BNAM2012
Joint Baltic-Nordic Acoustics Meeting
June 18th - 20th 2012 Odense, Denmark

Can components in distortion-product otoacoustic emissions be separated?

Anders Tornvig
Section of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, DK-9220 Aalborg Ø, Denmark, tornvig@es.aau.dk

David W. Purcell
National Centre for Audiology, Western University, 1201 Western Road, London, Ontario, N6G 1H1, Canada

Flemming Christensen and Dorte Hammershøi
Section of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, DK-9220 Aalborg Ø, Denmark

Otoacoustic emissions are signals emitted from the cochlea, either spontaneously or evoked by stimuli. Measured with an acoustic probe sealed in the ear-canal, they reveal information about a part of the mechanism of hearing that is otherwise inaccessible. Outer hair cells in the cochlea work to improve hearing sensitivity by means of nonlinear amplification, which produces distortion. In the measurement of otoacoustic emissions, two tones can be delivered to the cochlea to invoke this nonlinearity and elicit the distortion-product otoacoustic emission (DPOAE). DPOAEs arise mainly from two spatially separated generation mechanisms, thus making interpretation of DPOAE measurements complicated. In this study, we test whether or not source separation by group delays is equivalent to separation by time delays – either result is equally interesting to understand given the complexity of the cochlea. Specifically, DPOAEs from 500 Hz to 4000 Hz were measured in 20 normal-hearing human ears, using exponentially swept, fixed-ratio primaries. Source separation by group delays relied on gating in the DPOAE latency domain, obtained by an inverse Fourier transform of DPOAE magnitude and phase. Source separation by time delays relied on adding physical delays, suggested by the group delays, to a model used to estimate the swept DPOAEs.

1 Introduction

Otoacoustic emissions are sounds which can be recorded in the ear canal of most people with normal hearing [1]. They reflect the micro-mechanical activity of the outer hair cells within the cochlea and are useful for indirectly evaluating the health and function of the inner ear. They have played a major role worldwide in infant screening programs by reducing the age of identification of permanent hearing loss from three years, typically, to the first months of life.

Sending two tones into the cochlea invokes a nonlinear amplification by the outer hair cells. The nonlinearity produces inter-modulation distortion, most prominently at $f_{dp} = 2f_1 - f_2$, where f_1 and f_2 are the frequencies of the two-tone stimulus. The emissions produced at f_{dp} travel back out into the ear canal where they can be measured by a probe microphone. Emissions invoked in this way are called distortion-product otoacoustic emissions (DPOAEs). In a simplified view on the generation of

DPOAEs, known as the two-source model, DPOAE stimuli lead to emissions from two places in the cochlea. The emissions combine as a vector sum on their way back to the ear canal. Since the sources are spatially separated, one arrives before the other. The one arriving first is the *generator component*, enabled by the nonlinear distortion mechanism near the characteristic places of the stimuli. The *reflection component*, enabled by the linear reflection mechanism near the characteristic place of f_{dp}, arrives a little later. The difference in generation mechanism, identified by Shera and Guinan [2], gives rise to differences in the phase responses of the components. The relative phase of the generator component is largely independent of frequency, while the relative phase of the reflection component changes approximately linearly in a logarithmic frequency domain, as long as the frequency ratio between the stimulus tones is kept constant. These relatively well-documented properties of the two components are useful to know when trying to isolate them in a DPOAE measurement.

Separation of DPOAE components is interesting, because it may enable assessment of outer hair cells in very specific regions along the basilar membrane. The components likely reveal different information about the state of hearing. For instance, the reflection mechanism may be more affected by efferent activity, which is interesting in understanding feedback systems in the cochlea and brainstem.

In this paper, we investigate differences between two methods for isolating the generator component and the reflection component in a DPOAE measurement. The work is motivated by sparse documentation of the methods, custom implementations of them across laboratories, and a tendency in recent literature to assume them to be equivalent, or at least disregard differences.

