

# The Ear Tone Toolbox for Auditory Distortion Product Synthesis

Alex Chechile

CCRMA, Stanford University  
chechile@ccrma.stanford.edu

## ABSTRACT

*The Ear Tone Toolbox is a collection of open-source unit generators for the production of auditory distortion product synthesis. Auditory distortion products are sounds generated along the basilar membrane in the cochlea in response to specific pure-tone frequency combinations. The frequencies of the distortion products are separate from the provoking stimulus tones and are not present in the acoustic space. Until the release of the Ear Tone Toolbox, music software for the synthesis of auditory distortion products has not been widely available. This first release is a collection of six externals for Max, VST plugins, and patches for the hardware OWL synthesizer, all of which produce various combinations of distortion products and acoustic primary tones. Following an introduction to the phenomenon and an overview on the biomechanics involved, this paper outlines each unit generator, provides implementation examples, and discusses specifics for working with distortion product synthesis.*

## 1. INTRODUCTION

Auditory distortion products (DPs), also known as combination tones or Tartini tones, are intermodulation components generated along the basilar membrane that, under certain conditions, can be perceived as additional tones not present in the acoustic space. Specifically, upon the simultaneous presentation of two frequencies  $f_1$  and  $f_2$  ( $f_2 > f_1$ ) within close ratio (typically 1.22 in clinical settings [1]), DPs appear at combinations of the stimulus frequencies [2], of which the most prominent are  $f_2 - f_1$  (the quadratic difference tone, or QDT) and  $2f_1 - f_2$  (the cubic difference tone, or CDT) [3]. If the stimulus tones are presented through free-field loudspeakers at a moderate to loud amplitude, the resulting DPs can create additional harmonic content and add spatial depth when incorporated in music.

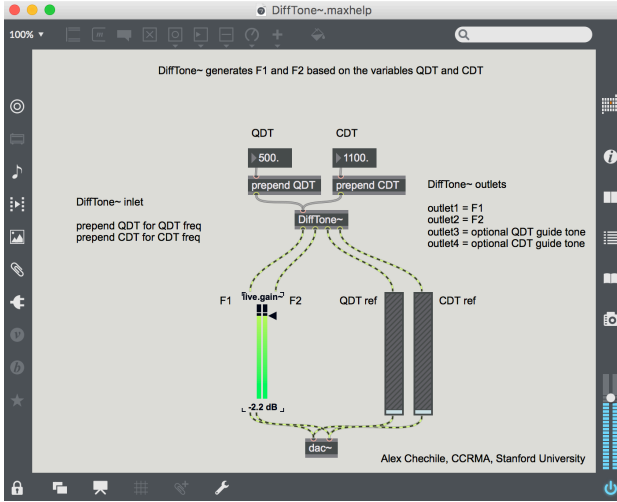
The auditory mechanisms causing DPs are primarily produced in the cochlea. When the cochlea receives sound, the basilar membrane works as a transducer to convey the sound vibrations in the fluids of the cochlea to inner hair cells, which then produce electrical signals that are relayed to the auditory brainstem through the auditory

nerve. At the same time, outer hair cells receive electric signals from the brainstem and mechanically vibrate at the frequencies of the sound [4]. This electromotility mechanically increases stimulus-specific vibrations on the basilar membrane, resulting in an increase of hearing sensitivity and frequency selectivity when transmitted to the inner hair cells [5, 6].

However, the outer hair cell movement does not occur exclusively at stimulus frequencies, but is somewhat irregular, thus making its frequency response nonlinear, extending to an audible range [4]. This nonlinear active process increases basilar membrane movement, which aids the loss of energy from damping, while an excess of the generated energy causes additional vibrations that travel backwards from the basilar membrane to the middle ear and the ear canal and creates what is known as otoacoustic emissions [7, 8]. While otoacoustic emissions can be recorded directly in the ear canal with a specially designed earpiece, DPs are specifically the intermodulation components in the inner ear.

