Interpreting otoacoustic emissions in humans: Evidence for multiple generating mechanisms

by

Radha Kalluri

B.Sc. Electrical Engineering
University of Massachusetts, Amherst, 1996

M.Sc. Electrical Engineering
University of Southern California, 1997

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

©Radha Kalluri, 2006, All rights reserved

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document
in whole or in part in any medium now known or hereafter created.

Author ........................................

Harvard-MIT Division of Health Sciences and Technology

May 18, 2006

Certified by ....................................

Christopher A. Shera
Associate Professor of Otology and Laryngology
Harvard Medical School
Thesis Supervisor

Accepted by .................................

Martha L. Gray
Edward Hood Taplin Professor of Medical and Electrical Engineering
Director Harvard-M.I.T. Division of Health Sciences and Technology
Interpreting otoacoustic emissions in humans: Evidence for multiple generating mechanisms

by

Radha Kalluri

B.Sc. Electrical Engineering
University of Massachusetts, Amherst, 1996

M.Sc. Electrical Engineering
University of Southern California, 1997

Submitted to the Harvard-MIT Division of Health Sciences and Technology on May 18, 2006, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

Healthy ears generate sounds known as otoacoustic emissions that can be evoked and measured in the ear-canal using small, low-noise microphones. The ability to measure acoustic signals that originate within the cochlea provides noninvasive access to what in humans is an almost inaccessible organ. Although otoacoustic emissions (OAEs) are frequently used as noninvasive probes of cochlear function in both the clinic and the laboratory, their utility is limited by incomplete knowledge of their generating mechanisms. A recently proposed model suggests that most OAEs are mixtures of emissions arising by two fundamentally different mechanisms: 1) nonlinear distortion induced by cochlear traveling waves and 2) linear reflection of those waves from pre-existing micromechanical impedance perturbations. The model predicts that OAEs generated by wave-induced perturbations manifest a phase that is nearly-frequency invariant whereas OAEs generated by reflection from pre-existing perturbations manifest a phase that rotates rapidly with frequency. The model suggests that the relative contribution from each mechanism to any emission measurement depends on factors such as the type and intensity of the evoking stimulus. In this thesis we tested the relationships between common OAE measurements and the two proposed mechanisms of OAE generation.

We tested the two-mechanism model by measuring and comparing OAEs evoked with single tones and broad-band clicks, as well as those evoked by two-tone complexes at frequencies not contained in the stimulus, so-called distortion-product emissions. Our results indicate that click-evoked and tone-evoked OAEs, previously regarded as different types of emission based on the characteristics of the stimuli used to evoke them, are really the same emission evoked in different ways. The phase characteristics of both emission types are consistent with those predicted for emissions originating by linear-reflection from pre-existing perturbations. In addition, we demonstrate that distortion-product OAEs are often mixtures of two components. By separating the two components we show that one component arises by linear reflection and the other
component arises by induced distortion. Our results provide strong empirical support for the two-mechanism model of OAE generation. Since the two emission mechanisms depend on fundamentally different aspects of cochlear mechanics, measurements that isolate each emission type should improve the power and specificity of OAEs as non-invasive probes of cochlear function.

Thesis Supervisor: Christopher A. Shera
Title: Associate Professor of Otology and Laryngology
Harvard Medical School
Acknowledgments

I wish to thank the many people who contributed to my academic and personal development. Above all, I thank my advisor Chris Shera. Without his patience, guidance, and “what can I do to help?” attitude, graduate school would have been too stressful, directionless, and ultimately unproductive. I am grateful to him for helping me define questions, refine ideas, and in general to think in a less chaotic way. Chris wrestled with the data-acquisition system, leaving me free to focus on other things, a gift that is most appreciated. Chris showed me by example that there is great value and satisfaction in being persistent and passionate about whatever task one undertakes, be it science or otherwise.

I thank the members of my thesis committee: John Guinan for taking interest in my research, critically reading my manuscripts, and for contributing to my overall development as a fledgling scientist, Sharon Kujawa for encouraging me to think about the clinical relevance of my work and for suggesting possible future pathways, and Prof. Chiang Mei for reminding me to think more deeply about cochlear nonlinearities and to make connections with non-auditory wave phenomena.

My thanks to: Bill Peake, who, perhaps without knowing, convinced me that it was okay follow in my brother’s footsteps, which eventually lead me to apply to the Speech and Hearing Program. Nelson Kiang for always providing interesting conversations about the ins and outs of scientific life.

My thanks to the engineering group at EPL: Mike, Ish, and Chris Scarpino, for providing me with the starship “jewel”ry and accessories that allowed me to complete the last part of my experiments, and Frank for often lending me a hand when labview was hard to handle. My thanks to Dianna Sands, without whom I’d still be wandering around hallways, looking for someone to let me in through locked doors. I thank my subjects for lending me their ears.

Thanks go to my dear friends: Courtney, Irina, Jocelyn, Ryuji, Tony, and Waty for unfailingly answering my afternoon calls to tea, for lingering over that tea and for talking about research and all things not research. Brad, Chandran, Dave and Domenica, for always encouraging words. Chris Bergevin, Prakash, and Annette for providing helpful comments on improving my presentations.

Thanks to my “bestest” friend Ray; for assuring me that growing up is simply an affliction of a dim imagination; and for showing me that the affliction was thankfully avoided by marrying him. Special thanks must go to my big brothers, Srinath and Sridhar, who are always ahead of me, but never leave me behind. Finally, I thank my parents for encouraging me to lead a balanced life, for supporting my desire to play, skip, dance, sing, and be merry, and for embracing me through all things bad and good.
As a parting gift to future SHBT students, I suggest the following cure for the afternoon blues:

Conceited brew;
tepid and bitter,
were it not for the
company it drew.
Sweetened by a cookie,
danish, and fritter,
hmmm, maybe two.
Soothing tonic for the
bruised ego of
the graduate student crew
Anyone? Tea for me
and tea for you?

Radha Kalluri
May 15, 2006

This work was supported by grants RO1DC003687 and T32DC00038 from the NIDCD, National Institutes of Health.
# Contents

1 Introduction 8
   1.1 Background ............................................. 9
      1.1.1 Forward traveling waves ............................ 9
      1.1.2 Reverse traveling waves ............................ 10
      1.1.3 Integrated two-mechanism framework .............. 11
   1.2 Overview ............................................. 14
      1.2.1 Near equivalence between SFOAEs measured by three methods 14
      1.2.2 Near equivalence between SFOAEs and TEOAEs .......... 15
      1.2.3 Distortion-product source unmixing ............... 15

2 Near equivalence between SFOAEs measured by suppression, compression, and spectral smoothing 17
   2.1 Introduction ........................................... 17
      2.1.1 Three SFOAE Measurement Techniques ............... 18
      2.1.2 Overview ........................................... 21
   2.2 Methods ............................................. 22
      2.2.1 Stimulus delivery and acquisition ................ 22
      2.2.2 Three-interval paradigm ........................... 22
      2.2.3 Measurement details ............................... 24
      2.2.4 Two-stage calibration ............................. 29
   2.3 Results .............................................. 35
   2.4 Discussion .......................................... 37
      2.4.1 Region of validity ............................... 41
      2.4.2 Differences between suppression and compression .... 41
      2.4.3 Advantages and disadvantages of SFOAE methods ..... 42
      2.4.4 Conclusion ....................................... 43

3 Equivalence between human click-evoked and stimulus-frequency otoacoustic emissions in humans 45
   3.1 Introduction .......................................... 45
   3.2 Measurement Methods ................................... 47
      3.2.1 Measuring CEOAEs ................................. 47
      3.2.2 Measuring SFOAEs ................................ 51
   3.3 Experimental Complications ............................. 52
      3.3.1 CEOAE transfer functions from linear and nonlinear methods 52
3.3.2 Synchronized spontaneous otoacoustic emissions .......................... 54
3.4 Comparison of SFOAE and CEOAE Transfer Functions ....................... 56
3.4.1 Equating stimulus intensities ........................................... 59
3.4.2 Comparison of the growth functions ..................................... 59
3.4.3 Comparison at equivalent intensities .................................... 61
3.5 Discussion .............................................................................. 61
3.5.1 Possible limits of application .............................................. 61
3.5.2 Quasilinearity of the transfer functions .................................. 65
3.5.3 Consistency with the DP-place component of DPOAEs ................. 67
3.5.4 Implications for emission mechanisms .................................... 68

4 Distortion-Product Source Unmixing .............................................. 70
4.1 Introduction .............................................................................. 70
4.2 Unmixing Via Selective Suppression ........................................... 75
4.2.1 Measurement methods ...................................................... 76
4.2.2 Detailed Measurement Methods .......................................... 78
4.2.3 Results: Unmixing via selective suppression ......................... 81
4.3 Unmixing Via Spectral Smoothing ............................................. 86
4.3.1 Analysis methods ............................................................. 87
4.3.2 Results: Unmixing via spectral smoothing ............................. 90
4.4 Errors due to Incomplete Unmixing .......................................... 90
4.4.1 Estimating δ and ρ .......................................................... 95
4.5 Discussion .............................................................................. 97
4.5.1 Source mechanism versus source location ............................... 98
4.5.2 Comparison with other work ............................................... 99
4.5.3 Region of validity of the two-mechanism model ...................... 100
4.5.4 Methods of unmixing ....................................................... 101
4.5.5 Implications of unmixing DPOAEs ..................................... 102

5 Summary and conclusions ............................................................. 104
5.1 Chapter Overview and Summaries .............................................. 104
5.2 Implications and Future Directions .......................................... 107
5.3 Conclusion .............................................................................. 109
Chapter 1

Introduction

Healthy cochleae generate sounds. These otoacoustic emissions (OAEs)\(^1\) can be evoked and measured in the ear-canal using small, sensitive, low-noise microphones (Kemp, 1978). Since the cochlea is embedded in the temporal bone, this ability to evoke and non-invasively measure signals that originate within it provides remarkable access to an otherwise inaccessible organ.

Until recently, all OAEs were commonly viewed as originating from a single non-linear mechanism associated with the mechanics of the cochlea (e.g., Kemp, 1978, 1997, 1998; Probst et al., 1991; Allen and Neely, 1992; Allen and Lonsbury-Martin, 1993; Patuzzi, 1996). Because it was believed that all OAEs shared this common origin, different OAE “types” were typically named and distinguished by the stimulus used to evoke them (e.g., Zurek, 1985; Probst et al., 1991); namely SFOAEs (stimulus frequency otoacoustic emissions evoked by tones of the same frequency as the stimulus), TEOAE (emissions evoked by transient stimuli), DPOAEs (distortion product otoacoustic emissions that occur at frequencies not contained in the stimulus) and SOAEs (spontaneous otoacoustic emissions present in the absence of external stimulation). It is now widely accepted that this one-mechanism view of OAE generation is incorrect.

Shera and Guinan (1999) proposed an alternate two-mechanism model in which OAEs originate by at least two mechanisms, namely: 1) by linear reflection from pre-existing mechanical perturbations that are arrayed across the length of the cochlea, and 2) by distortions that are induced by the stimulus acting on a nonlinear system. If, as suggested by the model, OAEs originate via two different mechanisms, it is reasonable to imagine that these mechanisms might be affected by cochlear pathologies in different ways. Separating and independently studying emissions generated by these two different mechanisms in normal hearing subjects should improve our understanding of OAEs and lead to improved interpretability of OAE measurements.

**Link to the cochlea** OAEs exhibit characteristics that strongly suggest their link to cochlear mechanics. OAEs are vulnerable to cochlear injury in a frequency specific

---

\(^1\)oto in reference to their origin in the ear, acoustic because they are sounds, and emissions because they are emitted or transmitted into the environment.
manner (e.g., Kim, 1980; Schmiedt and Adams, 1981). OAEs exhibit nonlinear characteristics that are very similar to the nonlinear response properties of the cochlear: For example, OAEs grow nonlinearly with level, OAEs at one frequency are suppressible by a tone at another frequency, and OAEs measured in response to complex stimuli contain spectral components not contained in the stimulus (for a review of OAE features see Probst et al. (1991)). In the next section we provide a brief background that provides the link between these commonly observed features of OAEs and cochlear mechanics.

1.1 Background

Auditory signals are generally described as going inward from the external ear-canal to the cochlea. Otoacoustic emissions travel in the reverse direction, out from the cochlea to the external ear canal. The two-mechanism model proposes two cochlear mechanisms by which the energy of the stimulus is reversed to be “re-emitted” as an otoacoustic emission. Before we describe the proposed model we begin with a brief description of the forward pathway. Readers interested in a more detailed study of cochlear mechanics theory are referred to de Boer (1980, 1984, 1991); Zweig (1991).

1.1.1 Forward traveling waves

The cochlea can be viewed as a coiled, fluid filled, rigid tube with nonuniform cross section that is lengthwise partitioned into two chambers by a flexible membrane (cochlear partition). Two covered windows (one in each chamber) look back into the middle-ear cavity. The input to the cochlea is provided by the piston-like stapes (the last in the chain of middle ear bones) pushing into the upper chamber of the cochlea. This motion of the stapes displaces fluid into the upper chamber, which is then vented by the window in the lower chamber. The resulting pressure difference across the two cochlear chambers creates a mechanical traveling wave that propagates down the length of the cochlear partition.

The nonuniform mechanical properties of the partition (e.g., its mass and stiffness, which vary along the length) spatially separate different frequency components of the stimulus so that high-frequency displacements are favored at locations near the stapes and low-frequency displacements at locations away from the stapes. As the vibration excited by a single tone propagates down the cochlea its amplitude envelope first increases, reaches a peak, and then quickly falls (Békésy, 1960). The location of the peak (known as the best-place, BP) is approximately an exponential function of the stimulus frequency (e.g., Békésy, 1960; Greenwood, 1990; Liberman, 1982).
Passive versus active mechanics  In healthy cochleae, specialized cells (outer hair cells) provide mechanical gain to the partition, thereby amplifying and sharpening the traveling wave envelope (Corey and Hudspeth, 1979). In contrast to the responses in cadaver cochleae (Békésy, 1960), which are rather broad due to the damping caused by fluid viscosity, responses in a healthy animal cochlea are tall and sharp (e.g., Rhode, 1970; Sellick et al., 1982). Because outer hair cells can provide only limited amplification, the resulting input/output gain function saturates at moderate stimulus intensities (e.g., 40 dB SPL), thereby creating a compressive nonlinearity. As a consequence of this compressive nonlinearity the peak of the traveling wave grows less than linearly with stimulus intensity (Johnstone et al., 1986). \(^2\)

Scaling symmetry  Although the cochlea’s response patterns appear to depend in a complex way on both frequency and space, measurements reveal a local scaling invariance. The cochlea’s response pattern is effectively a function of one dimensionless variable, \(f/f_{cf}(x)\), where \(f\) is the frequency of excitation and \(f_{cf}(x)\) defines the cochlear place versus characteristic frequency map. As a result of scaling-invariance and the exponential form of the map, the vibration pattern for a stimulus at one frequency is simply a shifted version of the pattern at another frequency. \(^3\)

1.1.2 Reverse traveling waves

The two-mechanism model suggests two fundamentally different ways in which the traveling-wave energy can be reversed. The following sketches the reasoning; for details refer to Shera and Guinan (1999).