2 Methods

We have measured DPOAEs from 500 Hz to 4000 Hz in 20 normal-hearing human ears using fixed frequency-ratio stimulus sweeps. Such sweeps are generally attractive, because they allow for rapid measurements of high-resolution DPOAEs across a broad band of frequencies. Components are separated based on the group delay, given by the slope of the DPOAE phase in the logarithmic frequency domain. We refer to this method as the group delay method. In addition, the components are separated based on the time delays, suggested by the found group delays. We refer to this method as the time delay method. Both implementations are based on bits of information spread across the literature. Several comparisons of the impact of specific configuration parameters were carried out along the way to make the most sensible choices for their values.

2.1 Experiment

DPOAEs were measured in 20 normal-hearing human ears. Both ears of 8 females and 2 males were included. Mean age was 25.6 years, ranging from 22 to 33 years. Otoscopy was performed before measurements to ensure nothing of significance would block the transmission of stimulus and response between probe tip and ear drum. Measurements were carried out in a sound attenuated and electromagnetically shielded room at the National Centre for Audiology at Western University.

The stimulus was two exponential sine sweeps, delivered simultaneously to one ear of the subject through two Etymotic ER2 tube insert earphones and an Etymotic ER10B+ emissions probe. The stimulus sweeps were designed to elicit the DPOAEs in 3 octaves from 500 Hz to 4000 Hz. The sweep rate was 8 s/octave, the f_2/f_1-ratio was constant at 1.22, and the levels were, on average, 65 dB SPL for the f_1 sweep and 55 dB SPL for the f_2 sweep. This 24 s stimulus was repeated 10 times in each ear to achieve very good signal-to-noise ratios after noise rejection and averaging.

The response was recorded in the ipsilateral ear using the Etymotic ER10B+ probe system with microphone and amplifier (total gain 60 dB with bandpass filtering from 200 Hz to 10 kHz).

2.2 DPOAE magnitude and phase estimation

The DPOAE magnitude and phase are estimated from the average recorded response to the sweep stimulus. A windowed, parametric sweep model is fit to the response at steps of 1/20 s by minimization of a least squares error. The magnitude and phase of the DPOAE can be estimated from the found, optimal parameter values. The basics of this least-squares fit algorithm is described by Long and Talmadge [3]. Long et al. [4] used it with continuously swept stimuli.

A Hanning window is used for windowing of the recorded signal as well as the model. A long window in the time domain gives a narrow main lobe in the frequency domain. Since the complex DPOAE data are estimated at discrete exponentially spaced frequencies over time, a narrow main lobe also gives a higher temporal resolution. With a high temporal resolution, delayed parts of the response will not be included in the magnitude and phase estimate. Therefore, making the main lobe sufficiently narrow will exclude the reflection component from the estimates.

To include both the generator component and the reflection component, it is customary to use a window size of 0.5 s [4]. As the analyzed signal changes frequency exponentially with time, the bandwidth of the signal increases relative to the constant main lobe size. In effect, the temporal resolution increases as the upward exponential sweep progresses. The 0.5 s Hanning window has a temporal resolution of 92.3 ms at 500 Hz, meaning that responses delayed up to 46 ms are included in the main lobe. At 4000 Hz the resolution is 11.5 ms. No attempt was made to correct the window attenuation of responses, which arrive near the edges of the main lobe, compared to responses at the center.

2.3 Source separation by group delay

The DPOAE response is recorded by sampling the time domain t at equally spaced intervals. The conjugate domain of t, connected via Fourier transformation, is the frequency domain f. While the phase of the generator component is largely independent of frequency, the slope of the DPOAE phase, the group delay, is approximately constant for the reflection component in the logarithmic frequency domain [5]. Separating sources by group delay over a broad frequency range therefore requires working in the logarithmic frequency domain ν, defined by

