banks** of oscillators. The first "bank" is a set of independent frequency transformation systems. The second "bank" is a set of stage-II frequency- and amplitude-modulated oscillators which are phase-incoherent, inside one frequency shifter. The third "bank" is a set of stage-I oscillators, which are controlled by real-time parameters from the analysis stage.

For an array of tactile inputs, touching each tactile control can thus lead to an independent acoustic response. The transformation bank, in a simple version, is based on Equation 1 implemented in an array for N inputs, with each $f_{c,n}$ chosen as part of a musical scale. This series of frequencies is selected by choosing a starting frequency $f_{c,1}$ such as 220 Hz (an 'A'), and calculating the remaining frequencies as:

$$f_{c,n} = f_{c,1} \cdot 2^{\frac{n}{12}} \tag{10}$$

while omitting selected frequencies depending on whether a minor scale, major scale, or other scale or mode is desired. Finally, swarm modulation reconstructs a time-varying spectrum around these center frequencies, which can vary, for example, between broad spectra and narrow spectra. This can be likened to the difference between the sound of a broadly-tuned timpani versus a narrowly-tuned bell.[2]

VIII. Implementation and Results

Spectral transformations were implemented in real-time using Matlab/Simulink. The swarm-modulated reconstruction is demonstrated in Fig. 7. In particular, the dynamically-adjustable phase-incoherent modulator array produces diversity in the modulation, avoiding significant beating or phase alignment. The swarm controls operated with zero latency, while overall temporal amplitude response latency tested better than 0.15 ms (@96kS/s), which was *independent* of the

[2]Note that Equation 10 describes a concept that is unrelated to Equation 1, but these two exponential equations share the fact that their frequencies are not rational multiples of each other, leading to an interesting observation that the equally-tempered musical tuning system (described by Equation 10), predominantly used around the world at this time, has a flaw that pairs of two notes are almost but not quite in tune with each other. For example, the note 'E' above the 220 Hz 'A' is tuned as $220 \cdot 2^{7/12} \doteq 329.6$Hz in the equally-tempered system, which deviates by $\sim 0.11\%$ from the ideal tuning of $220 \cdot \frac{3}{2} = 330$Hz. Musical consonance comes from frequencies with simple fractional relationships to each other. The slight mistunings from Equation 10 create the advantage that music transposed to a different key (*i.e.* with all frequencies scaled up or down) can be played using the same set of oscillators without requiring an instrument to be re-tuned according to re-computed frequencies. The same applies to a bank of swarm modulators used to create a musical instrument.

frequency-domain subsystem. This fundamentally overcomes the frequency-time tradeoffs in the BPF and FT/IFT methods.

Results of the system running in real-time, used in a tactile user-interface, are shown by way of example in Fig. 8. Overall results are summarized in the following table.

Latency of swarm modulation (reconstruction stage): Amplitude control: 0 samples by design Frequency control: 0 samples by design
Latency, full in/out transformation: Fast stratiresponsive component: $<$ 14 samples@96kS/s* Slow stratiresponsive component: 4096 samples* * caused by choice of precision in analysis stage, and can be reduced to zero with coarse temporal analysis

Fundamental improvements, by design, in comparison to the other outlined methods, of this stratiresponsive processing:

- Avoids the temporal quantization of the FT/IFT method, and the filtering latency of the BP filtering method. Swarm modulation responds within 1 sample to envelope controls, and frequency controls. This is the fast-responding component of the stratiresponsive input-output mapping.);
- Overcomes the limited spectral sensitivity of the GC method (only baseband-sensitive) and BP filtering method (only passband-sensitive);
- Reduces the erratic beating envelope from the BP filtering method, using incoherent diverse-frequency oscillators;
- Finally, it allows control cross-linkages between different spectral bands and temporal data, between analysis and reconstruction stages (*i.e.* flexibility of spectral mapping design, with application for acoustic user-interfaces), unlike the H, GC, BPF, and FT/IFT methods.

In terms of creative options and parametric flexibility, we were able to tune analysis/reconstruction parameters to cause a low-frequency thumping vibration (detected by a geophone) to be transformed into a percussive tuned frequency at 440 Hz with fast, continuous-amplitude/time response. The system replicated the difference between gently versus sharply striking a resonant string or solid block. A gentle touch on the geophone produced a well-tuned ringing sound, whereas a sharp tapping on the geophone (additional amplitude in high frequency ranges) causing a timpani-like sound (a timpani is a broadband, partially-tuned percussion instrument). That is, the output spectrum could be spread out to mimic the more greatly broadened spectrum of the input tapping sound. By choosing an input sampling rate of 96 kS/s, we were able to enter ultrasonic frequencies on the input sensitivity list, for transformation into the sonic range. Additional shifters (forming a shifter bank) handled the additional bands and transformed them to bands we set in the sonic range.