Pre-existing perturbations  Reflection-source emissions (Zweig and Shera, 1995; Talmadge et al., 1998) are generated via reflection off pre-existing “random” perturbations in the mechanics of the cochlea. A possible source of these perturbations comes, for example, from variations in the number of outer hairs, their orientation, and stiffness. The phase of each scattered wavelet depends on the phase of the forward-traveling wave at the location of scattering. In Fig. 1-1, at a site of reflection, the stimulus accumulates more phase as the frequency is increased. Since the micromechanical impedance perturbations are fixed in space, the phase of the incident wave at each perturbation changes as the frequency of the forward-traveling wave is

---

\(^2\)All mechanical systems if driven at high enough intensities will exhibit nonlinearities. Here the nonlinearities are manifest at relatively low levels because the linearity limit of the hair cells is lower than the limit of the passive portion of the partition.

\(^3\)We emphasize that the scaling invariance is a local feature and does not hold for widely separated frequencies. In addition scaling symmetry may not hold at very low frequencies.
varied. The reflection of a higher frequency wave from that site will have correspondingly larger phase than the reflection of a lower frequency wave. Consequently, OAEs generated by linear reflection from fixed perturbations manifest a phase that rotates rapidly with frequency.

**Distortion-induced perturbations** In contrast, distortion-source emissions are induced by the stimulus. As illustrated in Fig. 1-2, as the frequency of the stimulus is swept, this interaction moves with the waves (unlike pre-existing perturbations, which are fixed in space). Therefore the phase of the induced distortion source remains constant. Although for simplicity the figure illustrates the process for a single frequency, in the scaling symmetric cochlea, when the stimuli are frequency scaled, the combined stimulus waveforms merely shift such that the relative phase interaction within the different components of the stimulus remain constant.

1.1.3 **Integrated two-mechanism framework**

The two-mechanism model has the advantage that it resolves many previously unexplainable inconsistencies arising from the “one-mechanism” view. The two-mechanism view does not simply rename emissions, it restructures the way we understand emissions to be generated. The variability in OAE amplitudes across different species, for example, can be understood as a difference in the way the two mechanisms work in each species. In rodents, DPOAEs are robust, whereas SFOAEs are small. On the other hand, in primates, SFOAEs are large but DPOAEs are small (Zurek, 1985, e.g.). If the same mechanism was responsible for both, as suggested by the conventional view, one would expect emissions to have the same relative strength. Another instance where the two-mechanism model helps explain data is the following. SFOAEs and SOAEs are diminished by aspirin whereas DPOAEs are unaffected. In the context of the two-mechanism model, one might explain this effect as aspirin acting preferentially on one type of emission generator without effecting the other.

Although the two-mechanism model suggests that OAEs can be broadly understood as originating by either linear reflection or nonlinear distortion, traditional OAEs types (e.g., TEOAEs, SFOAEs, DPOAEs) might not be easily categorized into one or another group. Instead, evoked emissions measured in the ear canal are likely to be “mixtures” of emissions generated by both mechanisms. The different emission types probably contribute in different degrees dependent on species, stimulus parameters, and cochlear state.

---

4We clarify that induced sources, whether induced by distortions or otherwise, are predicted to have similar phase characteristics. In this case, distortions are one mechanism by which reverse traveling waves can be induced by the stimulus
Linear Reflection

wave shifts

Mechanical Perturbation fixed in space

phase changes at source

Cochlear Location

Figure 1-1: Schematic diagram illustrating that reflection sources from pre-existing perturbations are predicted to have a phase that varies rapidly with frequency. The figure shows a snapshot of a traveling wave at two different frequencies (top) and the corresponding phase lags (bottom) versus cochlear location. The two frequencies are denoted by black and gray lines, respectively. Although the theory indicates the emission arises from a coherent sum from densely packed perturbations that are randomly arrayed across the entire cochlea, for simplicity we consider the reflection from one perturbation. Since the perturbation (*) is fixed in space the phase of the wave scattered by the perturbation changes considerably (1) as the stimulus frequency is varied. Consequently, the phases of OAEs generated by linear reflection vary rapidly with frequency. Figure adapted from Kalluri and Shera (2001a)
Nonlinear Distortion

Induced source moves with wave

wave shifts

Figure 1-2: Analogous to Fig. 1-1, this figure illustrates that distortion sources are predicted to have a nearly frequency invariant phase. Distortion sources result from nonlinear distortion induced by the wave. The sources illustrated here (×) are idealized as points at the peak of the traveling wave. When the frequency is varied, the traveling wave (and thus the resulting distortion source) simply shifts along the cochlear partition, maintaining a nearly constant relative phase relationship as the stimulus frequencies are varied. Note that as the frequency of the stimulus is varied, the distortion source remains fixed to the peak as the wave pattern shifts (→) along the partition. As a result, the phases of all resulting distortion products are essentially independent of frequency. Figure adapted from Kalluri and Shera (2001a)
Our work in this thesis focuses on understanding how these two proposed “building-blocks” of OAE generation combine and interact to form common OAE types in humans.

1.2 Overview

This thesis is organized into five chapters. This first chapter provides the introduction, necessary background, and chapter abstracts. Chapters 2, 3, and 4 are the main body chapters (written as self-contained papers in collaboration with Christopher Shera) and are each effectively dedicated to one of the three OAE types. The chapters are ordered based on the complexity of the stimulus and the predicted relationship between emission types, rather than the order in which the work was actually completed. Chapter 2 (the most recently completed work) deals with SFOAEs, which are evoked by a single tone, the simplest of our stimuli. Chapter 3 (completed second) tests the relationship between SFOAEs and TEOAEs. Chapter 4 (completed first) deals with DPOAEs. In combination, the three chapters provide systematic tests of each branch of the two-mechanism model of OAEs. Chapter 5, the final chapter, includes a summary of our findings from each of the three body chapters, our conclusions, and recommendations for possible extensions of this work.

1.2.1 Near equivalence between SFOAEs measured by three methods

Stimulus frequency otoacoustic emissions (SFOAEs) have been measured in several different ways. These include: 1) suppression, 2) nonlinear-compression, and 3) spectral smoothing (or time domain windowing). Each of these methods exploits a different cochlear phenomenon or signal-processing technique to extract the emission. It is generally assumed that these techniques yield equivalent results, but this equivalence has never been verified. This paper reports, compares and contrasts SFOAEs measured in humans using each of these techniques. In the suppression technique the emission is computed as the complex difference between the ear canal pressure when a probe tone is presented alone and in the presence of a second higher intensity tone at a nearby frequency. The higher level tone is thought to suppress the emission. The compression technique makes use of the compressive growth of emission amplitude relative to the linear growth of the stimulus. The emission is computed by taking the complex difference between two measurements of ear-canal pressure; 1) a measurement made at a low stimulus intensity and 2) a scaled version of a measurement made at a high stimulus intensity. The spectral smoothing technique involves convolving the ear canal pressure with a smoothing function. The analysis is
equivalent to applying a time domain window the Fourier transform of the ear canal spectrum. We report that the emissions extracted by all three techniques were nearly indistinguishable; both in amplitude and in phase.

1.2.2 Near equivalence between SFOAEs and TEOAEs

Although linear reflection models of OAE generation predict that both click-evoked and stimulus-frequency otoacoustic emissions originate via essentially linear mechanisms, other models imply that differences in the spectrum of the evoking stimulus result in differences in the mechanisms of OAE generation. The work reported here was motivated by these basic disagreements about the influence of stimulus spectrum on the mechanisms of OAE generation.

The experiments reported here test for bandwidth dependent differences in mechanisms of otoacoustic emission generation. Click-evoked and stimulus-frequency OAE input/output transfer (or describing) functions were obtained as a function stimulus frequency at various intensities. We find that at low and moderate intensities human CEOAE and SFOAE input/output transfer functions are nearly identical. When stimulus intensity is measured in component-equivalent sound pressure level, CEOAE and SFOAE transfer functions have equivalent growth functions at fixed frequency and equivalent spectral characteristics at fixed intensity. This near-equivalence suggests that the OAEs evoked by broad and narrow band stimuli (CEOAEs and SFOAEs) are generated by the same mechanism. The spectral characteristics of CEOAE and SFOAE transfer functions are consistent with the predictions of coherent-reflection models of OAE generation. We conclude that although CEOAEs and SFOAEs are conveniently given different names based on the characteristics of the stimuli used to evoke them, the two OAEs “types” are better understood as members of the same emission family; namely, linear-reflection.

1.2.3 Distortion-product source unmixing

This paper tests key predictions of the “two-mechanism model” for the generation of distortion-product otoacoustic emissions (DPOAEs). The two-mechanism model asserts that lower-side band DPOAEs constitute a mixture of emissions arising not simply from two distinct cochlear locations (as is now well established) but, more importantly, by two fundamentally different mechanisms: nonlinear distortion induced by the traveling wave and linear coherent reflection off pre-existing micromechanical impedance perturbations. The model predicts that (1) DPOAEs evoked by frequency-scaled stimuli (e.g., at fixed-$f_2/f_1$) can be unmixed into putative distortion- and reflection-source components with the frequency dependence of their phases consistent with the presumed mechanisms of generation; (2) The putative reflection-source
component of the total DPOAE closely matches the reflection-source emission (e.g., low level stimulus-frequency emission) measured at the same frequency under similar conditions. These predictions were tested by unmixing DPOAEs into components using two completely different methods: (a) selective suppression of the putative reflection source using a third tone near the distortion-product frequency and (b) spectral smoothing (or, equivalently, time-domain windowing). Although the two methods unmix in very different ways, they yield similar DPOAE components. The properties of the two DPOAE components are consistent with the predictions of the two-mechanism model.
Chapter 2

Near equivalence between SFOAEs measured by suppression, compression, and spectral smoothing

2.1 Introduction

To deal with the problem of separating two signals that overlap in time and in frequency, stimulus frequency otoacoustic emissions have typically been measured using methods based on (1) compression (e.g., Kemp and Chum, 1980), (2) suppression (e.g., Guinan, 1990), and (3) spectral smoothing (also known as time-domain windowing, e.g., Shera and Zweig 1993a). Each of the three SFOAE measurement methods exploits a different cochlear phenomenon to extract the emission; the suppression technique makes use of cochlear two-tone suppression, the compression technique makes use of the OAE’s compressive growth, and the spectral smoothing technique makes use of the difference in delay between the OAE and its stimulus. Although it is generally assumed that these various techniques for measuring SFOAEs all yield equivalent results, this equivalence has never been established.

Nonlinear techniques like suppression and compression are commonly used because standard techniques used to measure other kinds of OAEs cannot be easily adapted to measure SFOAEs. For example, distortion-product otoacoustic emissions (DPOAEs) are extracted by Fourier analysis in the frequency domain because the stimulus and the emission do not overlap in frequency. Transient-evoked otoacoustic emissions (TEOAEs) are extracted by windowing in the time domain because the stimulus and
emission do not overlap in time. SFOAEs and the tone used to evoke them overlap both in time and in frequency; making them difficult to extract by either Fourier analysis or by windowing in the time domain.

Since the stimulus and emission are not distinguishable either in frequency or in time, what is the evidence that there is an emission? The presence of the emission is evident from the emergence of a quasi-periodic spectral fine structure in ear-canal pressure as the probing frequency is varied (e.g., Zwicker and Schloth, 1984; Shera and Zweig, 1993a). This fine structure is understood as arising from the interference between the stimulus and the emission. When SFOAEs are absent this fine structure is also absent (e.g. when measurements are made in a simple cavity). In this view, when a normally functioning ear is excited by a continuous tone, the ear-canal pressure \( P_p \) at the probing frequency \( f_p \) contains two components: the pressure of the stimulus, \( P_{ST}(f_p) \), and the pressure of the stimulus-frequency otoacoustic emission, \( P_{SF}(f_p) \):

\[
P_p(f_p) = P_{SF}(f_p) + P_{ST}(f_p).
\]

As the frequency of excitation is varied, the relative phase between the stimulus and emission changes. As a result the maxima of the interference pattern occur when stimulus and emission add in phase, the minima when they add out of phase (Fig. 2-1).

Several features of this fine structure are similar other OAEs and attach the origin of these interfering components (i.e. the putative SFOAE) to the cochlea. First, the fine structure amplitude grows compressively with increasing stimulus level, suggesting that the nonlinearities of the cochlea shape the SFOAEs. Figure 2-1 illustrates the intensity dependence of this fine structure in our data. While the quasi-periodic oscillations are prominent at low stimulus levels, they are almost imperceptible at higher stimulus levels. Second, the addition of a second tone suppresses the response at the probe frequency (much like cochlear two-tone suppression) (e.g., Guinan, 1990). Third, the frequency spacing between fine-structure peaks suggests that the putative SFOAE has long frequency delays (Zwicker and Schloth, 1984; Zweig and Shera, 1995; Talmadge et al., 1998). These long and frequency dependent delays further link the SFOAE to the cochlea whose mechanical response is known to be frequency dispersive (e.g., Kemp, 1978; Wilson, 1980; Norton and Neely, 1987; Neely et al., 1988).

### 2.1.1 Three SFOAE Measurement Techniques

Although nonlinear features of the cochlea shape the SFOAE, the acoustic waveforms of the stimulus and emission superpose in the ear canal (expressed as Eq. 2.1).

\(^1\)Ability to temporally separate is dependent on frequency content and stimulus type.
Quasi-periodic fine structure in the total ear-canal pressure is level dependent. Spectral oscillations that result from the interference between the stimulus and emission are more prominent at low stimulus levels (20 dB SPL). As stimulus levels increase the magnitude of the oscillations decrease.

Superposition suggests that the problem of measuring SFOAEs is equivalent to the problem of accurately estimating the pressure of the stimulus $P_{ST}$. In other words if we can develop a good way to estimate the pressure of the stimulus, then we can easily derive an estimate for the emission. In the following sections we briefly describe how common features of the SFOAE (as described above) can be used to generate three different estimates of the stimulus pressure $P_{ST}$. These three estimates of $P_{ST}$ can then be used with eq. (2.1) to compute the pressure of the SFOAE $P_{SF}$:

$$P_{SF}^{c,s,w} = P_{p} - P_{ST}^{c,s,w}, \quad (2.2)$$

where the superscripts $c,s,w$ represent $P_{SF}$ and $P_{ST}$ estimated by compression, suppression and spectral-smoothing (time-domain windowing).