It is not surprising that musicians were the first to discover the perception of DPs. Long before the physiological mechanisms behind combination tones were fully understood, musicians Sorge, his colleague Romieu, and Tartini individually found “third tones” produced from two acoustic tones during the middle of the 18<sup>th</sup> century [3]. In music, evoking DPs can affect the perception of the overall harmony (see Campbell and Greated for an analysis of QDT and CDT in the finale of Sibelius’ Symphony No. 1 (1899) [9]). Extending beyond harmonic content, DPs also provide additional spatial depth in music as the acoustic stimulus tones are generated apart from the DPs in the ear, and the DPs are sensed as originating in the listener’s head. Composer and artist Maryanne Amacher, known for her use of combination tones in electronic music, discussed the spatial dimension of DPs as a part of a “perceptual geography,” and she evoked such environments in immersive compositions and installations [10].

The synthesis of auditory distortion products allows for the precise calculation of perceptual tones which enables the composer and performer access to additional harmonic content, produces spatial depth between sound sources, and creates an intimate interactive listening experience. Until this point, no widely released music software allows the direct synthesis of auditory distortion products. The Ear Tone Toolbox (ETT) serves to fill this gap, while also operating as an educational tool for hearing DPs and understanding the underlying principles.



**Figure 1.** The Max help file for *DiffTone~* provides an overview of the input parameters and output signals of the external object.

## 2. EAR TONE TOOLBOX

The *Ear Tone Toolbox* is a collection of unit generators for the production of auditory distortion product synthesis. The toolbox generates the necessary acoustic primary tone combinations for evoking perceived DPs, which are not present in the acoustic space. The open-source software was written in the FAUST (Functional Audio Stream) programming language for real-time audio signal processing, and can easily compile to many architectures and formats.<sup>1</sup> In its current state, the *ETT* offers external objects for Max, VST plugins, and patches for the hardware OWL synthesizer.

The toolbox consists of six instruments that allow the user to input various combinations of evoked distortion products and acoustic primary tones. The examples discussed in this paper are in the format of Max external objects. The parameters of each object are sent to the single input, and are specified using the prepend Max object. For example, to control the QDT value, the user passes a numerical value through a prepend object with the argument QDT. The following section provides an overview of each generator in the toolbox.

### 2.1 Distortion Product Focus with *DiffTone*

The *DiffTone* generator allows the direct synthesis of user defined auditory distortion products. By specifying the desired QDT ( $f_Q$ ) and CDT ( $f_C$ ) frequencies, the instrument produces the acoustic primary tones  $f_1$  and  $f_2$  for evoking the distortion products with the equations  $f_1 = f_Q + f_C$  and  $f_2 = 2f_Q + f_C$ .

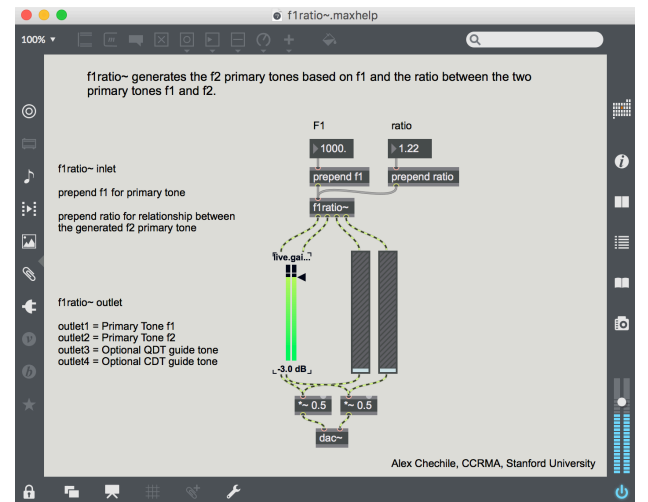
For example, if a 500 Hz QDT and an 1100 Hz CDT were input, the object would generate two sine tones at  $f_1=1600$  Hz and  $f_2=2100$  Hz. In reverse, we see the two primary tones create the desired combination tones with our original equation for the QDT as  $2100-1600 = 500$  Hz and the CDT as  $2*1600-2100 = 1100$  Hz.

The *DiffTone* generator, along with the other instruments in the toolbox, contains an optional guide tone that can be used for testing, demonstrating, or educational purposes, but otherwise should remain absent from the acoustic signal during normal use. The first and second outlets of the Max object provide the respective  $f_1$  and  $f_2$  sine waves and the third and fourth outlets provide sine wave guide tones at the QDT and CDT frequencies. Figure 1 depicts the Max help file for *DiffTone~*. Like all Max objects in the *ETT*, the input parameters are specified using the prepend object indicating which parameter the user would like to change, which in this case include prepend QDT and prepend CDT.