Cross-dependent parameters allowed creative sonic responses such as one drawn in Fig. 2(right). These parameters included: number and position of input spectral ranges, input envelope, output bandwidth, output center frequency (or multiple frequencies when layering more than one shifter instance), stage-I oscillator diversity, and stage-I oscillator base FM rate.

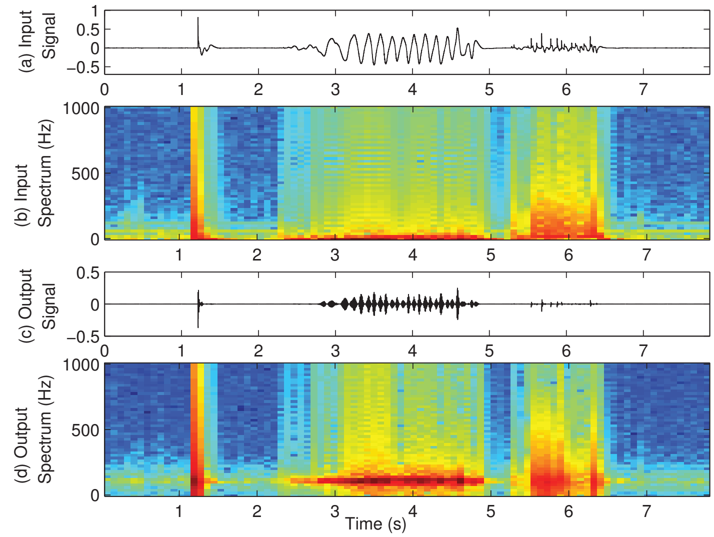

Fig. 8: Results, successfully showing stratiresponsive, fast-responding coarse control and slow-responding intricate spectral control, simultaneously. In this real-time test, subsonic vibrations from human touch (top graphs) are shifted into the audible range (lower graphs), with 0.11 ± 0.02 ms latency. Our system vastly beats the order-of-magnitude 10 ms resolution from merely using IFT/FT to resolve frequency in 100 Hz increments, for example. Input: (@1.2126sec) tap fingernail; (2.4-4.9sec) repeated dull pressing from the skin of the finger; (5.2-6.4sec) repeated tapping and scratching of the sensor.

IX. Conclusion

Swarm modulation was designed to enhance the human senses, as an augmented-reality hearing aid, and to create user-interfaces with content outside the range of human hearing.

Multimedia and gaming technologies could benefit from user-interfaces with this precise and fast acoustic response, to provide multi-sensory feedback to the user, for fast and accurate eyes-up control of buttons and keys — which is particularly desired in critical control environments such as flight-control in cockpits, automobile dashboards, or gaming environments. The human senses can be enhanced further by combining this "Composite Frequency Range" (CFR) sensing, with Composite Dynamic Range (CDR) sensing [34], Composite Acoustic Impedance Range (CAIR) sensing [35], and the sensing of sensing itself (veillametrics) [35].

Swarm modulation is demonstrated to be an effective method for real-time spectral transformation. It could have far-reaching applications in communication systems and musical synthesis, while allowing humans to hear phenomenology that is outside the range of human hearing, in real-time.

X. Demonstrations

Demonstrations, information and sample code is available at:

http://eyetap.org/swarm

References

[1] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, I," *Bell System Technical Journal*, vol. 40, pp. 43–64, Jan. 1961.