**Compression**

The compression technique (e.g., Kemp and Chum, 1980) exploits the nonlinear compressive growth of the emission together with the linear growth of the stimulus. Because the emission grows compressively while the stimulus grows linearly, the ratio of the SFOAE amplitude to the stimulus amplitude is larger at low-stimulus levels than at high-stimulus levels (if both grew linearly this ratio would be independent of stimulus intensity). We can estimate the pressure of the stimulus at low-levels $P_{ST}^{c}$ by appropriately scaling the ear-canal pressure measured at high levels. If the high-level measurement is made at sufficiently high stimulus levels the contribution from emissions is small after scaling. The SFOAE can then be defined as the complex difference between the probe-frequency pressure measured at one probe intensity and a linearly scaled-down version of the pressure remeasured at a higher probe intensity.
Suppression

The suppression technique makes use of cochlear two-tone suppression to separate the emission from the total ear-canal pressure. The SFOAE is defined as the complex difference between the probe-frequency ear-canal pressure measured first using the probe tone alone $P_p(f_p)$ and then remeasured in the presence of an additional "suppressor" tone at a nearby frequency $P_{ps}(f_p; f_s)$ (e.g., Guinan, 1990; Kemp et al., 1990; Brass and Kemp, 1991, 1993; Shera and Guinan, 1999). The suppressor tone is assumed to substantially reduce or eliminate the emission evoked by the probe. The SFOAE can then be defined as the complex difference between the probe-alone pressure and the probe-plus-suppressor pressure.

Spectral Smoothing

Spectral smoothing (also known as time-domain windowing) makes use of the difference in time delay between the stimulus and emission to separate the two signals. As the frequency of the probe tone is increased, the phase of the stimulus and the phase of the emission have different dependencies on frequency (i.e. they have different phase-vs-frequency slope). According to linear-systems theory phase slope in the frequency domain corresponds to latency in the time domain. Therefore, by applying Fourier analysis to our frequency-domain measurements of ear-canal pressure, we expect to see two components with different latencies; namely a short-latency component corresponding to the pressure of the stimulus and a long-latency component corresponding to the pressure of the SFOAE. These two components of the total-ear-canal pressure should be separable by appropriate windowing in the "latency domain." ² Widowing in the latency domain corresponds to convolution with a "smoothing function", $S$, in the frequency domain. (The notion that the stimulus pressure can be estimated by smoothing is intuitively consistent with the interpretation that fine structure is created by an interference between the stimulus and emission.)

Although this method is based on linear-systems theory, this procedure does not make priori assumptions about the mechanism via which the emissions are generated; rather it is a signal processing technique that make use of particular feature of the signal to separate it into two components. Techniques for analyzing OAEs in this way were introduced by Shera and Zweig (1993a; Zweig and Shera, 1995), who applied them to the study of SFOAEs; similar methods have since been applied to other emissions (e.g., L.J.Stover et al., 1996; Brown et al., 1996; Fahey and Allen, 1997; Knight and Kemp, 2000a; Ren et al., 2000).

²We put "latency-domain" in quotes because the signal we obtain by Fourier transforming the frequency response does not correspond with the time-domain impulse response of the system.
2.1.2 Overview

Understanding differences or equivalence between the emissions extracted by different measurement techniques should improve our understanding of the phenomena that underly each technique and the governing emission generating mechanisms. In addition to impacting the development of future SFOAE measurement systems, this study should clarify the interpretation of existing SFOAE studies. Recent disagreements about the origin of SFOAE sources (Siegel et al., 2003), the unknown influence of suppressor tones (Shera et al., 2004), and the possible fundamental difference between SFOAEs and other types of OAEs (Patuzzi, 1996) further motivate the need for this study.

At low stimulus intensities, SFOAEs appear to arise by linear mechanisms. Several features of SFOAEs support this interpretation. For example at low levels SFOAE amplitudes scale with stimulus intensity (e.g., Zwicker and Schloth, 1984; Zweig and Shera, 1995). And, in agreement with the predictions of linear-reflection models, the phase of SFOAEs vary rapidly with frequency, and the SFOAE spectrum has sharp spectral notches (e.g., Shera and Guinan, 1999; Talmadge et al., 1998). Although these features of SFOAEs support the idea that they originate from linear mechanisms, this interpretation is weakened by the use of nonlinear measurement methods. The weakness of the argument is based on a simple notion; can one really use a nonlinear phenomenon to make a measurement and still interpret the results within the context of a linear model?

SFOAEs have sometimes been characterized as an “error of extrapolation” (Patuzzi, 1996) and a distinction is made between SFOAEs and other emission types. In contrast to the interpretation that DPOAEs, TEOAEs, and SOAEs are true signals evoked by the ear, SFOAEs are regarded as an artifact of nonlinear residual techniques. For example, a similar error of extrapolation would result if we deliver our stimuli with nonlinear sources and then use the compression technique to measure SFOAEs in a simple cavity. By comparing SFOAEs measured by spectral smoothing (an entirely linear technique that does not depend on system nonlinearities) and SFOAEs measured by compression, we can determine whether SFOAEs are also artifacts of the nonlinear-residual technique.

Modeling studies (e.g., Shera et al. 2004) suggest that suppressor tones can both suppress and induce SFOAE sources. The degree to which suppressors induce sources depends on the frequency of the suppressor; the sources are predicted to be greatest for suppressor frequencies that are much higher than the probe and small for near-probe suppressors. The possibility that the suppressor adds components to the measured SFOAE clouds the interpretation of SFOAEs.

To address these questions, in this paper we report, compare, and contrast the SFOAEs measured in humans using each of three different techniques.
2.2 Methods

2.2.1 Stimulus delivery and acquisition

Signals were delivered and recorded in the ear-canal in one ear of four \( (n = 4) \) normal-hearing human subjects who were comfortably seated in a sound-isolated chamber. All procedures were approved by the human studies committee at the Massachusetts Eye and Ear Infirmary and at the Massachusetts Institute of Technology. Stimulus waveforms were generated and responses acquired and averaged digitally using a custom-built data-acquisition system. The system consists of a National Instruments 4461 data-acquisition board, two Shure-E2c earphones, an ER-10c preamplifier, a Knowles EK3103 microphone, and a custom-built sound-delivery apparatus. The Shure-E2c insert earphones were chosen because, unlike ER10c earphones, their response at the fundamental frequency grows linearly at levels below approximately 75 dB SPL. Additionally, their compact design made them easy to adapt for insertion into the ear-canal.

Stimuli were digitally generated using a fixed sampling rate of 48 kHz and fft-lengths of 4096, resulting in a frequency resolution of approximately 11 Hz.

Potential artifacts in a data buffer were detected by comparing the rms-difference between a previously stored artifact-free reference buffer and the current data buffer against a criterion. If the difference was more than the criterion, then the data buffer was rejected, otherwise the data buffer was added to the averaging buffer. Continual replacement of the reference buffer minimized the effects of slowly varying drifts in the baseline signal. To reduce potential low frequency noise in the stimulus delivery, we applied a high pass filter with a cutoff at 150 Hz before delivering signals to the ear canal. To reduce spurious rejection of buffers due to low-frequency noise, we applied a bandpass filter with cutoff frequencies at 150 Hz and 15 kHz.

2.2.2 Three-interval paradigm

We presented our stimuli in three equal-length interleaved segments (Fig. 2-2). Because three different stimulus configurations can be presented in close succession, interleaving can minimize the effects of time-dependent drifts. Similar interleaving techniques were previously used to measure SFOAEs via suppression (e.g., Guinan, 1990; Kalluri and Shera, 2001a).

The three-interval paradigm is implemented as follows: 1) In the first segment we present a single tone (the probe) at frequency \( f_p \) and level \( L_p \). We refer to this first segment as the probe-alone segment. During this segment the ear-canal pressure at the probe frequency contains both the stimulus and the emission. 2) During the second segment we simultaneously present the probe tone and a suppressor tone at
Figure 2-2: Schematic diagram of the three-interval interleaved paradigm. The emission is computed as the Fourier component at the probe frequency by taking the complex difference between the probe-alone segment and: 1) the probe-plus-suppressor segment, 2) a smoothed version of the probe-alone segment, or 3) compressor segment. The probe-alone, probe-plus-suppressor, and compressor waveforms are extracted using rectangular windows $w_p$, $w_{ps}$, $w_c$ (only $w_p$ is shown in the figure). The window duration, $T_w$, contains an integral number of cycles of probe, suppressor, and compressor.

A slightly lower frequency ($f_s = f_p - 50$) and higher stimulus level (typically $L_s$ was between 55 and 60 dB SPL). We refer to this second segment as the probe-plus-suppressor segment. Because the simultaneous presentation of the probe and suppressor is assumed to substantially reduce or eliminate the emission, the ear-canal pressure measured during this segment presumably contains just the stimulus. To reduce sound-source distortions we present the probe and suppressor tones using two different sound sources. 3) Finally, in the third segment we present a single high level tone (the compressor) at the same frequency as the probe tone and at the same level as the suppressor tone ($L_{ph} = L_s$). Because the OAE grows compressively, the pressure of the low-level stimulus can be estimated by scaling the pressure measured during this third segment.

The probe-alone waveform, $p_p(t)$, the probe-suppressor waveform, $p_{ps}(t)$, and the high-level probe or compressor waveform, $p_{ph}(t)$ are extracted from the measured ear-canal pressure, $p(t)$, by applying windows $w_p(t)$, $w_{ps}(t)$ and $w_c(t)$ to the ear-canal pressure waveform:

$$p_p(t) = \frac{1}{2} [w_p(t) \cdot p(t) + w_p(t - T_w) \cdot p(t)] \quad (2.3)$$

$$p_{ps}(t) = \frac{1}{2} [w_{ps}(t) \cdot p(t) + w_{ps}(t - T_w) \cdot p(t)] \quad (2.4)$$

$$p_{ph}(t) = \frac{1}{2} [w_c(t) \cdot p(t) + w_c(t - T_w)) \cdot p(t)] \quad (2.5)$$
The windows are rectangular boxcar windows \(^3\),

\[ w_p(t) = w_{\text{box}}(t - T_o; T_w) \quad (2.6) \]

\[ w_{ps}(t) = w_{\text{box}}(t - T_o - T_1; T_w) \quad (2.7) \]

\[ w_c(t) = w_{\text{box}}(t - T_o - T_2; T_w). \quad (2.8) \]

The window offsets \( T_o \) and \( T_1 \) were chosen to allow the system time to return to steady state after ramping the suppressor and compressor tones on or off. The window duration \( T_w \) contains an integral number of cycles of both probe and suppressor. For the measurements reported here \( \{T_o, T_1, T_2, T\} = \{\frac{1}{4}, \frac{9}{4}, \frac{9}{2}, 7\} T_w \text{ ms} \), where \( T_w = \frac{N}{F_s} \).

From these three waveforms, \( p_p(t), p_{ps}(t), \) and \( p_{ph}(t) \), we derive three estimates of the stimulus pressure at the probe frequency:

\[ P_{ST}^S(f_p) = \mathcal{F}\{p_{ps}(t)\}e^{i2\pi f_p T_1} \quad \text{(by suppression)}, \quad (2.9) \]

\[ P_{ST}^C(f_p) = 10^{-(L_{ph} - L_p)/20} \mathcal{F}\{p_{ph}(t)\}e^{i2\pi f_p T_2} \quad \text{(by compression)}, \quad (2.10) \]

\[ P_{ST}^w(f_p) = \mathcal{F}\{p_p(t)\}e^{i2\pi f_p T_o} * S \quad \text{(by smoothing)}, \quad (2.11) \]

where \( * \) indicates convolution by the smoothing filter \( S \), and \( \mathcal{F}\{\cdot\} \) indicates the \( N = 4096 \) point discrete Fourier transform.

These three estimates of the stimulus pressure together with Eq. 2.2 can be used to generate three estimates of the SFOAE.

Probe levels were varied from 20–50 dB SPL. The suppressor frequency was fixed at approximately 50 Hz below the probe frequency. Suppressor levels were typically fixed between 55 and 60 dB SPL. To enable comparisons between suppression-derived and compression-derived SFOAEs, we always chose suppressor levels that were equal to the compressor level \( (L_s = L_p) \).

### 2.2.3 Measurement details

**Compression**

The compression technique relies on our ability to generate stimuli that scale linearly with stimulus intensity. In a nonlinear measurement system a scaled-down version of high-level ear-canal pressure would not accurately represent the pressure of the stimulus at low levels. To minimize sound-source nonlinearities we designed a sound-delivery system that included Shure E2c sound sources, whose response at the fundamental frequency grows linearly for stimulus levels below approximately 75 dB.

\(^3\)\( w_{\text{box}} = 1 \) for \( 0 \leq t \leq T_w \) and \( w_{\text{box}} = 0 \), otherwise
SFOAE methods

SPL. In addition, we calibrated our earphones using moderate level chirps to reduce nonlinearities introduced by driving the sound sources with large broadband chirps. We appraised the linearity of our measurement system by measuring $P_{SF}$ in an acoustic cavity with dimensions comparable to the ear canal space; if the emission level was below the noise floor at a particular stimulus configuration then we considered the system to be sufficiently linear. In all the measurements we report, the system distortion was in the noise.

**Suppression**

Because maximal suppression in humans occurs for suppressors with frequencies near the probe (e.g., Brass and Kemp, 1993, 1991; Backus, 2005), we fixed the suppressor frequency at 50 Hz below the probe. Although high-side suppressors (suppressors with frequencies above the probe) are thought to generate somewhat greater suppression, we chose low-side suppressors to be consistent with our previous measurements (Kalluri and Shera, 2001). However, since our suppressor is only 50 Hz away from the probe, the differences between high-side- and low-side- suppressors (especially at relatively high suppressor levels) are probably small. At suppressor levels of approximately 60 dB SPL further decreasing the frequency difference between the suppressor and probe did not result in discernable differences in OAE level.  