$$\nu(f) = -\log(f/f_{ref}), \tag{1}$$

where $\log$ is the natural logarithm and f_{ref} can be chosen to be the upper limit of hearing at approximately 20 kHz. The approximate scaling symmetry of the cochlea – any cochlear wave completes the same number of cycles N within the cochlea, regardless of its frequency and travel distance – gives a phase, proportional to the logarithmic frequency variable

$$\theta(f) = -N\log(f/f_{ref}) \iff \theta(\nu) = N\nu \tag{2}$$

The group delay τ is then

$$\tau(\nu) = -\frac{\partial\theta(\nu)}{\partial\nu} = -N \qquad \wedge \qquad \tau(f) = -\frac{\partial\theta(f)}{\partial f} = N/f \tag{3}$$

Like t and f, N and ν are variables of conjugate domains. ν is sampled linearly by sampling f exponentially. Taking the Fourier transformation of an exponentially sampled f domain brings the data into the N domain, termed the latency domain [6]. The latency is measured in periods of the stimulus frequency. When the frequency ratio between the stimulus sweeps is constant, the latency of the reflection component is approximately constant at 15 cycles in humans [5]. The sampling interval of the N domain is $1/(\log(2^n) \cdot 2)$, where n is the number of octaves sampled.

Windows can be multiplied onto the DPOAE latency domain to isolate the components. We found it convenient to implement this as a circular convolution in the frequency domain. The recorded DPOAE data represents a complex, one-sided spectrum of the logarithmic frequency domain. Before convolution, a mirrored and complex conjugated version of it is appended at the end. The middle point and the DC point are not included. The window is designed in the latency domain. It is a recursive, exponential filter used by Kalluri and Shera [6]. A window order of 10 is used here to obtain a sharp cutoff in the N domain, while not weighting parts within the window too much. A window of 18 cycles is placed around a 0 group delay for isolation of the generator. For the reflection, a window of 16 cycles is placed around a group delay of 15 cycles. These windows overlap slightly, meeting where they both attenuate 3 dB in the N domain. In the ν domain, the bandwidths are ~0.1 octaves.

After separation of the components by windowing in the latency domain, a linear regression is carried out on the phase response of each component in the logarithmic frequency domain. The phase slopes, N_G and N_R, are found by fitting the unwrapped phase curves to

$$\theta_G(f) = N_G \log(f/f_{ref}) + k_G \qquad \wedge \qquad \theta_R(f) = N_R \log(f/f_{ref}) + k_R, \tag{4}$$

where k_G and k_R are constants, disregarded here. f is sampled exponentially. The regressions gave correlation coefficients closely around 0.886 and 0.996 – straight lines fitted the data nicely.

2.4 Source separation by time delay

Source separation by time delays is described by Long et al. [4]. The model, fit to the recorded signal by the least-squares fit algorithm, can be modified to assume a frequency dependent delay of the response. This delay is included by subtraction of the integral of equation (3) from the phase in the original model. Then the model expects a frequency slightly lower than f_{dp} at a given point in time.

The delays added to the model are given by N_G and N_R, which were found by linear regression in the group delay method. The Hanning window is also changed to 2 s long, giving it four times the temporal resolution, compared to the previously described 0.5 s window. This effectively excludes the component outside the window from the magnitude and phase estimates.

2.5 Noise estimation

Noise is estimated by applying the methods above with the negative of the group delay found for the reflection. In effect, the noise is estimated from what was present a little before the stimulus evoke responses there. There is a risk that these noise estimates include parts of higher-frequency responses.

3 Results

Figures 1, 2, and 3 show measurements, estimated generator and reflection components, and the noise.