[2] H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, II," *Bell System Technical Journal*, vol. 40, pp. 65–84, Jan. 1961.
[3] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, III: The dimension of essentially time-and band-limited signals," *Bell System Technical Journal*, vol. 41, pp. 1295–1336, Jul. 1962.
[4] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, IV:Extensions to many dimensions;generalized prolate spheroidal functions," *Bell System Technical Journal*, vol. 43, pp. 3009–3058, Nov. 1964.
[5] ——, "Prolate spheroidal wave functions, Fourier analysis and uncertainty, V:The discrete case," *Bell System Technical Journal*, vol. 57, pp. 1371–1430, may-jun 1978.
[6] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.*, vol. Vol.93, no. Part III, pp. 429–457, 1946.
[7] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, pp. 292–300, Mar. 1976.
[8] S. Mann and S. Haykin, "Time-frequency perspectives: The 'chirplet' transform," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 3. IEEE, 1992, pp. 417–420.
[9] C. Starr, C. A. Evers, and L. Starr, *Biology: Concepts And Applications*. Brooks/Cole, 2006.
[10] S. Mann, *Intelligent Image Processing*. John Wiley and Sons, November 2 2001, iSBN: 0-471-40637-6.
[11] T. Hermann and A. Hunt, "An introd. to interactive sonification," in *IEEE Multimedia*, vol. 12(2), 2005, pp. 20–24.
[12] P. Cook, *Real sound synthesis for interactive applications*. Peters, Natick, MA, USA, 2002.
[13] J. Williamson, R. Murray-Smith, and S. Hughes, "Shoogle: excitatory multimodal interaction on mobile devices," in *Proc. SIGCHI conference on human factors in computing systems*, 2007, p. 121124.
[14] J. Silver, E. Rosenbaum, and D. Shaw, "Makey makey: Improvising tangible and nature-based user interfaces," in *Proceedings of the sixth international conference on tangible, embedded and embodied interaction*. ACM, 2012, pp. 367–370.
[15] A. Camurri and G. Volpe, "Multimodal and cross-modal analysis of expressive gesture in tangible acoustic interfaces," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 781–6.
[16] B. Kapralos, "The sonel mapping acoustical modeling method," Ph.D. dissertation, York University, Toronto, 2006.
[17] J. R. Parker and J. Heerema, "Audio interaction in computer mediated games," *International Journal of Computer Games Technology*, vol. 2008, p. 1, 2008.
[18] S. Mann, "Physiphones..." in *Proc. New Interfaces for Musical Expression*, 2007.
[19] C. Harrison and S. E. Hudson, "Scratch input: Creating large, inexpensive, unpowered and mobile finger input surfaces," in *ACM UIST*, Oct. 19-22 2008, pp. 205–208.
[20] S. Mann, ""fluid streams": fountains that are keyboards with nozzle spray as keys that give rich tactile feedback and are more expressive and more fun than plastic keys," in *Proc. 13th ACM int. conf. on Multimedia*, Singapore, 2005, pp. 181 – 190.
[21] S. Mann, R. Janzen, R. Lo, and C. Aimone, "Inventing new instruments based on a computational "hack" to make a badly tuned or unpitched instrument play in perfect harmony," in *Proc. International Computer Music Conference, ICMC '07, August 27-31, Copenhagen, Denmark*, vol. 1, 2007, pp. 105–112.
[22] S. Mann, R. Janzen, and R. Lo, "Hyperacoustic instruments: Computer-controlled instruments that are not electrophones," in *Proc. of the 2008 IEEE International Conference on Mulimedia and Expo, ICME2008*, June 2008, pp. 89–92.
[23] T. Lo and H. So, "Theoretical study of impact localisation limits in tangible acoustic interface applications," in *Electronics Letters*, July 30 2009.
[24] S. Mann, "The andantephone: A musical instrument that you play by simply walking," in *Proc. ACM Multimedia, October 23-27, Santa Barbara, USA*, 2006, pp. 181–4.
[25] R. Janzen and S. Mann, "*Stray:* a new multimedia music composition using the andantephone," in *Proc. ACM Multimedia 2010*, 2010, pp. 683–686.
[26] S. Mann, R. Janzen, and T. Hobson, "Multisensor broadband high dynamic range sensing..." in *Proc. Tangible and Embedded Interaction (TEI 2011)*, 2011, pp. 21–24.
[27] S. Mann, A. Sharifi, M. Hung, and R. Verbeeten, "The hydraulophone: Instrumentation for tactile feedback from fluid streams as a new multimedia interface," *ICME 2006. Proceedings.*, pp. 409–412, July 9-12, Toronto, Ontario, Canada 2006.
[28] F. Ribeiro *et al.*, "Auditory augmented reality: Object sonification for the visually impaired," in *Proc. IEEE MMSP*, 2012, pp. 319–324.
[29] S. Mann, R. Janzen, R. Lo, and J. Fung, "Non-electrophonic cyborg instruments: Playing on everyday things as if the whole world were one giant musical instrument," in *Proceedings of the 15th annual ACM international conference on Multimedia, September 24-29, Augsburg, Germany*, 2007, pp. 932–941.
[30] S. Yim, Y. Ding, and A. McCree, "Spectral transformation for musical tones via time domain filtering," in *Proc. IEEE ASSP: Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997.
[31] J. J. Wells and D. T. Murphy, "Real time spectral expansion for creative and remedial sound transformation," in *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK*, Sept. 8-11 2003.
[32] K. Yasu, M. Hishitani, T. Arai, and Y. Murahara, "Critical-band based frequency compression for digital hearing aids," *Acoustical Science and Technology*, vol. 25, no. 1, pp. 61–63, 2004.
[33] A. Simpson, A. A. Hersbach, and H. J. McDermott, "Improvements in speech perception with an experimental nonlinear frequency compression hearing device," *International journal of audiology*, vol. 44, no. 5, pp. 281–292, 2005.
[34] R. Janzen and S. Mann, "High dynamic range simultaneous signal compositing, applied to audio," in *Proc. IEEE CCECE 2012, Montreal*, April 29 to May 2 2012.
[35] ——, "Veillance flux, vixels, veillons: An information-bearing extramissive formulation of sensing, to measure surveillance and sousveillance," *Proc. IEEE CCECE2014*, May 4-7 2014, 10 pages.