We chose suppressor levels that were likely to remove all or most of the SFOAE. SFOAE-suppression input/output curves (such as the example shown in Figure 2-3 at two different frequencies) reach an approximate plateau by 60 dB SPL. The plateau region indicates levels at which further increasing the suppressor level does not extract more of the SFOAE. Although 55-to-60 dB SPL suppressors extract most of the emission when the probe was at 30 dB SPL, they did not remove all of the SFOAE at 50 dB SPL. In Fig. 2-3, at each probe level, the effective level corresponding to a 60 dB SPL suppressor is marked by a $\forall$. As illustrated effectiveness of a 60 dB SPL suppressor is dependent on the probe level. In retrospect, we could have used more intense suppressors with the higher probe levels to ensure maximal suppression, but we were concerned that suppressors greater than 60 dB SPL would evoke large efferent effects.

**Spectral smoothing**

Fourier transforms were performed with respect to a log-frequency coordinate (Zweig and Shera, 1995; Kalluri and Shera, 2001; Knight and Kemp, 2001). Fourier trans-

---

$^4$Note we only tested the affect of suppression level and frequency on two subjects but our results are generally consistent with finding of Backus (2005) who measured the frequency dependence of suppression in several subjects.
Figure 2-3: SFOAE suppression and compression input/output curves. This figure plots the normalized SFOAE amplitude (normalized by the probe level to form an SFOAE “transfer function”) as a function of the effective level of the suppressor (open symbols) and compressor (filled symbols). The “effective stimulus level” of the compression stimulus is $10 \log\left( \frac{A_{p}}{A_{ph}} \right)$ and $10 \log\left( \frac{A_{s}}{A_{p}} + \frac{A_{ph}}{A_{p}} \right)$ for suppression. Here $A_p, A_{ph}$ and $A_s$ are the pressures of the probe, compressor and suppressor tones respectively. The three triangles in the left hand panel indicate the effective level corresponding to $L_s, L_{ph} = 60$ dB SPL. At low effective levels the SFOAE amplitude is systematically greater for suppression than for compression. At high effective levels the result from both techniques asymptote to the same value. Data are in Subject #2.
Figure 2-4: The figure shows the latency-domain representations of $P_p$ and $P_{ps}$ at probe levels of 50 dB SPL. The Fourier transforms $F\{P_p\}$ (squares) and $F\{P_{ps}\}$ (circles) are shown as a function of $\tau$, the emission latency expressed in stimulus periods. Filled vs. unfilled symbols distinguish two different measurements. The $\tau$-domain representation of both $P_p$ show a strong peak of about the same height centered at a latency of about 10 cycles. The addition of a 55 dB SPL (filled symbols) suppressor tone reduces this peak to approximately 35% of its original height, the addition of a 60 dB SPL tone (open symbols) reduces the height to about 10% of its original height.

formation with respect to a modified frequency coordinate results in sharper, more well-defined peaks in the latency domain; resulting in a cleaner extraction by windowing.\textsuperscript{5} The resulting Fourier-conjugate variable, $\tau$ represents the emission latency in periods of the stimulus frequency.\textsuperscript{6}

Figure 2-4 shows the probe-alone in the latency domain and the probe+suppressor in the latency domain. The peak belongs to the SFOAE (observe that the addition of a suppressor tone reduces this peak). Consistent with Fig. 2-3 the degree of peak reduction depends on the suppressor level; as a result the 60 dB SPL suppressor almost completely eliminates the peak, a 55 dB SPL suppressor does not.

Convolution in the frequency domain is equivalent to windowing latency domain.

\textsuperscript{5}Measurements of tone-burst-evoked OAE and ABR latency (Neely \textit{et al.}, 1988), as well as measurements of SFOAE group delay (Shera and Guinan, 2003), all indicate a gradual breaking of scaling symmetry in the basal turns of the mammalian cochlea. For near optimal compensation for traveling-wave dispersion, the measurements suggest working with the variable $-\sqrt{f/f_{ref}}$ (see also Shera and Guinan, 2003).

\textsuperscript{6}The smoothing technique as applied here is similar to previously used techniques to unmix the two-components of distortion-product otoacoustic emissions (DPOAE) (Kalluri and Shera, 2001a).
The pressure of the stimulus in the latency domain is given by:

\[ \mathcal{F}\{P_{ST}^m\} = \mathcal{F}\{P_p\} \cdot \hat{S} \quad \text{where,} \quad \hat{S} \equiv \mathcal{F}\{S\}. \]

where here \( \mathcal{F} \) represents the Fourier transform with respect to the log-frequency coordinate and is different from the discrete time Fourier transform used earlier. To separate emission from the stimulus while avoiding excessive “ringing” in the frequency response, we use a recursive-exponential window for \( \hat{S} \) (Shera and Zweig, 1993a).\(^7\)

In practice, measurements are only available over a finite frequency range, and the smoothing operation is complicated by end effects.\(^8\) Throughout this paper, the analyzed frequency range was chosen to include an approximately integral number of spectral cycles and smoothing was performed using periodic boundary conditions (the data were effectively wrapped around a cylinder). When necessary, a linear ramp was subtracted, and subsequently restored after smoothing, to remove any discontinuity at the “seam.” The need to remove linear ramps was generally minimized by our use of a two-stage calibration procedure (described below) which reduced time dependent drifts in the baseline signal.

In these measurements, the duration of the time window (bandwidth of the smoothing function) \( \tau_{\text{cut}} \) was 5 stimulus periods, which was sufficient to separate the stimulus from the emission at all stimulus levels used.\(^9\)

---

\(^7\)The \( n \)-th order recursive-exponential filtering window is defined by (Shera and Zweig, 1993a):

\[ \hat{S}_n(\tau; \tau_{\text{cut}}) \equiv \frac{1}{\Gamma_n(\lambda_n \tau/\tau_{\text{cut}})} \]

where the parameter \( \tau_{\text{cut}} \) is the cutoff time (length of the window) and the function \( \Gamma_n(\tau) \) is defined recursively:

\[ \Gamma_{n+1}(\tau) = e^{\Gamma_n(\tau) - 1}, \quad \text{with} \quad \Gamma_1(\tau) = e^{\tau^2}. \]

The window \( \hat{S}_n(\tau; \tau_{\text{cut}}) \) has a maximum value of 1 at \( \tau = 0 \); the scale factor \( \lambda_n \) is chosen so that the window falls to the value \( 1/e \) at \( \tau = \tau_{\text{cut}} \):

\[ \lambda_n = \sqrt{\gamma_n}, \quad \text{where} \quad \gamma_{n+1} = \ln(\gamma_n + 1) \quad \text{with} \quad \gamma_1 = 1. \]

Note that the 1st-order filtering window is a simple Gaussian; in the limit \( n \to \infty \), \( \hat{S}_n \) approaches a rectangular (or boxcar) window. For intermediate \( n \) (e.g., the value \( n = 10 \) used here), \( \hat{S}_n \) has a much sharper cutoff than standard windowing functions (e.g., the Hamming, Blackman, etc.) and considerably less “ringing” in the smoothing function than the simple boxcar.

\(^8\)The frequency range of the end effects are typically on the order of width of the smoothing function, whose approximate width is given by (Shera and Zweig, 1993a)

\[ \Delta \nu = \Delta f/f \geq 1/\pi \tau_{\text{cut}}. \]

\(^9\)At lower stimulus intensities we could have used larger values for \( \tau_{\text{cut}} \) because the latency of the
2.2.4 Two-stage calibration

Generally we calibrate with broad-band chirps before each measurement series. Calibrations are used to guarantee that the stimulus tones at all frequencies have constant level and start at sine phase in the ear canal. When using this standard calibration technique (we will refer to this as 1-stage calibration) we find that our data contains spurious spectral structure. This structure probably results from a combination of factors including; time dependent variations due to calibration drifts, shifts in the probe fit, and as shown below, contamination of the calibration by otoacoustic emissions.

To prevent spurious spectral structure from contaminating our implementation of spectral smoothing we adopted a two-stage calibration procedure. In the new procedure, we first determine an approximate driving voltage by calibrating with a broad-band chirp. Then, preceding each measurement point, we perform a second calibration by simultaneously presenting high level tones at the probe and suppressor frequencies (at 60 dB SPL unless otherwise noted). This second calibration is used to make adjustments to the driving voltage. Because the tone calibrations occur immediately preceding each measurement frequency, errors that might result from time dependent drifts are minimized.

Figure 2-5 illustrates the benefits of this two-stage calibration procedure at two different probe levels (50 dB SPL in the top row, and 20 dB SPL in the bottom row). The two panels on the left show the ear-canal pressure measured after single-stage calibration and the two panels on the right show the ear-canal measured after two-stage calibration. Each point is the pressure measured at one frequency. $P_p(f_p)$ are drawn with filled symbols and $P_{ps}(f_p)$ are drawn with open symbols. Note that the nominal stimulus pressure was subtracted from each curve to approximately center each curve about the origin.

According to Eq. 1, when drawn in the complex domain, $P_p$ should rotate around $P_{ps}$. Since we subtracted the nominal pressure of the stimulus from both $P_p$ and $P_{ps}$, we expect $P_{ps}$ to be centered around the origin, and $P_p$ to rotate around the origin. $P_p$ measured after one-stage calibration at 50 dB SPL does not rotate around the origin and instead forms a complex pattern. During suppression, we expect to see the circular pattern to grow smaller and cluster about the origin; again these data do not conform to the expectation. Because spectral-smoothing attributes all emission peak seems in general to be greater for the lower stimulus intensities. Although the latency domain data suggests a level dependent trend consistent with the idea that latencies become shorter with increasing stimulus intensity, we resist making these conclusions based on our limited data set. The comparison of emission latency over stimulus intensity would be more compelling if we analyzed a larger frequency range which would improve the resolution of the latency domain peaks.
spectral structure to SFOAEs, naively smoothing signals that might contain unknown sources of spectral structure can generate large errors in the estimated SFOAE. The filled circles in Fig. 2-6 show the SFOAE estimated by spectral smoothing at 50 dB SPL. For comparison the open diamonds show the emission extracted by suppression. These two estimates are very different.

In contrast, after two-stage calibration, $P_p$ at both 20 and 50 dB SPL form a neat spiral whose amplitude is reduced after suppression so that $P_{ps}$ clusters about the origin. The open circles in Fig. 2-6 show the SFOAE estimated by spectral smoothing at 50 dB SPL. Now the SFOAE estimated by suppression and by smoothing are nearly indistinguishable.

Choosing the calibration level  Simply using two-stage calibration does not ensure that spectral smoothing will work properly. It is also important to calibrate at the appropriate level. Figure 2-7 illustrates the effects of calibration level on the estimates of SFOAE that are computed by spectral smoothing. Again these data were collected using the two-interval paradigm with two-stage calibration. While the probe and suppressor levels were fixed at $L_p = 50$ dB SPL and $L_s = 60$ dB SPL, the levels of the calibration tones $L_{clb}$ were varied; in the top row $L_{clb}$ equals 40 dB SPL, in the middle row $L_{clb}$ equals 50 dB SPL, and in the bottom row $L_{clb}$ equals 60 dB SPL. Although spectral smoothing uses just the probe-alone segment, for comparison we show both the probe-alone and probe+suppressor curves. The left-most panels in all rows show the real and imaginary parts of the ear-canal pressure, both in the absence of suppression (black circles) and during suppression (open circles).

At different calibration levels we found key differences in the form of the amplitude fine-structure and in how suppression affects that fine structure. In the top row (when $L_{clb} = 40$ dB SPL) the oscillation amplitude of the fine structure is relatively large, adding suppression further enhances the amplitude of the fine structure. In the middle row (when $L_{clb} = 50$ dB SPL), the amplitude of the fine-structure is small, suppression slightly enhances the fine structure, and more creates a phase difference between the two spectral waveforms. In the last row, the amplitude of the fine structure is moderate and suppression reduces this amplitude. Although the spectral fine structure of the probe-alone and probe-plus-suppressor pressures is dependent on calibration level, the complex difference between the probe-alone and probe-plus-suppressor pressure (i.e. the SFOAE estimated by suppression) remains approximately the same.

Though calibration does not significantly affect the estimates of $P_{SF}^a$, the estimates of $P_{SF}^m$ are strongly dependent on calibration level (Fig. 2-7). Since the technique of spectral smoothing blindly assigns all spectral structure to the estimate of the SFOAE, when spectral structure differs, estimates of SFOAE differ. If $P_{SF}^m$ and $P_{SF}^a$
Figure 2-5: Two-stage calibration reduces spurious spectral structure. The two left panels show complex-domain plots of ear-canal pressure at the probe-frequency measured after single-stage calibration. The two right panels show the ear-canal pressure after two-stage calibration. In all four panels the nominal stimulus pressure was subtracted so that the data are approximately centered around the origin. In the bottom and top panels the probe levels $L_p$ are 20 and 50 dB SPL, respectively and the suppressor levels were fixed at $L_s = 60$ dB SPL. $L_{clb} = 60$ dB SPL tones were used for calibration. Closed symbols are used to draw the ear-canal pressure without suppression. Open symbols are used to draw the ear-canal pressure with suppression. Data from Subject #1
at $L_{c1b} = 40$ dB SPL are directly compared (Fig. 2-7, first row, right panel), $P_{SF}^w$ has additional spectral structure, is larger at some frequencies, and has a less than one cycle difference in phase. At $L_{c1b} = 50$ dB SPL, although the differences in phase are indistinguishable, $P_{SF}^w$ is much smaller than $P_{SF}^s$. Finally at $L_{c1b} = 60$, $P_{SF}^w$ and $P_{SF}^s$ are comparable.

The most obvious source of variation with calibration level comes from the SFOAE itself. In the following we use a simple model to explain the effect we see in the middle panel of (Fig. 2-7).\(^1\) First we begin with a description of that data and the procedure used to collect it: The second stage of the two-stage calibration was performed by simultaneously presenting two tones, both at 50 dB SPL, and one at the probe and the other at the suppressor frequencies. Following calibration, the actual measurement is made with a suppressor tone at a higher level of 55 dB SPL.

When the ear-phones are driven by some nominal voltage to produce approximately 50 dB SPL, the ear-canal pressure contains both the stimulus and the stimulus frequency otoacoustic emission, $P_{ec} = P_{ST} + P_{SF}$. However, during calibration we simultaneously present two 50 dB SPL tones, one at the suppressor frequency and the at the probe frequency. The simultaneous presentation of the second tone partially suppresses the emission, so that the probe-frequency component of the ear-canal pressure is now $P_{ST} + \alpha P_{SF}$, here $\alpha$ is the unsuppressed fraction of the $P_{SF}$. As described earlier, the $P_{SF}$ or as in this case some fraction of it, $\alpha P_{SF}$, creates spectral fine structure. Our calibration procedure adjusts the driving voltage to flatten this structure. We call this flattened pressure $P_{flat} = P_{ec} - \alpha P_{SF}$.

During the measurement, we first present the probe-alone; the lack of the suppressor tone suppresses the emission.