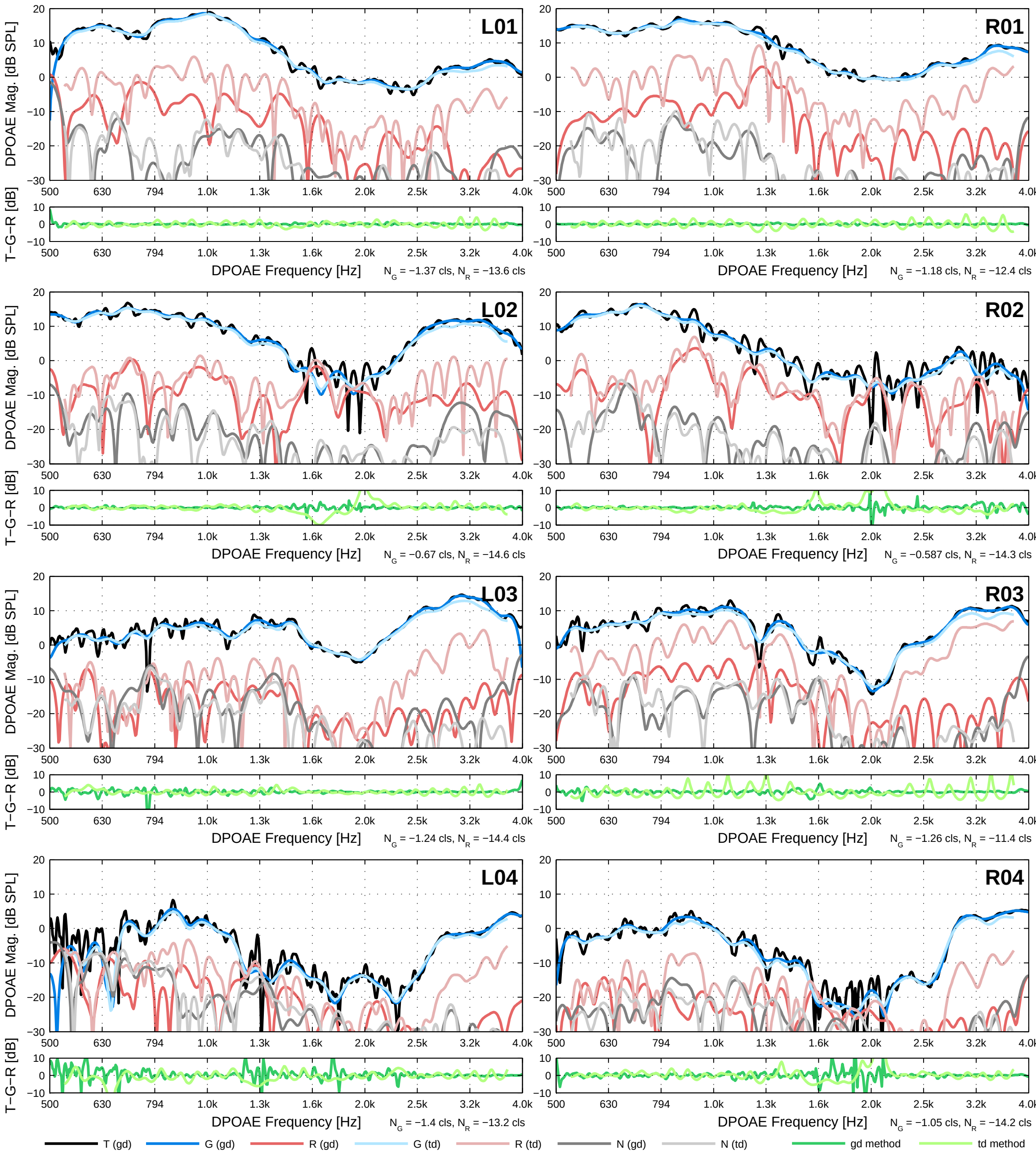

Figure 1: DPOAE magnitudes of left and right ear of subjects 1-4. The upper plots show the total (T) DPOAE (black line), the generator component (G) for both separation methods (blue lines), and the reflection component (R) for both methods (red lines). The noise (N) estimates from both methods are plotted as well (gray lines). The green lines indicate T-G-R differences. A zero T-G-R difference means that the separated components make up the total entirely. N_G and N_R are the phase slopes estimated by linear regression.