### 2.2 Primary Tone Focus with *flratio*

Conversely, the *flratio* unit generator is primary-tone focused, allowing the user to specify the  $f_1$  frequency and the ratio between the second primary tone. The resulting DPs occur as a byproduct of the given acoustic primary tone  $f_1$  and  $f_2$ . Although clinical applications for recording otoacoustic emissions typically use a  $f_2/f_1$  ratio of 1.22, it is possible to achieve more robust DPs at lower interval ratios, with the CDT more sensitive to ratio than the QDT [11]. Hence, *flratio* allows for experimentation between stimulus tones as the ratios can dynamically change. The  $f_2$  is calculated with the equation  $f_2 = f_1 * r$ .

With *flratio~* the input parameters are specified with prepend  $f_1$  and prepend ratio, and it produces the  $f_1$  and  $f_2$  sine wave signals from the first two outlets, and the optional guide tones for the QDT and CDT from the third and fourth outlets. The help file for *flratio~* is shown in Figure 2.



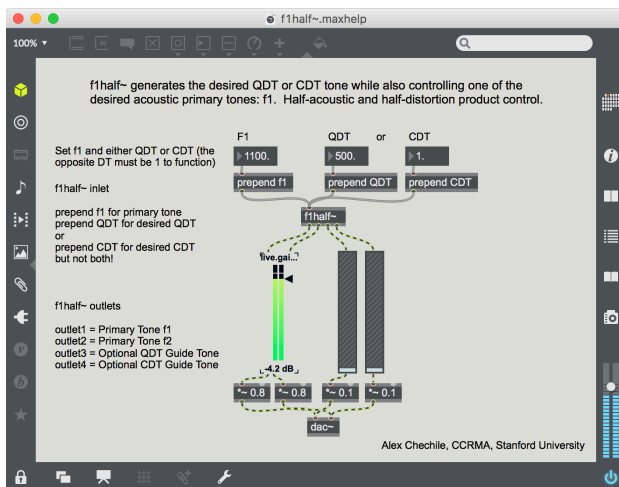
**Figure 2.** The Max help file for *flratio~* illustrates the input and output parameters of the external object.

### 2.3 Simultaneous DP and Primary Tone Control with *f1half* and *f2half*

For applications where the user requires specific control over both an acoustic component and a distortion product, the following two objects are optimal. The unit generators *f1half* and *f2half* allow the user to specify one of the acoustic primary tones (the  $f_1$  in *f1half* and  $f_2$  in

<sup>1</sup> <http://faust.grame.fr>

$f2half$ ) and either the QDT or the CDT. When the QDT is specified,  $f1half$  calculates the  $f_2$  with the equation  $f_2 = f_0 + f_1$ , and when the CDT is specified,  $f1half$  calculates the  $f_2$  with the equation  $f_2 = (2f_1) - f_c$ . Similarly,  $f2half$  calculates the  $f_1$  frequency with the equation  $f_1 = f_2 - f_0$  when the QDT is specified, and  $f_1 = (f_c + f_2)/2$  when the CDT is specified. It is important to note that both objects require only one specified combination tone while the other must be set to “1” (yet the uncalculated second DP will still be produced). The Max help file for  $f1half$  is shown in Figure 3. The first two outlets provide the sine wave primary tones  $f_1$  and  $f_2$ , and outlets three and four provide optional guide tones for the respective QDT and CDT frequencies.



**Figure 3.** The Max help file for  $f1half$ ~, which is similar to  $f2half$ ~. Note one of the QDT or CDT values in this patch must be set to “1” for the object to calculate the second primary tone.

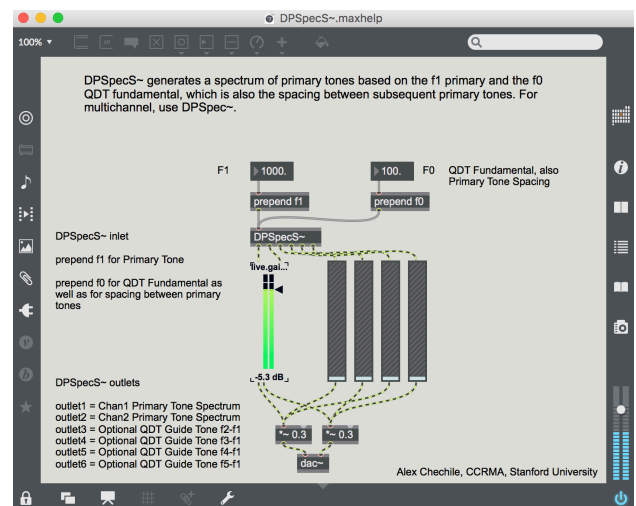
## 2.4 Distortion Product Spectrum with $DPSpecS$ and $DPSpec$