\(^1\)Note that this explanation can be generalized to explain the results in the other panels. This generalization is not incorporated into this document.
Figure 2-7: Low-level calibrations can affect the results of spectral smoothing. The column on the left shows the real and imaginary parts of the probe-frequency component of the ear-canal pressure during the probe-only (filled black symbols) and the probe-plus-suppressor (open gray symbols) segments. While $L_p = 50$ dB SPL and $L_s = 55$ dB SPL, $L_{clb}$ the level of calibration tone in our two-stage calibration procedure was varied; $L_{clb} = 40$ (top row, squares), $L_{clb} = 50$ (middle row, circles) and $L_{clb} = 60$ (bottom row, diamonds) dB SPL. In each of the curves in the left column the nominal stimulus pressure was subtracted. The column on the right shows the magnitude and phase of SFOAEs that are extracted by suppression (open gray symbols) and spectral smoothing (filled black symbols). Data from Subject #2
pressor restores the previously suppressed portion of the emission,

\[ P_p = P_{flat} + (1 - \alpha)P_{SF}. \]  \hspace{1cm} (2.13)

Next, we present a suppressor tone at a higher level than used during calibration, (55 dB SPL). This higher level tone further suppresses the remaining portion of the emission \( \alpha P_{SF} \). Although for simplicity we assume that the 55 dB SPL suppressor completely removes the remaining portion of the emission, to be more accurate we should really scale that by some factor \( \beta < 1 \) which indicates the fraction of the emission that is suppressed. In this example, we consider that \( \beta = 1 \), or in other words the 55 dB SPL suppressor was intense enough to suppress the rest of the emission. Then we can express the suppressed ear-canal pressure by:

\[ P_{ps} = P_{flat} - \alpha P_{SF}. \]  \hspace{1cm} (2.14)

The \( P_{ps} \) and \( P_p \) segments both contain some fraction of the emission and therefore have similar quasi-periodic spectral structure. The opposing signs in front of \( \alpha P_{SF} \) and \( (1 - \alpha)P_{SF} \) accounts for the phase difference between the two curves.

Applying spectral smoothing to \( P_p \) extracts only a small fraction of the emission, \( (1 - \alpha)P_{SF} \). However, calibration doesn't affect the emission extracted by suppression; notice that the difference between \( P_p - P_{ps} \) equals \( P_{SF} \).

Since contamination by OAEs will be smallest at high calibration levels (because the OAE grows compressively), using relatively high levels during the second calibration minimizes contamination.

**Summary of measurement paradigms**  In two of four subjects we initially collected data using the one-stage calibration procedure with the three-interval paradigm (3I-1S). After realizing that two-stage calibration benefits spectral smoothing, we recollected the data in these subjects with the new calibration procedure. Since two-stage calibration was implemented solely for improving spectral smoothing, we reduced our data collection time by removing the compressor stage (i.e., we reduced the paradigm to a two-interval paradigm with two-stage calibration 2I-2S). We retained the probe+suppressor interval to facilitate cross-paradigm comparisons. Thus in subject #1 and subject #2 we have four estimates of \( P_{SF} \); two estimates by suppression \( P_{SF}^s \) (one via 3I-1S and the second via 2I-2S), one estimate by compression \( P_{SF}^c \) via 3I-1S, and one by spectral smoothing \( P_{SF}^s \) via 2I-2S.

The optimized paradigm (three-interval two-stage calibration) was used to generated all three estimates of \( P_{SF} \) in subjects #3 and #4.
Figure 2-8: SFOAEs by suppression, compression, and spectral-smoothing. This figure shows magnitude (top), phase (middle) and detrended phase (bottom) of $P_{SF}$ measured by compression, suppression or spectral smoothing in one subject. The different symbols correspond to the different methods used to derive the estimates. The bottom panel was computed from the data shown in the middle panel by removing a smooth trend line that captures the large variations in phase (the same trend line was removed from all the curves). The noise floor in this measurement was approximately $-25$ dB SPL.

2.3 Results

Figure 2-8 shows $P_{SF}$ measured in one subject (subject #1) using suppression, compression, and spectral smoothing. In the top two panels, the four curves in each correspond to the magnitude and phase of: 1) the SFOAE measured by compression (squares), 2) the SFOAE measured by suppression in the three interval paradigm (open diamonds), 3) the SFOAE measured by suppression in the two interval paradigm (filled diamonds), and 4) the SFOAE measured by spectral smoothing (circles). Because the large OAE delay and phase unwrapping hides the details of phase variation, the bottom panel shows the phase after subtracting a smooth trend line that captures the broad variations of phase. The detrended phase curves were limited between $\pm 0.5$
cycles. In these measurements the noise floor was typically between $-20$ and $-25$ dB SPL.

The differences between the SFOAE measured in the two-interval (2I) paradigm (open symbols) and three-interval (3I) paradigm (filled symbols) are small and on the order of $1$ dB (at least at frequencies outside the spectral notches). Since the two are measured on different days, these differences are an example of the day-to-day variability of the data (variability that likely originates from small variations in the probe’s fit from day-to-day). The near-equivalence between emissions measured by 3I and 2I paradigms shows that the addition of the third interval does not significantly affect the derived SFOAE. The differences between the SFOAE measured by 1) suppression and compression, and 2) suppression and spectral smoothing were smaller than the day-to-day differences reported above. In this example case, we illustrated that the four estimates of the SFOAE pressure, $P_{SF}^{c,s,n}$ at $L_p = 30$ dB SPL, are nearly identical in magnitude, phase, and in the fine details of the phase.

Figure 2-9 extends our comparison of the three estimates of the SFOAE pressure ($P_{SF}^{c,s,n}$) to four different stimulus intensities in two subjects. The top four rows are the estimates of $P_{SF}$ at probe levels $L_p = 20, 30, 40, \text{ and } 50$ dB SPL respectively. The bottom row shows the detrended phase for $L_p = 30$ dB SPL. Figure 2-10 shows SFOAEs measured in the same two subjects in a higher frequency region. In both subjects and in both frequency ranges, the differences between the SFOAE measured by suppression, compression, and spectral smoothing are small at $L_p = 20$ and 30 dB SPL. We do note a difference in the fine details of the phase in this subjects data (bottom row, left panel), however since the largest deviations occur at locations that correspond to the large spectral notches, these differences probably arise because the phase of the emission is not well characterized in the notch.

Although the differences between the three estimates are small at low probe levels, there are differences at higher probe levels. Take note of the difference between the three estimates at 50 dB SPL in subject #2 (right panel, 4th row, Fig. 2-9 and Fig. 2-10). In both these examples, the amplitude of the emission estimated by suppression (filled diamonds) is systematically greater than the emission estimated by compression. In discussion we attempt to understand these differences in terms of a simple model. In addition, in this same subject estimates of $P_{SF}^{w}$ also deviate at higher levels. We believe this difference comes from our use of a 55 dB SPL suppressor tone. As shown in Fig. 2-4, a 55 dB SPL tone does not completely suppress the emission. To properly match compare suppression and spectral smoothing, we should compare with data that are measured with higher suppressor levels.

In subjects #3 and #4 (Fig. 2-11) the data were collected in the 3-interval paradigm after 2-stage calibration. The format in this figure is the same as in Fig. 2-9 where each column shows the data for one subject, the top four rows are the estimates of $P_{SF}$ at different probe levels and the bottom row shows the detrended $P_{SF}$ phase for
the 30 dB SPL probe level. In this figure each panel has three estimates of $P_{SF}$; $P_{SF}^c$ by compression, $P_{SF}^s$ by suppression, and $P_{SF}^w$ by spectral-smoothing. Because subject #4 was particularly noisy we were unable to collect data over the entire frequency range. Frequency regions below 1.5 kHz were particularly difficult because in this region this subject’s emissions were small. However, despite the noisy measurements, our finding of approximate equivalence between the three SFOAE estimates at low levels in other subjects also holds in this subject. Notice, for example, the ability of all three techniques to capture the sharp amplitude peaks at 20 dB SPL, possibly indicating frequencies of spontaneous emissions. Again, as we found in previous subjects, all three techniques capture the same fine variations in phase. Because the amplitude of the spectral structure is relatively small at high stimulus intensities (e.g., Fig. 2-1), its was difficult to cleanly estimate $P_{SF}^w$ at 50 dB SPL. Because the spectral structure is relatively small, minor calibration errors can skew the spectral structure. Therefore there are larger differences between $P_{SF}^w$ and $P_{SF}^c$ or $P_{SF}^s$ at higher intensities.

2.4 Discussion

In this paper we directly compared three techniques for measuring stimulus-frequency otoacoustic emissions (SFOAEs) in the same human subjects. At low probe levels the emissions extracted by all three techniques were nearly indistinguishable; both in amplitude and in phase. However this good match between the three SFOAE estimates deviates at higher probe levels, where we found systematic differences between the SFOAE estimated by suppression and the SFOAE estimated by compression. Specifically we find that at the same effective level, the SFOAEs estimated by suppression have a larger magnitude. Although suppression and compression generate differences in the absolute magnitude of the emission, the form of the emissions – i.e. the details of the spectral shape, the locations of spectral notches and the frequency dependence of the phase – remain similar. This similarity in form suggests that although there are small differences between the techniques, the underlying emission generating mechanism are not different.

In some subjects (e.g., subject #4 in Fig. 2-1) spectral smoothing fails to match the SFOAEs estimated by suppression and compression. The greatest deviations occur at higher probe intensities and in subjects whose emissions levels are relatively small (compare subject #1 to subject #4). Since spectral-smoothing attributes all spectral variations as originating from the emission, small deviations arising from measurement artifacts can have large effects on the extracted emission. This is particularly true at high probe levels where the quasi-periodic spectral oscillations are small (see for example the 50 dB SPL curve in Fig. 2-1). Whereas small irregularities would not greatly distort the quasiperiodic structure seen at the low probe levels, they
Figure 2-9: SFOAEs by suppression, compression, and spectral-smoothing at different probe levels. This figure shows $P_{SF}$ measured in two different subjects using suppression (open and filled diamonds), compression (filled squares) and spectral smoothing (open circles). The symbols drawn with closed symbols were measured in the 3-interval paradigm and the symbols drawn with open symbols were measured in the 2-interval paradigm (see text for details). The top four rows in each column show the magnitude of the SFOAE spectra at probe levels of 20, 30, 40 and 50 dB SPL. Because the large OAE delay and phase unwrapping hides the details of phase variation, the bottom panel shows the phase of the 30 dB SPL SFOAE after subtracting a smooth trend line that captures the broad variations of phase.
Figure 2-10: SFOAEs by suppression, compression, and spectral-smoothing at a higher frequency region. The description of this figure is similar to Fig. 2-9; it shows $P_{SF}$ measured in two different subjects using suppression (open and filled diamonds), compression (filled squares) and spectral smoothing (open circles). The symbols drawn with closed symbols were measured in the 3-interval paradigm and the symbols drawn with open symbols were measured in the 2-interval paradigm (see text for details). The top four rows in each column show the magnitude of the SFOAE spectra at probe levels of 20, 30, 40 and 50 dB SPL. Because the large OAE delay and phase unwrapping hides the details of phase variation, the bottom panel shows the phase of the 30 dB SPL SFOAE after subtracting a smooth trend line that captures the broad variations of phase.
Figure 2-11: SFOAEs by suppression, compression, and spectral-smoothing at different probe levels. Similar to Figs. 2-9, 2-10, this figure shows $P_{50}$ measured in two more subjects using suppression (open diamonds), compression (squares) and spectral smoothing (circles). Unlike the previous two figures, here we made all measurements using the optimized 3-interval paradigm. The top four rows in each column show the magnitude of the SFOAE spectra at probe levels of 20, 30, 40 and 50 dB SPL. Because the large OAE delay and phase unwrapping hides the details of phase variation, the bottom panel shows the phase of the 30 dB SPL SFOAE after subtracting a smooth trend line that captures the broad variations of phase.
would greatly distort the spectral structure at the higher probe levels. As a result the estimates of SFOAEs by spectral smoothing do not always match the SFOAEs estimated by suppression and compression.

2.4.1 Region of validity

We did not systematically explore the effects of suppressor frequency on the SFOAE, therefore we cannot rule out the possibility that very-high frequency suppressors add components to the SFOAE. However, our finding that near-probe suppressors, compression, and spectral smoothing generate similar results suggests that near probe-suppressors are successful at suppressing almost the entire emission, and are not adding components to the emission.

As our measurements were limited to frequencies between 1 and 4 kHz, understanding whether this equivalence between SFOAE measurement methods extends to lower frequencies remains an important open question. Differences in the mechanics of the low-frequency cochlear apex and the high frequency base may translate to differences in the effectiveness of the three SFOAE measurement techniques. For example direct measurements of basilar membrane motion show that the peak of the traveling wave is not as compressive at low frequencies. Reduced compression in the growth of basilar membrane responses presumably reduces the effectiveness of the compression technique for measuring SFOAEs. Two-tone suppression in auditory nerve responses show that suppression is greater at high frequencies than at lower frequencies (Delgutte, 1990; Javel et al., 1978). Presumably these differences in suppression translate to differences in the effectiveness of suppression at low frequencies.

The use of very high intensity suppressors or compressor levels has the added complication of possibly exciting middle-ear muscle and/or olivo-cochlear efferent effects. Shairer et. al. reported an increased noise at the probe frequency which they attributed to an unknown biological source; possibly created by temporal variations resulting from efferent effects. We cannot exclude that these variations may also be present in our data. However since the measurements of the three techniques were essentially made at the same time, we assume that if these effects are present then they are equally present in all three of our estimates.

2.4.2 Differences between suppression and compression

Our measurements were limited to probe levels below 50 dB SPL. Although suppression and compression yield nearly equivalent results for probe levels below 40 dB SPL, there are systematic differences between the emissions extracted by suppression and compression at 50 dB SPL probe levels; as in the data of subject #2 in Fig. 2-10.
This difference comes from our decision to fix the suppressor level at 60 dB SPL (in one case at 55 dB SPL).

Figure 2-3 shows the normalized emission magnitude (i.e., the “transfer” function amplitude), as a function of “effective” stimulus level. The effective-stimulus level is defined as 

\[10 \log \left( \frac{A_p^2 + A_s^2}{A_p^2} \right)\]

for suppression, and

\[10 \log \left( \frac{A_{ph}^2}{A_p^2} \right)\]

for compression; \(A_p\), \(A_s\), and \(A_{ph}\) are the amplitudes of the probe, suppressor, and high level probe, respectively. Each curve is parameterized by the level of the probe \(L_p\). The results illustrate that a 60 dB SPL suppressor tone is not sufficient to maximally extract the emission generated by a 50 dB SPL probe. These data show that below a threshold effective level suppression always extracts a greater portion of the emission than compression. Above threshold the two techniques asymptote to the same value. Extrapolating the 50 dB SPL input/output curve suggests that to reach a plateau we would have had to use suppressor/compressor levels of 80 dB SPL.