The pairs of colors indicate how similar the two methods are. The green lines indicate how close the sum of blue and red is to the black line. No phase response is shown here, but the group delay method generally had the phase of the total DPOAE response completely reconstructed after addition of the separated components. The time delay method does not preserve the phase well enough to do so.

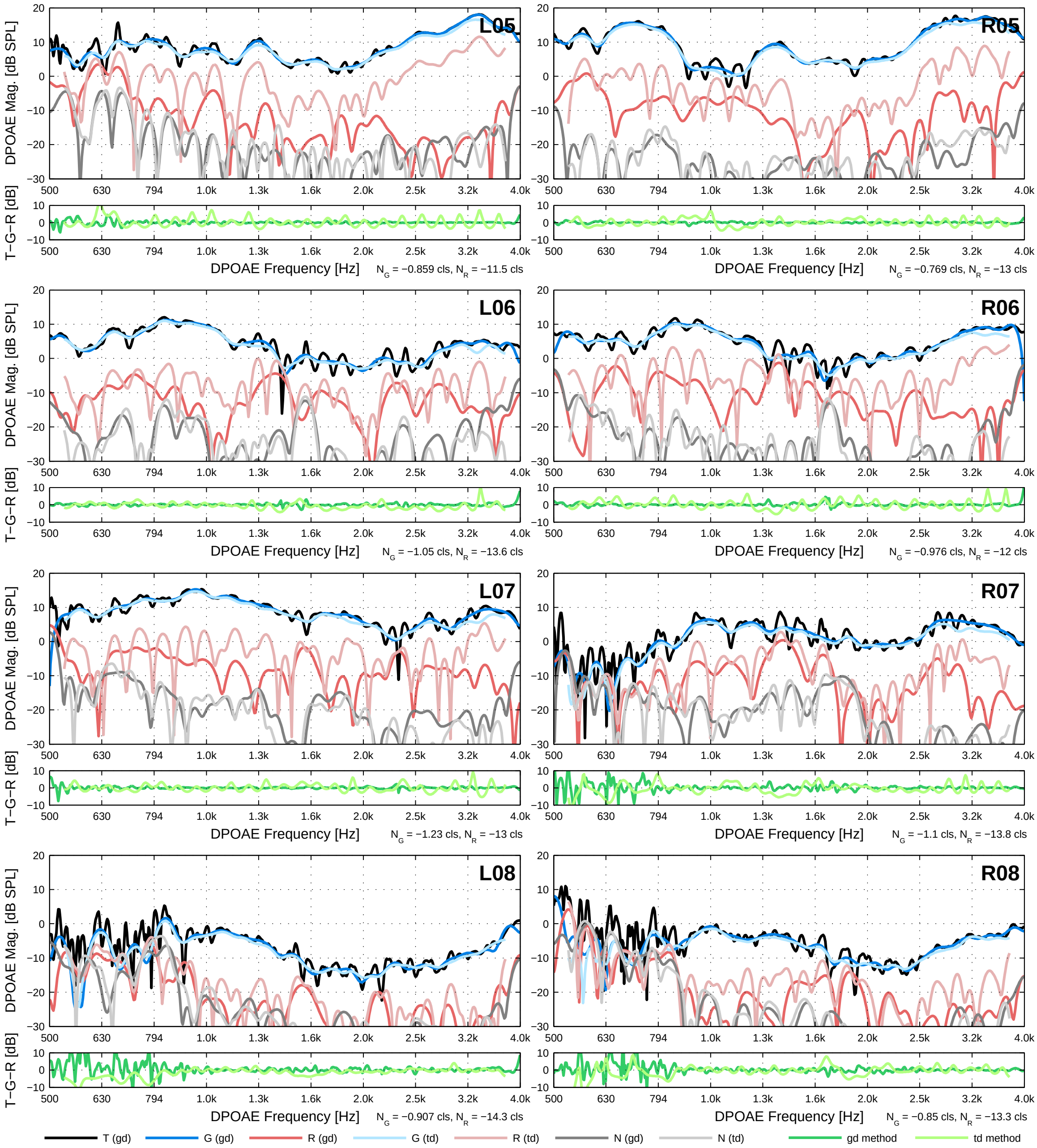

Figure 2: DPOAE magnitudes of left and right ear of subjects 5-8. See caption of Figure 1.