In a series of studies investigating the relationship between combination tones and the missing fundamental, Pressnitzer and Patterson found that the QDT could be perceived at lower primary tone amplitude levels if the stimulus tones were presented in a harmonic spectrum where each primary tone is spaced evenly by a constant value, and the spacing becomes the DP fundamental [12]. Since each subsequent pair of primary tones produces the same QDT, the DP spectrum is perceived at lower amplitude levels due to the vector sum of the various primary tone pairs. The study also found the level of the perceived DP increases with the number of primary tones used. As combination tones and the missing fundamental are produced by different mechanisms, DPs are perceived with primary tones that are in both harmonic and inharmonic relationship to the DP fundamental [13].

The  $DPSpecS$  unit generator creates a distortion product spectrum following the Pressnitzer and Patterson studies. The user specifies the  $f_1$  acoustic primary tone as well as the DP fundamental  $f_0$ , and the synthesizer produces a spectrum of sine waves spaced by the value of the  $f_0$ . For example, if the user specifies a  $f_1$  of 1000 Hz

and a 100 Hz QDT  $f_0$ , the object will output twelve sine waves in stereo (alternating six tones from the first outlet and the other six from the second outlet) spaced by 100 Hz. The third through sixth outlets of the Max object provide optional guide tones for the distortion product fundamental and the next three harmonics. A multichannel version of the instrument,  $DPSpec$  is also included in the  $ETT$  and provides individual outlets for each sine wave in the spectrum. The spectrum of primary tones is calculated by  $f_n = (n-1)f_0 + f_1$  where  $f_0$  is the distortion product fundamental and  $n \geq 2, 3 \dots 12$ .

The QDT between  $f_2$  and  $f_1$  equals the 100 Hz DP fundamental  $f_0$ , as does the combination tone between  $f_3$  and  $f_2$ , and so forth. The CDT between  $f_1$  and  $f_2$  is also generated (900 Hz with our example), and is emphasized by the subsequent combinations between  $f_2$  and  $f_4$ , and again between  $f_3$  and  $f_6$ , etc. Distortion products between the harmonics are also produced, although at a lower amplitude. For example, a 200 Hz QDT is produced between  $f_3$  and  $f_1$ , and between  $f_4$  and  $f_2$ , etc. The Max help file for  $DPSpecS$ ~ is found in Figure 4.



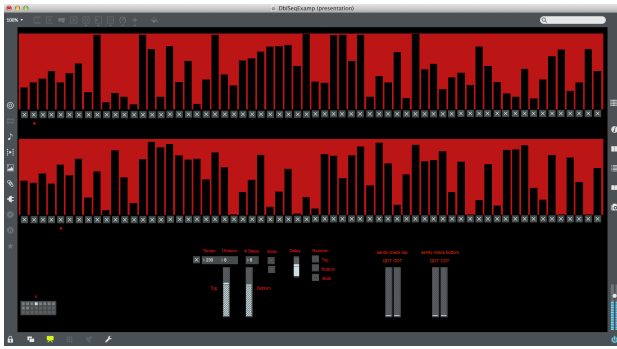
**Figure 4.** The Max help file for  $DPSpecS$ ~ illustrating the input parameters and the output signals.

## 3. IMPLEMENTATION

Synthesizing distortion products with unit generators enables the user to apply fundamental electronic music techniques<sup>2</sup> for creating larger instrument systems with a high level of creative freedom. For example, Figure 5 depicts a dual auditory distortion product sequencer system in which specific QDT and CDT frequencies can be arranged and manipulated. The two simultaneously running sequencers produce four acoustic primary tones using two  $DiffTone$ ~ objects, creating a harmonically complex distortion product spectrum. The two sequences can run in synchronous or asynchronous time, of which the latter produces an evolving spectrum of distortion products. The author uses this technique in his *On the Sensations of Tone* (2010-present) series of compositions [14].