Kanis and de Boer (1993) view compression as an extension of suppression, or “self-suppression”. If we view suppression and compression in the same spirit, we can suggest a possible, although “hand wavey”, explanation for why suppression extracts a greater portion of the emission than compression. The explanation is based first on the assumption that our suppressors and compressors are sufficiently close in frequency and effectively act on the same nonlinearity. The total output (by total we mean the output including all distortion products) of the nonlinearity is the same regardless of whether the input is at one frequency or distributed at more than one frequency. During self-suppression or compression most of the output of the nonlinearity is at the frequency of the probe frequency, during suppression the output is distributed between the probe frequency and the distortion product frequencies. Since the total output is distributed between the probe and distortion product frequencies, the amplitude of the probe component during suppression is smaller than during compression. Therefore, in this situation one can imagine that the probe component during suppression grows more compressively than during self-suppression.

### 2.4.3 Advantages and disadvantages of SFOAE methods

Each of the three measurement techniques used in this study have clear advantages and disadvantages. The disadvantage of the compression technique was the need for a highly linear sound source. In the absence of a linear sound source, the compression technique would generate artifactual SFOAEs. The greatest benefit of both the suppression and compression techniques is the ability to make measurements at isolated frequencies. In contrast, spectral smoothing requires measurement of ear-canal pressure at several closely spaced frequency points that cover a large interval (at least several cycles of the quasi-periodic spectral structure). Although spectral smoothing
requires a large number of frequency measurements, it does not require the additional probe+suppressor or probe+compressor segments. In that regard, although spectral smoothing cannot provide measurements at isolated frequencies, it would provide a faster measurement of SFOAE over a broad frequency range.

Unlike the spectral smoothing which requires careful calibration to minimize irregularities arising from calibration errors, contamination by emissions, and measurement artifacts, compression and suppression were considerably more insensitive.

2.4.4 Conclusion

Our finding of near equivalence between the SFOAE measured by suppression, compression, and spectral smoothing provides an assurance that those features that are now commonly regarded as characteristic of SFOAEs are not simply an artifact of measurement technique but rather represent a “true” otoacoustic emission. Based on our finding that both linear and nonlinear techniques for measuring SFOAEs generate the same results, we conclude that the use of nonlinear measurement techniques does not alter previous interpretations of SFOAEs as originating via linear-reflection mechanisms. Finally, our previous findings that SFOAEs and CEOAEs are nearly identical at low and moderate stimulus levels (Kalluri and Shera, 2004) indicates that measuring CEOAEs provides a fourth independent way to measure SFOAEs.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFOAE</td>
<td>stimulus-frequency otoacoustic emission</td>
</tr>
<tr>
<td>$A_s$</td>
<td>amplitude of suppressor</td>
</tr>
<tr>
<td>$A_p$</td>
<td>amplitude of probe</td>
</tr>
<tr>
<td>$A_{ph}$</td>
<td>amplitude of high-level/compressor probe</td>
</tr>
<tr>
<td>$L_s$</td>
<td>level of the suppressor in dB SPL</td>
</tr>
<tr>
<td>$L_p$</td>
<td>level of the probe tone in dB SPL</td>
</tr>
<tr>
<td>$L_{ph}$</td>
<td>level of the high-level/compressor tone in dB SPL</td>
</tr>
<tr>
<td>$L_{clb}$</td>
<td>level of the calibration tones in dB SPL</td>
</tr>
<tr>
<td>$f_p$</td>
<td>probe/compressor frequency</td>
</tr>
<tr>
<td>$f_s$</td>
<td>suppressor frequency</td>
</tr>
<tr>
<td>$f_{dp}$</td>
<td>distortion-product frequency, $2f_s - f_p$</td>
</tr>
<tr>
<td>$p_{ps}$</td>
<td>probe+suppressor segment time waveform</td>
</tr>
<tr>
<td>$p_p$</td>
<td>probe-only time waveform</td>
</tr>
<tr>
<td>$p_{ph}$</td>
<td>high-level probe time waveform</td>
</tr>
<tr>
<td>$w_{rex}$</td>
<td>recursive exponential window</td>
</tr>
<tr>
<td>$w_{box}$</td>
<td>rectangular window</td>
</tr>
<tr>
<td>$w_p$</td>
<td>window to extract probe-only segment</td>
</tr>
<tr>
<td>$w_{ps}$</td>
<td>window to extract probe+suppressor segment</td>
</tr>
<tr>
<td>$w_c$</td>
<td>window to extract the high-level/compressor segment</td>
</tr>
<tr>
<td>$P_{ec}$</td>
<td>complex ear-canal pressure</td>
</tr>
<tr>
<td>$P_{ST}$</td>
<td>complex pressure of the stimulus</td>
</tr>
<tr>
<td>$P_{SF}$</td>
<td>complex pressure of the SFOAE</td>
</tr>
<tr>
<td>$P_p$</td>
<td>complex pressure of the probe-only segment at $f_p$</td>
</tr>
<tr>
<td>$P_{ps}$</td>
<td>complex pressure of the probe+suppressor segment at $f_p$</td>
</tr>
<tr>
<td>$P_{ph}$</td>
<td>complex pressure of the high-level/compressor probe at $f_p$</td>
</tr>
</tbody>
</table>
Chapter 3

Equivalence between human click-evoked and stimulus-frequency otoacoustic emissions in humans

3.1 Introduction

The stimuli used to evoke otoacoustic emissions (OAEs) range from the spectrally dense (e.g., the broadband clicks used to elicit click-evoked OAEs) to the spectrally sparse (e.g., the single pure tones used to evoke stimulus-frequency OAEs). Although linear reflection models of OAE generation (e.g., Zweig and Shera, 1995; Talmadge et al., 1998) predict that both click-evoked and stimulus-frequency otoacoustic emissions (CEOAEs and SFOAEs) originate via essentially linear mechanisms (i.e., wave reflection off pre-existing mechanical perturbations), other models imply that differences in the spectrum of the evoking stimulus result in differences in the mechanisms of OAE generation. Nobili et al. (2003a), for example, use model simulations to argue that the mechanisms responsible for CEOAE generation are both inherently nonlinear and fundamentally different from those responsible for generating SFOAEs. In Nobili et al.’s simulations, CEOAEs result from spatially complex “residual oscillations” of the basilar membrane that trace their origin to spectral irregularities in middle-ear transmission (see also Nobili, 2000; Nobili et al., 2003b). Based on OAE measurements in guinea pig, Yates and Withnell (1999b) also posit a distinction between OAEs evoked by narrow- and broadband stimuli. They argue that although SFOAEs may originate from the independent “emission channels” predicted by linear reflection models, CEOAEs are essentially broadband distortion-product emissions (broadband...
DPOAEs). In this view, CEOAEs arise not from independent channels but from intermodulation distortion sources induced as a consequence of nonlinear interactions among the multiple frequency components of the broadband click stimulus (see also Withnell and Yates, 1998; Yates and Withnell, 1999a; Carvalho et al., 2003).

The work reported here was motivated by these basic disagreements about the influence of stimulus spectrum on mechanisms of OAE generation. Our goal was to determine the relationship between the OAEs evoked by stimuli with the most dissimilar temporal and spectral structure (i.e., CEOAEs and SFOAEs). Interpretation of the experiments assumes that differences in OAE spectral characteristics imply differences in OAE generating mechanisms, and conversely. Similar logic has been used to distinguish “reflection-” and “distortion-source” OAEs. Whereas reflection-source OAEs (e.g., SFOAEs at low levels) have a rapidly varying phase and a slowly-varying amplitude occasionally punctuated by sharp notches, distortion-source OAEs (e.g., DPOAEs evoked at fixed, near-optimal primary-frequency ratios) have an almost constant amplitude and phase. These differences in OAE spectral characteristics are taken as indicative of fundamental differences in their mechanisms of generation (e.g., Kemp and Brown, 1983; Shera and Guinan, 1999).

Despite its fundamental importance, only a handful of studies have addressed the comparison between CEOAEs and SFOAEs. Although Zwicker and Schloth (1984) measured tone- and click-evoked frequency responses in the same human subject, the uncertain reliability of the tone-evoked data makes compelling comparisons difficult. Unlike the tone-evoked responses observed in subsequent studies, the ‘synchronous-evoked’ OAEs reported by Zwicker and Schloth appear inconsistent with an origin in a causal system (Shera and Zweig, 1993a). Furthermore, the emission data for the two stimulus types are presented in different ways: although the CEOAE data represent the emission alone, the tone-evoked data represent the combined pressure of the stimulus and the emission. In what appears to be the only other study to explicitly address the issue, Prieve et al. (1996) found that the OAEs evoked by clicks and by tone bursts have similar intensity dependence, consistent with a common mechanism of generation. Unfortunately, the bandwidths of their tone-bursts were not all that narrow (they typically spanned an octave or more), and their data do not allow a comparison of spectral structure or phase.

The experiments reported here examine the effect of stimulus bandwidth on OAE generation mechanisms by measuring and appropriately comparing the emissions evoked by wide-band clicks (CEOAEs) with those evoked by tones (SFOAEs). Comparisons are made across stimulus frequency and intensity in the same human subjects.
3.2 Measurement Methods

Measurements were made in one ear of each of four \((n = 4)\) normal-hearing human subjects who were comfortably seated in a sound-isolated chamber. All procedures were approved by human studies committees at the Massachusetts Eye and Ear Infirmary and the Massachusetts Institute of Technology.

Stimuli were digitally generated and recorded using implementations of standard OAE measurement protocols on the Mimosa Acoustics measurement system. The measurement system consists of a DSP-board (CAC Bullet) installed in a laptop computer, an Etymotic Research ER10c probe system, and two software programs—one for measuring CEOAEs (T2001 v3.1.3) and another for measuring SFOAEs (SF2003 v2.1.18).

Signals were delivered and recorded in the ear canal. In-the-ear calibrations were made before each measurement. Stimuli were digitally generated using a fixed sampling rate of 48 kHz and data buffer lengths of 4096 samples, resulting in a frequency resolution of approximately 11 Hz. Potential artifacts were detected by computing the difference between the current data buffer and an artifact-free reference buffer. The current data buffer was discarded whenever the rms value of the difference waveform exceeded a subject-specific criterion. Accepted data buffers were added to the averaging buffer. Continual replacement of the reference buffer minimized the effects of slowly varying drifts in the baseline.

We briefly outline the procedures for measuring each type of OAE below. Interested readers can consult Mimosa Acoustics technical documentation for more detailed descriptions of the measurement system (see also Lapsley-Miller et al., 2004a,b).

3.2.1 Measuring CEOAEs

CEOAEs were evoked using broadband clicks (0.5–5 kHz) ranging in intensity from 35 to 80 dB pSPL (peak-equivalent SPL). To enable comparisons with SFOAEs, which are evoked using iso-intensity pure tones, the click waveform was adjusted using the in-the-ear calibration data to produce a flat-spectrum microphone signal. Responses were averaged across 500–4000 repetitions, depending on the stimulus level. Noise floors for the measurements typically ranged from −25 to −33 dB SPL.

A typical response waveform is shown in Fig. 3-1. The large pulse is the acoustic click, the smaller, more temporally dispersed portion of the waveform is the CEOAE. CEOAEs were extracted from the ear-canal pressure waveform by using either the linear-windowing or the nonlinear-residual method. The following sections describe each method in turn. Our standard protocol used a click repetition period \(T_s\) of approximately 26 ms (1253 samples). As a check for possible efferent effects (e.g., Guinan et al., 2003), we varied the interstimulus time from roughly 20 ms up to 100
ms but found no significant dependence on repetition period.

The linear windowing method

In the linear windowing paradigm (e.g., Kemp, 1978) the stimulus and emission, \( p_{ST}(t) \) and \( p_{CE}(t) \), are extracted from the total ear-canal pressure, \( p(t) \), by applying stimulus and emission windows, \( w_s(t) \) and \( w_e(t) \):

\[
\begin{align*}
p_{ST}(t) &= w_s(t)p(t) ; \\
p_{CE}(t) &= w_e(t)p(t) . 
\end{align*}
\]

The stimulus and emission spectra are then computed by taking the 4096-point discrete Fourier transform \( \mathcal{F}\{\cdot\} \) of zero-padded waveforms \( p_{CE}(t) \) and \( p_{ST}(t) \):

\[
\begin{align*}
P_{ST}(f) &= \mathcal{F}\{p_{ST}(t)\} ; \\
P_{CE}(f) &= \mathcal{F}\{p_{CE}(t)\} . 
\end{align*}
\]

The input/output CEOAE transfer function \( T_{CE}(f; A) \) is defined as the ratio of the two spectra:

\[
T_{CE}(f; A) = \frac{P_{CE}(f; A)}{P_{ST}(f)} ,
\]

where \( A \equiv |P_{ST}| \) is the stimulus amplitude. Although we refer to the ratio as a transfer function, \( T_{CE}(f; A) \) depends on the stimulus amplitude and is therefore more correctly known as a “describing” function (e.g., Krylov and Bogolyubov, 1947; Gelb and Vander Velde, 1968).