The generator components of subject L01, L03, R06, L07, R09, L10 and R10 rise or fall steeply near the ends of the spectra because the circular convolution wraps around in the complex, mirrored frequency domain. Other convolution methods were tested, giving other undesired artifacts. Eventually, the basic method, described in section 2.3, was selected.

In most measurements (e.g. R06), the time delay separated reflection component is larger than the group delay separated reflection component at higher frequencies. The narrow window applied is apparently still too broad at the high frequencies, making it include parts of both components. Related, the estimated group delays N_G and N_R lie consistently around -1.06 cycles across subjects for the generator and -13.3 cycles for the reflection. This is a result of actual delay in the cochlea, but biased by the latency windows applied, not allowing large deviations from their centres at 0 and 15 cycles.

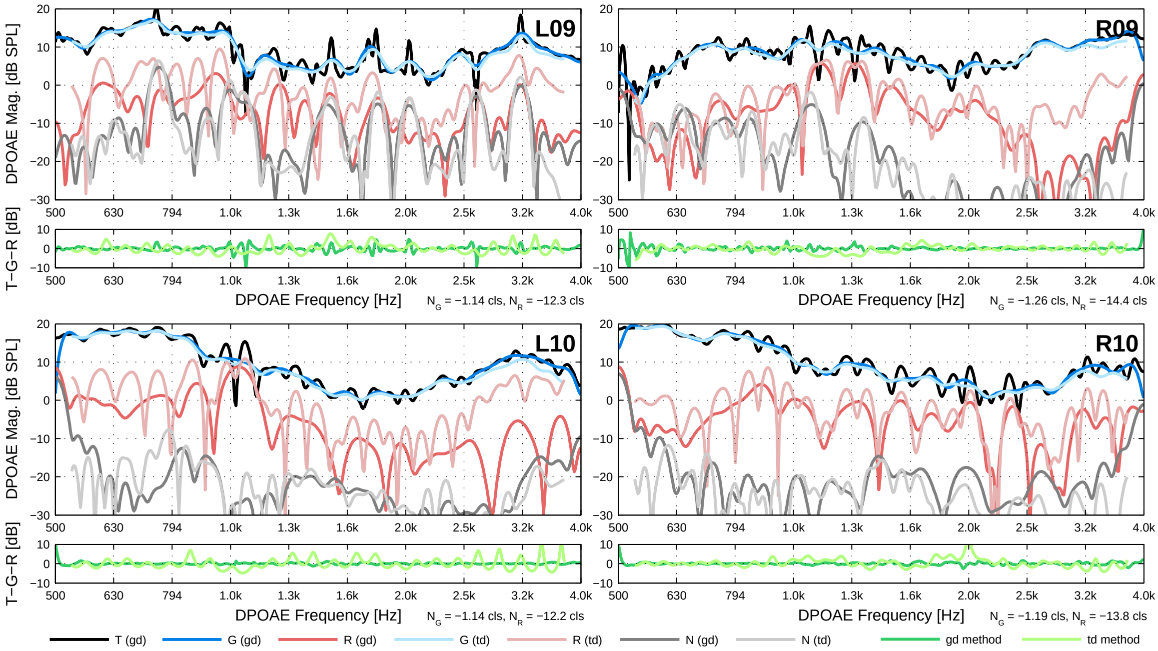

Figure 3: DPOAE magnitudes of the ears of subjects 9-10. See caption of Figure 1.