<sup>2</sup>For an overview of select DP synthesis techniques, with audio examples, see “Sound Synthesis with Auditory Distortion Products” by Kendall, Haworth, and Cádiz [13].





**Figure 5.** *DiffTone~* implemented in a larger Max patch featuring two sequencers for producing a complex distortion product spectrum.

### 3.1 On the Sensations of Tone IX: The Descent

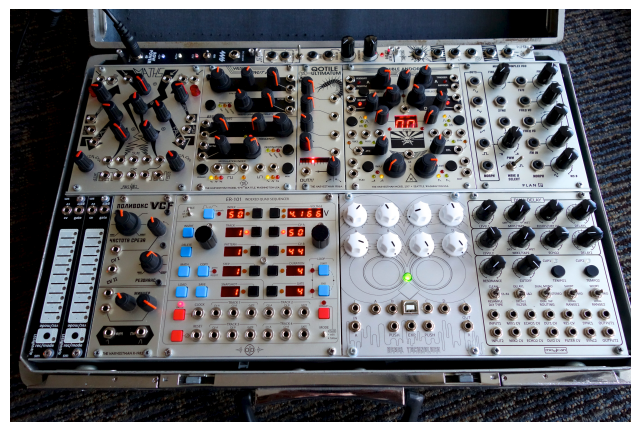
*On the Sensations of Tone* is a series of electronic and electroacoustic pieces that explore the physicality of sound and spatial depth through auditory distortion product synthesis. Presenting multiple acoustic primary tones through multichannel sound systems, the pieces evoke a complex distortion product spectrum while immersing the listener in an interactive sound field where slight head movement causes distortion products to appear, disappear, and change timbre. The structure of each entry alternates between sections that produce DPs and sections that provide contrasting non-DP material. The non-DP sections consist of live or arranged material performed on a modular synthesizer or acoustic instruments.

*On the Sensations of Tone IX: The Descent* (2015) departs from the aforementioned alternating structure as the majority of the piece is built from field recordings made in the Paris catacombs and the DP material is integrated into the recordings. Using Soundman in-ear binaural microphones<sup>3</sup> and arranged in Ableton Live, the piece relays an auditory narrative of walking through the underground tunnels below Paris. Emerging from the field recordings are two sections built using the multichannel *DPSpec* unit generator from the *Ear Tone Toolbox*. The unique character of the DP synthesis in the piece is the result of integrating the core unit generator into a larger Max patch for further processing. For example, in the second of the two parts containing DPs, the twelve individual sine waves are first amplitude modulated in unison and the resulting individual signals are modulated again in asynchrony. Barely perceivable frequency randomization was applied to both the distortion products as well as the stimulus frequency spectrum. The result of such processing creates an uneven jitter between the primary tones, and evokes fluctuating combinations of DPs.

### 3.2 Modular Synthesis

In addition to unit generators for computer-based sound synthesis, the *ETT* has been compiled for use on the open-source and open-hardware OWL synthesizer. Built

by the London-based collective Rebel Technology, the programmable synthesizer is available as a standalone pedal or as a eurorack module. Both versions contain a STM32F4 microcontroller with a 168 MHz 32bit ARM Cortex M4, 192Kb RAM, 1MB of flash memory and a sampling rate adjustable up to 96kHz.<sup>4</sup> The eurorack version (see Figure 6) allows for control voltage (CV) control over each instrument parameter in the toolbox. The OWL patches are written in C/C++, or the synthesizer can run patches using PureData with Heavy, and FAUST code using faust2owl. An online repository hosts the OWL *Ear Tone Toolbox* patches.<sup>5</sup> Given the form factor of the hardware, the multichannel version of *DPSpec* is unavailable, and the primary tone spectrum is generated with a reduced number of sine tone oscillators. Apart from *DPSpec*, the rest of the instruments in the *Ear Tone Toolbox* run on the OWL similarly to their software counterparts.



**Figure 6.** The *Ear Tone Toolbox* on the OWL synthesizer eurorack module (bottom row, second from the right).

## 4. DISCUSSION AND CONCLUSION

Auditory distortion product synthesis offers additional parameters for consideration during music composition and performance, and the result provides a unique listening experience for the audience. The direct control over combination tones creates more harmonic and melodic material, which can enhance or intentionally disrupt acoustic material. The distinction between sounds emerging from speakers and other sounds generated within the listener's ears creates an added spatial depth to the work. DPs encourage exploration and interaction with sound as head position and listener location in the acoustic field can produce different combination tones.