For the windowing technique to work, the stimulus click must be sufficiently localized in time so that the end of the stimulus does not significantly overlap with the early components of the emission. Unless otherwise noted, the clicks used in these experiments were bandlimited from 0.5 to 5 kHz—the broadest flat spectrum click without notches that the measurement system was able to generate. Interference between stimulus and emission can be further reduced by the proper choice of windows. We used 10th-order recursive-exponential windows \( w_{rex}(t; \Delta t) \) (Shera and Zweig, 1993a; Kalluri and Shera, 2001a) with time offsets and widths chosen to reduce interactions between the stimulus and emission. Thus,

\[
\begin{align*}
w_s(t) &= w_{rex}(t - t_s; \Delta t_s) , \\
w_e(t) &= w_{rex}(t - t_e; \Delta t_e) ,
\end{align*}
\]

with standard offsets \( \{t_s, t_e\} = \{0, 10\} \) ms and widths \( \{\Delta t_s, \Delta t_e\} = \{5, 10\} \) ms. All offsets are relative to the center of the stimulus click at \( t = 0 \). The recursive-
Figure 3-1: Schematic diagrams of the measurement paradigms. (A) For CEOAEs, the stimulus and emission are measured using the linear windowing technique by applying recursive-exponential windows \( w_s \) and \( w_e \), to the ear-canal pressure, \( p(t) \). The windows’ center positions, \( t_s \) and \( t_e \), and widths, \( \Delta t_s \) and \( \Delta t_e \), are chosen to optimize the separation between the stimulus and emission. (B) For SFOAEs, the stimulus and emission are measured using the interleaved suppression technique. The emission is computed as the Fourier component at the probe frequency by taking the complex difference between the probe-alone and probe-suppressor segments of the ear-canal pressure. The probe-alone and probe-suppressor waveforms are extracted using rectangular windows \( w_p^{(n)} \) and \( w_{ps}^{(n)} \); only \( w_p^{(0)} \) is shown in the figure. The Fourier analysis buffer (duration \( T_w \)) contains an integral number of cycles of both probe and suppressor.
Figure 3-2: Dependence of CEOAE spectra on window offset, \( O_e \). The inset illustrates the windows \( w_e(t - \Delta t_e; \Delta t_e) \) with post-stimulus offsets ranging from 2 to 8 ms. The spectral structure of the CEOAE transfer function varies with window offset. For window offsets between 5–7 ms, CEOAE transfer functions are almost independent of \( O_e \) (thick lines). For shorter offsets (< 5 ms) the transfer functions manifest additional spectral structure (*), presumably due to interference-like interactions between the stimulus and the emission. For offsets greater than 7 ms the short latency, high-frequency components of the CEOAE degrade (*).

The exponential window is defined in footnote 10 of Kalluri and Shera (2001a).

The location of the emission analysis window must be carefully chosen to reduce interference caused by interactions between the stimulus and the early components of the emission. The window \( w_e(t) \) begins at time \( O_e = t_e - \Delta t_e/2 \) after the click (see Fig. 3-1). To determine the optimal window offset, we varied \( O_e \) until small shifts had negligible effects on the transfer function within the frequency range of interest (1–4 kHz). Offsets smaller than about 5 ms or larger than 7 ms produced significant changes in the magnitude of the transfer function (see Fig. 3-2). We adopted the value \( O_e = 5 \) ms for all the results shown here. Because CEOAEs are dispersed in time, with high frequency components arriving before the low frequency components, the optimal window for the 1–4 kHz region will not be optimal for emissions in other frequency bands.

The nonlinear residual method

The nonlinear residual method is an alternate and generally more popular procedure for measuring CEOAEs. In this method CEOAEs are extracted by exploiting the nonlinear compressive growth of the emissions in conjunction with the linear growth of the stimulus. Three identical clicks are followed by a fourth click that is three times larger and of the opposite polarity. The average of these four components is then computed, and the resulting residual is the CEOAE estimate.
Unlike the linear windowing technique, in which short-latency components of the emission are typically eliminated, the nonlinear residual method separates the emission from the stimulus without eliminating the early arriving components of the emission. However, to avoid confusion, reduce potential artifacts due to system distortion, and enable direct comparison between the two CEOAE techniques we apply the standard emission window, $w_{e}(t)$, to the nonlinear-derived emission as well. Therefore, just as in the linear technique, early arriving components of the emission are eliminated.

### 3.2.2 Measuring SFOAEs

We measured the SFOAE pressure, $P_{SF}(f)$, using a variant of the suppression method (Shera and Guinan, 1999; Kalluri and Shera, 2001b). As illustrated in Fig. 3-1, the emission is obtained as the complex difference between the ear-canal pressure at the probe frequency ($f_{p}$) measured first with the probe tone alone and then in the presence of a more intense (55 dB SPL) suppressor tone at a nearby frequency, $f_{s}$, roughly 47 Hz below the probe frequency (Fig. 3-1). The suppressor was presented in interleaved time segments to minimize possible artifactual contamination from time-varying drifts in the base signal. To reduce spurious contamination by earphone distortion, the probe and suppressor were generated using separate sounds sources.

The probe-alone waveform, $p_{p}(t)$, and probe-suppressor waveform, $p_{ps}(t)$, are obtained from the measured ear-canal pressure, $p(t)$, by averaging over two sub-segments extracted using windows $w_{p}(t)$ and $w_{ps}(t)$:

$$
\begin{align*}
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \q
steady state after switching the suppressor tone on or off. The window duration $T_w$ equals that of the Fourier analysis buffer. Stimulus frequencies were chosen so that the analysis buffer (duration $T_w = N\Delta t$, where $N$ is the buffer size and $\Delta t$ is the reciprocal of the sampling rate) always contained an integral number of cycles of both probe and suppressor. For the measurements reported here $\{T_o, T_1, T_b\} = \{\frac{1}{4}, \frac{5}{2}, 5\} T_w$. The SFOAE pressure is computed as

$$P_{SF}(f) = \mathcal{F}\{p_p(t)\} - \mathcal{F}\{p_{ps}(t)\}e^{i2\pi f_pT_1} ,$$

(3.13)

where $\mathcal{F}\{\cdot\}$ indicates the 4096-point discrete Fourier transform at the probe frequency, $f_p$. The stimulus pressure is extracted from the probe-suppressor segment:

$$P_{ST}(f) = \mathcal{F}\{p_{ps}(t)\}e^{i2\pi f_pT_1} .$$

(3.14)

By analogy with Eq. (3.5) for $T_{CE}(f; A)$, the transfer function $T_{SF}(f; A)$ is defined as the ratio of probe-frequency spectral components

$$T_{SF}(f; A) = \frac{P_{SF}(f; A)}{P_{ST}(f)} ,$$

(3.15)

where we have now explicitly indicated the dependence on stimulus amplitude ($A \equiv |P_{ST}|$). We measured $T_{SF}(f; A)$ with a frequency resolution of approximately 23 Hz using probe-tone levels ranging from approximately 10 to 40 dB SPL. We typically employed 32 averages at the highest probe level and 128 averages at the lowest.

### 3.3 Experimental Complications

Before describing our main results, we first address two measurement issues that complicate the comparison between $T_{CE}(f; A)$ and $T_{SF}(f; A)$. The first pertains to differences between the two different CEOAE measurement methods. The second deals with complications arising from synchronized spontaneous otoacoustic emissions (SSOAE).

#### 3.3.1 CEOAE transfer functions from linear and nonlinear methods

Figure 3-3 compares the CEOAE transfer functions $T_{CE}(f; A)$ measured using the linear-windowing and nonlinear-residual methods. We denote transfer functions measured using the two methods by $T_{CE}(f; A)$ and $T_{CE}^{NL}(f; A)$, respectively. For brevity, we show measurements from one subject; similar results were obtained in all.
Equivalence between CEOAEs and SFOAEs

-20

(A) Linear windowing

Kalluri & Shera

-25
-30
-35
-40
-45
-50
-55

1 1.2 1.4 1.6 1.8 2 2.2 2.4

Frequency [kHz]

Magnitude of $T\text{CE}(f; A)$ dB pSPL

Spectral Level [dB pSPL]

43 50 60 70 80

(B) Nonlinear derived

Figure 3-3: Comparison of $T\text{CE}(f; A)$ and $T_{\text{CE}}^{\text{NL}}(f; A)$ at stimulus levels ranging from 43 to 80 dB pSPL. The two techniques yield qualitatively similar results at high stimulus levels (70–80 dB pSPL) but diverge at lower levels. Whereas $T_{\text{CE}}^{\text{NL}}(f; A)$ falls below the noise floor (gray shaded area), $T\text{CE}(f; A)$ continues to increase until it becomes independent of level. The noise floor shown here was measured at the lowest level. Note that because we are computing transfer functions, the transfer-function noise floor is scaled by the stimulus and is therefore much lower at higher stimulus levels.
Although both the linear-windowing and nonlinear-residual techniques yield qualitatively similar values of $T_{CE}(f; A)$ at high stimulus levels, CEOAEs at low levels can only be extracted using the linear technique. As stimulus levels are decreased from 80 to 60 dB pSPL, the magnitudes of both $T_{CE}(f; A)$ and $T_{CE}^{NL}(f; A)$ increase. At these levels $T_{CE}(f; A)$ and $T_{CE}^{NL}(f; A)$ have similar peaks, notches, and phase behaviors. This similarity in behavior does not carry through to the lowest levels. As stimulus levels are further reduced, the magnitude of $T_{CE}(f; A)$ continues to grow and eventually becomes nearly independent of level. By contrast, $T_{CE}^{NL}(f; A)$ reaches a maximum value and then falls quickly into the noise floor. The combination of results—near level independence of $T_{CE}(f; A)$ and the rapid fall of $T_{CE}^{NL}(f; A)$ at low stimulus levels—suggests that CEOAEs grow almost linearly at the lowest stimulus levels. Note, however, that by using the nonlinear-derived method, Withnell and McKinley (2005) found short-latency CEOAE components in guinea pigs that appear to result from nonlinear mechanisms within the cochlea. When measured using the linear-windowing protocol, these short latency components would typically be obscured by the stimulus. Since our measurements had a residual short-latency stimulus artifact due to earphone nonlinearities (e.g., Kapadia et al., 2005), we cannot rule out the possibility that human CEOAEs also contain small short-latency nonlinear components buried beneath the stimulus artifact. Because the nonlinear technique cannot be used to measure $T_{CE}(f; A)$ at the lowest stimulus levels, all subsequent CEOAE measurements presented in this paper were made using the linear protocol.

3.3.2 Synchronized spontaneous otoacoustic emissions

Some of our subjects had synchronized spontaneous otoacoustic emissions (SSOAEs). SSOAEs are long-lasting transient responses that are not always identifiable by conventional SOAE searches, in which no external stimulus is presented. SSOAEs can, however, be detected when they are evoked by or synchronized to an applied stimulus, in this case the click used to evoke CEOAEs.

We measured SSOAEs using a variant of the standard linear-windowing technique for measuring CEOAEs. The variant employed an inter-click time of 100 ms rather than the standard 20 ms used in the CEOAE measurements. To detect SSOAEs we then computed and compared the response spectra, $P_{CE}^{i}(f)$, in five partially overlapping analysis windows centered at different post-stimulus times ($t_{i}^{e} = 15, 30, 50, 70,$ and $90 \text{ ms}$):

$$P_{CE}^{i}(f) = \mathcal{F} \{ p(t) w(t - t_{i}^{e}; \Delta t) \} \quad i = 1, 2, \ldots, 5,$$

where the nominal window duration $\Delta t$ is 20 ms. Figure 3-4 shows the spectra for the five windowed segments in two subjects. The dark thick line gives the spectrum of the
Figure 3-4: Synchronized SOAEs are characterized by sustained activity in the CEOAE spectrum. SSOAEs were identified using the linear-windowing technique with an inter-stimulus time of 100 ms. Five analysis windows with center offsets $t_{e1}^1$ through $t_{e5}^5$ ranging from 15-90 ms after the stimulus were applied to the measured ear-canal pressure. In subjects without measurable SSOAEs, only the response in the first window ($t_{e1}^1$, thick line) contains significant emission energy; responses in all subsequent windows ($t_{e2}^2$ through $t_{e5}^5$, thin lines) are small by comparison. In subjects with SSOAEs a response at SSOAE frequencies appears as a peak in all the windows.
response during the first window, centered at 15 ms after the click. The narrow lines show the spectra measured during subsequent windows. In two of the four subjects, significant response energy occurs only within the first 20 ms, and we considered these subjects to have unmeasurable SSOAEs. In the remaining two subjects, some peaks in the spectrum [e.g., those identified by asterisks (*) in Fig. 3-4] disappear more slowly over the five windows. We identified these long-lasting transient responses to the click as SSOAEs.

The existence of SSOAEs complicates the measurement of $T_{CE}(f; A)$, and to a lesser extent $T_{SF}(f; A)$, in at least two ways. First, SSOAEs make it more difficult to determine the stimulus spectrum [i.e., the denominator in Eq. (3.5)]. In subjects with long-lasting SSOAEs, the response in the stimulus window is contaminated by responses which have not fully decayed by the time the next stimulus presentation occurs. Contamination by SSOAEs typically creates spurious ripples in the measured stimulus spectrum. To reduce errors in the computation of the transfer function, we estimate the stimulus spectrum at low stimulus levels (i.e., at 40–70 dB pSPL) by appropriately rescaling the stimulus spectrum measured at high levels (i.e., at 80 dB pSPL). This rescaling procedure reduces the error because the relative influence of SSOAE ripples is smallest at high levels.

Second, SSOAEs increase variability in the measurements, as shown in Fig. 3-5. The gray lines in the bottom panel show individual measurements of $T_{CE}(f; A)$ made during several sessions on two different days. Note the increased variability near SSOAE frequencies, indicated by asterisks in the top panel. The variability presumably reflects an instability in the relative phase with which the stimulus initiates or synchronizes the SSOAE. For example, sometimes the SSOAE seems to add to the CEOAE; at other times it appears to subtract. To reduce this variability in subjects with SSOAE we calculate the (complex) average of measurements made during multiple sessions at each frequency (stars in the figure) before comparing with measurements of $T_{SF}(f; A)$. We find that matches between $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are generally improved significantly by this ensemble averaging.

### 3.4 Comparison of SFOAE and CEOAE Transfer Functions

Figure 3-6 shows measurements of $T_{CE}(f; A)$ and $T_{SF}(f; A)$ versus frequency in a subject lacking the complications introduced by the existence of SSOAEs. Error bars on the magnitude represent the standard deviation of the mean and are drawn one
Figure 3-5: Averaging out variability due to SSOAEs. The top panel shows the results of the SSOAE identification procedure in a subject with strong SSOAEs. The format is identical to Fig. 3-4; the frequencies of significant SSOAEs are marked by asterisks (*). The bottom panel shows the magnitude of individual $T_{CE}(f; A)$ measurements (thin gray lines) made at 40 dB pSPL. The $T_{CE}(f; A)$ measurements were made during two different days (dashed and solid gray, respectively). Each gray line represents the average of over 1000 consecutive presentations. The magnitude of the complex ensemble average of these individual measurements of $T_{CE}(f; A)$ is shown by the stars. The diamonds are the $T_{SF}(f; A)$ at a comparable stimulus level down in the low-level linear regime (20 dB SPL).
Equivalence between CEOAEs and SFOAEs

Kalluri & Shera

Figure 3-6: CEOAE and SFOAE transfer functions. The left and right columns show the magnitude (top) and phase (bottom) of $T_{CE}(f; A)$ and $T_{SF}(f; A)$, respectively, for a subject without SSOAEs (Subject #1). The symbols identify the stimulus intensity, which ranged from 40 to 80 dB pSPL for clicks and from 10 to 40 dB SPL for tones. Error bars on the magnitude represent the standard error of the mean.

standard deviation above and below the mean value. As the figure demonstrates, $T_{CE}(f; A)$ and $T_{SF}(f; A)$ have qualitatively similar spectral structure, including a rapidly varying phase and magnitude peaks and notches that occur at approximately the same frequencies in both transfer functions. Both transfer functions also share a qualitatively similar dependence on stimulus intensity. At the lowest levels, transfer function magnitudes appear nearly independent of level, consistent with a region of approximate linearity near threshold. At higher intensities, the transfer-function magnitudes generally decrease, consistent with compressive nonlinear growth in emission amplitudes. Although the qualitative similarity of these prominent features in the two transfer functions suggests that clicks and tones evoke emissions via similar mechanisms, definitive conclusions require a more careful comparison.