4 Discussion

The patterns of the separated components are generally comparable with those reported in previous literature [6, 7]. The generator component includes the slowly varying nonlinearity of different outer hair cells along the basilar membrane. Its magnitude is relatively large and without sudden changes. The reflection component on the other hand shows what resembles a fine-structure pattern, predicted by linear coherent reflection theory [5]. It should be the same pattern as that evoked by just one stimulus sweep at the distortion-product frequency. When the two components are similar in magnitude, the total response exhibits another layer of fine-structure because the two components, one having a rapidly changing phase, are close in frequency when they sum in the ear canal. Constructive and destructive interference between the components give the fine-structure, which our data shows.

Some interesting observations were made, not shown by the figures here. Changing the N's to be slightly different (1 cycle for instance) between the two methods gave significantly different results in the estimation of the reflection component. This verifies that the methods are related via the latency parameter N. Furthermore, changing the assumed latency N of the reflection component to be a few cycles different from 15 cycles gave much larger T-G-R differences. This means that the separated components do not describe the total DPOAE well. Components of similar magnitudes give fine-structure in the total response. This important characteristic also largely vanished when changing N to be slightly different. Window placements around 0 and 15 cycles in the latency domain are good initial choices for our data – data which are to some extent representative of adult, healthy ears.

A major difference between the methods is the point of application. The group delay method is applied after the least-squares fit estimation of the DPOAE response, so that the method depends on the success of the least-squares fit algorithm. The estimates from the least-squares fit algorithm are affected by the window attenuating delayed responses. Also, the phase response is dominated by the larger-magnitude component. This can render the reflection inseparable by a group delay method.

The time delay method on the other hand is applied as a configuration of the least-squares fit estimation. But it requires assumptions about individual delays of responses which are generally not known.

Another difference between the methods is their operation with different windows in two different conjugate pairs of domains. Here, the windows were made similar in terms of bandwidth and application point in time. However, convolution artifacts still occur when converting between the domains.

In conclusion, separation of components in DPOAE measurements is motivated by the possibility of assessing the activity of outer hair cells in very specific regions of the cochlea. Two ways of separating DPOAE components have been implemented. The group delay method separates the components based on phase slopes. The time delay method separates the components based on times of arrival at the ear drum. Our analysis of data, measured in normal-hearing human ears, shows that similarly configured implementations give identical results for separation of the generator component. They do not give identical results for separation of the reflection component. They are, however, similar in magnitude and differences may be attributed to the fact that the methods are applied after and before the least-squares fit DPOAE magnitude and phase estimation. Dissimilarity is also caused by different window types being applied in different domains. It remains unanswered what method is most suitable for source separation. More needs to be learned about the limitations of the methods.

5 Acknowledgements

Thanks to Sriram Boothalingam for his help transitioning the first author into the rapidly evolving DPOAE literature. Sincere thanks to all who participated in our experiment.

References

[1] D. T. Kemp. “Stimulated acoustic emissions from within the human auditory system”. *J. Acoust. Soc. Am.*, 64:1386–1391, 1978.

[2] C. A. Shera and J. J. Guinan Jr. “Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs”. *J. Acoust. Soc. Am.*, 105:782–798, 1999.

[3] G. R. Long and C. L. Talmadge. “Spontaneous otoacoustic emission frequency is modulated by heartbeat”. *J. Acoust. Soc. Am.*, 102:2831–2848, 1997.

[4] G. R. Long, C. L. Talmadge, and J. Lee. “Measuring distortion product otoacoustic emissions using continuously sweeping primaries”. *J. Acoust. Soc. Am.*, 124:1613–1626, 2008.

[5] G. Zweig and C. A. Shera. “The origin of periodicity in the spectrum of evoked otoacoustic emissions”. *J. Acoust. Soc. Am.*, 98:2018–2047, 1995.

[6] R. Kalluri and C. A. Shera. “Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation”. *J. Acoust. Soc. Am.*, 109:622–637, 2001.

[7] R. H. Withnell, L. A. Shaffer, and C. L. Talmadge. “Generation of DPOAEs in the guinea pig”. *Hearing Research*, 178:106–117, 2003.