After a performance, members of the audience unfamiliar with DP synthesis often report the experience as unlike any other listening experience. The intimacy of allowing a layer of the music to be generated within one's own ears is, however, not for everyone. Audience members suffering from tinnitus or hearing loss

<sup>3</sup>Although recorded in binaural, the final composition with the DPs is intended for concert presentation using free-field speakers.

<sup>4</sup><http://www.rebeltech.org/products/owl-modular/>

<sup>5</sup><https://hoxtonowl.com/patch-library/>

occasionally report the experience as disagreeable, or in the latter case, not perceivable.

Aside from the benefits it provides, DP synthesis is not without limitations. The amplitude levels for evoking a strong DP response through free-field speakers often lies within the approximate range of 84-95 dB SPL, which makes reproduction for home recordings or internet distribution difficult. The author prefers to reserve compositions with DP synthesis for concert settings where he can control the amplitude, and make accommodations for the acoustics of the venue. Preparing the audience with a pre-concert introduction to the phenomenon allows the listener to understand the unique aspects of the experience. Under controlled conditions, the author finds the majority of the audience's reaction is positive.

The *Ear Tone Toolbox* is the first widely available software package for distortion product synthesis. The six unit generators described in this paper comprise the initial release of the toolbox, with regular updates and additional generators planned. By releasing the software open-source, the author intends to encourage the future development of the field of distortion product synthesis, and to provide educational tools for listening and understanding the fundamentals of combination tones.

## Acknowledgments

The author would like to thank Maryanne Amacher, Chris Chafe, Brian Ferneyhough, Takako Fujioka and Pauline Oliveros for their mentorship and support, as well as Cathleen Grado, Romain Michon, and his colleagues at Stanford's Center for Computer Research in Music and Acoustics.

## 5. REFERENCES

- [1] J. W. Hall III, *Handbook of Otoacoustic Emissions*. San Diego, California: Singular Publishing Group, 2000.
- [2] H. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 2nd English ed. New York: Dover Publications, 1954.
- [3] R. Plomp, *Experiments on tone perception*. Soesterberg: National Defense Research Organization TNO. Institute for Perception RVO-TNO, 1966.
- [4] W. E. Brownell, "Outer hair cell electromotility and otoacoustic emissions," *Ear Hear*, vol. 11, no. 2, pp. 82–92, Apr. 1990.
- [5] T. Gold, "Hearing. II. The Physical Basis of the Action of the Cochlea," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 135, no. 881, pp. 492–498, Dec. 1948.
- [6] H. Davis, "An active process in cochlear mechanics," *Hear. Res.*, vol. 9, no. 1, pp. 79–90, Jan. 1983.
- [7] D. T. Kemp, "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.*, vol. 64, no. 5, pp. 1386–1391, Nov. 1978.
- [8] D. T. Kemp, *The OAE Story*. Hatfield, UK: Otodynamics, 2003.
- [9] M. Campbell and C. A. Greated, *The musicians' guide to acoustics*, 1st American ed. New York: Schirmer Books, 1988.
- [10] M. Amacher, "Psychoacoustic Phenomena in Musical Composition: Some Features of a 'Perceptual Geography'," in *Arcana III*, New York, NY: Hips Road, 2008.
- [11] J. L. Goldstein, "Auditory Nonlinearity," *The Journal of the Acoustical Society of America*, vol. 41, no. 3, pp. 676–699, Mar. 1967.
- [12] D. Pressnitzer and R. D. Patterson, "Distortion Products and the Perceived Pitch of Harmonic Complex Tones," in *Physiological and Psychophysical Bases of Auditory Function*, D. J. Breebart et al., eds., The Netherlands: Shaker Publishing BV, 2001.
- [13] G. Kendall, C. Haworth, and R. F. Cadiz, "Sound Synthesis with Auditory Distortion Products," *Computer Music Journal*, vol. 38, no. 4, Winter 2014.
- [14] A. Chechile, "Creating Spatial Depth Using Distortion Product Otoacoustic Emissions in Music Composition," presented at the International Conference on Auditory Display, Graz, Austria, 2015, pp. 50–53.