In subjects without SSOAE, the transfer function’s standard deviation is computed either from the noise floor or, when possible, by finding the deviation of multiple runs. In subjects with SSOAE, we always made multiple measurements, particularly at low stimulus levels where the SSOAE were likely to have the most influence. Because SOAEs typically have constant magnitude but random phase, complex averaging of multiple runs allows one to partially eliminate the SSOAE. The standard deviation represents the uncertainty of the mean value.
3.4.1 Equating stimulus intensities

Assuming that meaningful and more quantitative comparisons between the responses to click and tonal stimuli are even possible in this nonlinear system, the comparisons need to take into account that both CEOAEs and SFOAEs depend on stimulus intensity. Even if the responses are comparable in principle, comparisons made at different effective intensities may amount to comparing apples and oranges. Complicating the situation is the fact that click and tone intensities are conventionally specified in different ways. Whereas pure-tone intensities are measured in SPL, click intensities are measured in peak-equivalent SPL (pSPL), defined as the SPL of a pure tone with the same peak pressure as the click waveform. At what click intensity (in dB pSPL) should one measure CEOAEs in order best to compare them with SFOAEs measured at a given probe level (in dB SPL)? Whether or not this question has a meaningful answer depends on the nature of the nonlinearities involved in OAE generation.

Figure 3-7 indicates that not only does the question have a compelling empirical answer, but that the answer, although approximate, is rather simple. The figure shows relative CEOAE levels (Subject #2, also without SSOAEs) measured in a narrow frequency range using clicks of various stimulus intensities (pSPL) and bandwidths. The data can be made to fall approximately along a single continuous growth function if the stimulus intensity is expressed in what we call component-equivalent SPL (dB cSPL). For any stimulus, the cSPL at any given frequency is the SPL of the Fourier component of the stimulus at that frequency. cSPL and SPL are equivalent for pure tones. As the figure illustrates, the cSPL of any component of a click stimulus of fixed pSPL decreases as the click bandwidth increases.

3.4.2 Comparison of the growth functions

Figure 3-8 demonstrates that the single growth function shown to characterize the intensity dependence of $T_{CE}(f; A)$ in Fig. 3-7 applies also to SFOAEs at the same frequency. (In the context of Fig. 3-7, the SFOAE stimulus might be thought of as a “click” in which the bandwidth has been reduced so much that the stimulus comprises nothing but a single pure tone.) Although the measurements of $T_{SFOAE}(f; A)$ are limited to stimulus intensities less than 45 dB SPL, the agreement between the two growth functions (plotted vs cSPL) is excellent throughout the measured range. Although variability in the measurements increases at the lowest intensities (especially for SFOAEs), the common growth function manifests a region of approximately linear growth below about 10–15 dB cSPL. The growth function then gradually transitions into a compressive regime whose slope is approximately $-1$ dB/dB at higher intensities.
Figure 3-7: Equivalent levels for clicks of different bandwidth. The figure shows the intensity dependence of $T_{CE}(f; A)$ as measured with clicks of four different bandwidths (namely, 1–5, 1–3, 1–2, and 1–1.5 kHz). Open symbols give $T_{CE}(f; A)$ versus click intensity expressed in peak-equivalent SPL (pSPL). Filled symbols show the same $T_{CE}(f; A)$ data versus component-equivalent SPL (cSPL). The data were taken in a narrow frequency band near a peak in the $T_{CE}(f; A)$ amplitude (1.15–1.2 kHz). Multiple points at the same intensity and bandwidth represent values of $T_{CE}(f; A)$ at different frequencies in the range.

Figure 3-8: The transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ share a common intensity dependence. The $T_{CE}(f; A)$ data from Fig. 3-7 are shown here with filled symbols; $T_{SF}(f; A)$ are shown as open squares. All stimulus levels are expressed in dB cSPL. Both $T_{CE}(f; A)$ and $T_{SF}(f; A)$ were measured over the same frequency range (1.15–1.2 kHz).
3.4.3 Comparison at equivalent intensities

Figure 3-9 compares transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ across frequency at two different stimulus intensities matched by expressing both in component-equivalent SPL (see Fig. 3-8). The lower of the two intensities (20 dB cSPL) falls within or just above the low-level linear regime; the higher intensity (40 dB cSPL) evokes responses in the region of compressive OAE growth. At each intensity, the figure shows transfer functions from two subjects (right and left columns), neither of whom had identifiable SSOAEs in the measured frequency range (1-2.4 kHz). As the figure demonstrates, the transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are almost identical at matched intensities. The agreement extends even to the spectral notches, regions where one might expect responses to be especially sensitive to small changes. Since details of the phase are obscured by the large OAE delay and phase unwrapping, Fig. 3-10 replots the transfer-function phases after subtracting out smooth trend lines that capture the secular variation of the phase. The resulting comparisons show that the agreement between $\angle T_{CE}(f; A)$ and $\angle T_{SF}(f; A)$ is generally excellent, even in microstructural detail.

Figure 3-11 extends the comparison to subjects with SSOAE. In these subjects, CEOAE transfer functions were obtained by averaging across measurement sessions, as described in Sec. 3.3.3.3. Although small differences between $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are found, especially at the lower intensity (e.g., in Subject #4, where sharp peaks in $|T_{SF}(f; A)|$ are can be seen at SSOAE frequencies), the overall agreement is still excellent.

3.5 Discussion

At low and moderate stimulus intensities human CEOAE and SFOAE input/output transfer functions are nearly identical. When stimulus intensity is measured in component-equivalent SPL (cSPL), we found that CEOAE and SFOAE transfer functions have equivalent growth functions at fixed frequency and equivalent spectral characteristics at fixed intensity. This strong similarity suggests that the OAEs evoked by broad and narrowband stimuli (clicks and tones) are generated by the same mechanism.

3.5.1 Possible limits of application

Although our conclusions may apply more widely, we summarize below the known limitations of our study. (1) Our comparisons between CEOAEs and SFOAEs were time consuming, and were therefore performed in a relatively small number of subjects ($n = 4$). Nevertheless, since the subjects were selected only for having measurable...
Figure 3-9: Transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ in two subjects without SSOAE. Each of the two columns shows measurements from a different subject. The top row shows transfer-function magnitudes measured at 20 dB cSPL; the middle row shows magnitudes at 40 dB cSPL. The bottom row shows unwrapped phases at both intensities. In all panels, $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are shown with filled and open symbols, respectively.
Equivalence between CEOAEs and SFOAEs

Kalluri & Shera

Figure 3-10: Details of transfer-function phase. The figure shows the values of $\angle T_{CE}(f; A)$ and $\angle T_{SF}(f; A)$ reproduced from Fig. 3-9 after subtracting out smooth curves that capture the secular variation of the phase. Each of the two columns shows measurements from a different subject. The top row shows detrended phases measured at 20 dB cSPL; the bottom row shows detrended phase at 40 dB cSPL. In any given panel, the same trend curve was subtracted from both $\angle T_{CE}(f; A)$ and $\angle T_{SF}(f; A)$, which are shown using filled and open symbols, respectively.
Figure 3-11: Transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ in two subjects with SSOAE. Each of the two columns shows measurements from a different subject. The top row shows transfer-function magnitudes measured at 20 dB cSPL; the middle row shows magnitudes at 40 dB cSPL. The bottom row shows unwrapped phases at both intensities. In all panels, $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are shown with filled and open symbols, respectively. SSOAE frequencies are identified by asterisks in the top panels.
emissions, and because we found similar results in all, it seems unlikely that the near equivalence we report is merely a statistical fluke. (2) All of our subjects had normal hearing. Although additional studies are needed to determine whether our findings generalize, we have no reason to suspect that similar conclusions will not apply to impaired ears, so long as their emissions remain measurable. (3) We used low to moderate stimulus levels (35–80 dB pSPL for broadband clicks, 10–40 dB SPL for tones) and cannot rule out the possibility that SFOAE and CEOAE transfer functions differ more significantly at higher stimulus levels. (4) In order to reduce interference between the stimulus and the emission, we used time widows to eliminate CEOAE components arriving earlier than about 5 ms after the stimulus peak. In addition to removing high-frequency components of the response, this windowing may also have removed possible short-latency low-frequency components generated in the base of the cochlea. Although accurate estimates of the magnitudes of these components were compromised by system nonlinearities (see below), measurements in test cavities imply that any such short-latency components must be small relative to the long-latency components. (5) Our comparisons are limited to the frequency range of 1 to 3 kHz. In particular, we did not explore the behavior in more apical regions of the cochlea, where emission mechanisms may differ from those in the base (Shera and Guinan, 2003; Siegel et al., 2005). (6) We did not systematically explore a wide range of stimulus presentation rates (e.g., for the click stimuli) in every subject. Since high-rate clicks are generally much more effective elicitors of efferent activity than the stimuli used to measure SFOAEs (Guinan et al., 2003), we checked for differences related to efferent effects by varying the click-repetition period in two subjects. Although we found no obvious effects of click-repetition period in these subjects, the strength of otoacoustic efferent effects varies from individual to individual, and we may simply have “gotten lucky.” It remains possible, even likely, that differences in the strength of efferent feedback elicited by the two stimuli can produce differences in $T_{CE}(f; A)$ and $T_{SF}(f; A)$ in some subjects, at least when $T_{CE}(f; A)$ is measured using high-rate clicks. (7) Finally, our measurements are in humans, a species whose OAE characteristics differ in some respects from those of many laboratory animals (e.g., humans have longer OAE latencies and smaller distortion-source emissions). The near-equivalence we find between $T_{CE}(f; A)$ and $T_{SF}(f; A)$ remains to be examined in other species.

### 3.5.2 Quasilinearity of the transfer functions

If the mechanisms of OAE generation and propagation were completely linear, the near-equivalence we find between CEOAE and SFOAE transfer functions would be entirely expected. The response of a linear system can be equivalently characterized either in the time domain using broadband stimuli (such as the click stimulus used to
Evoke CEOAEs) or in the frequency domain using narrowband stimuli (such as the pure tone used to evoke SFOAEs). If the cochlea were a linear system, the principle of superposition requires that transfer functions measured in the time and frequency domains be identical, regardless of the details of emission generation.

Our data support the notion that cochlear responses are nearly linear at levels approaching the threshold of hearing. For example, we find that the transfer functions $T_{CE}(f; A)$ and $T_{SF}(f; A)$ are almost identical and independent of stimulus intensity at low levels. Furthermore, CEOAE transfer functions obtained using the nonlinear-residual method, a method that relies on nonlinear OAE growth to extract the emission, fall into the noise floor at low intensities. These results are consistent with previous OAE measurements (e.g., Zwicker and Schloth, 1984; Shera and Zweig, 1993a), including those demonstrating approximate linear superposition among OAEs evoked by various low-level stimuli (Zwicker, 1983; Probst et al., 1986; Xu et al., 1994). Linearity of CEOAE and SFOAE responses at low levels is also consistent with basilar-membrane mechanical responses (reviewed in Robles and Ruggero, 2001), which manifest approximate linearity at levels approaching threshold.

Since the operation of the cochlea is certainly nonlinear at intensities not far above threshold, our finding that CEOAE and SFOAE transfer functions continue to match even at moderate levels is more unexpected. The continuing match suggests that as stimulus intensities rise the cochlea emerges gracefully from the low-level linear regime. In particular, the observation that growth functions expressed in terms of component-equivalent SPL fall along a continuous curve suggests that to a good approximation CEOAEs are generated within a nonlinear system in which the stimulus is first decomposed into narrowband components, each of which is then “passed through” the mechanisms of OAE generation and nonlinearity almost independently before being recombined with the others to form the output signal. (The transformation to cSPL would hold exactly if the nonlinearity acted independently on each frequency component of the stimulus.) In this picture, interactions among the different frequency components of the stimulus (e.g., via suppression or intermodulation distortion) appear to play only a secondary role in CEOAE and SFOAE generation. Our results are consistent with studies of the OAEs evoked by broadband noise (Maat et al., 2000), where Wiener-kernel analysis indicates that although the overall emission amplitude varies with stimulus intensity, the cochlear response appears approximately linear at each level. Analogous results, including strong if imperfect matches between responses evoked by broad and narrowband stimuli, are found in measurements of basilar-membrane motion (e.g., Recio and Rhode, 2000; de Boer and Nuttall, 2002).

Our findings are also consistent with those of Prieve et al. (1996), who found that CEOAEs and tone-burst evoked OAEs (TBOAEs) have similar growth functions. They concluded that emissions evoked by the two stimuli share common mechanisms.
of generation and, in particular, that both are generated by mechanisms acting in independent frequency channels. This conclusion was questioned by Yates and Withnell (1999b), who pointed out that although the tone-burst bandwidths (which generally spanned an octave or more) were narrower than those of the clicks, they were still broad enough to excite the same complex cross-frequency interactions as the click. They therefore argued that the growth functions matched not because of OAE generation via independent frequency channels, but precisely the opposite: because both stimuli produce nonlinear interactions among the different spectral components of the stimulus. Our data do not support Yates and Withnell’s suggestion: We measured SFOAEs using continuous narrow-band pure tones (rather than the relatively broadband TBOAEs used by Prieve et al.) and still found the reported match between wide- and narrowband growth functions.

### 3.5.3 Consistency with the DP-place component of DPOAEs

The match we find between $T_{CE}(f; A)$ and $T_{SF}(f; A)$ is consistent with emission measurements that report strong similarities between CEOAEs and certain DPOAEs (Knight and Kemp, 1999), in particular upper-sideband DPOAEs and lower-sideband DPOAEs measured at $f_2/f_1$ ratios close to 1. DPOAEs are typically mixtures of emissions originating from at least two different regions of the cochlea, namely the region where the responses to the primaries overlap and the region tuned to the distortion-product frequency (e.g., Kim, 1980; Kemp and Brown, 1983; Gaskill and Brown, 1990; Brown et al., 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Heitmann et al., 1998; Talmadge et al., 1999; Kalluri and Shera, 2001b; Knight and Kemp, 2001). Theory and experiment both indicate that the relative contribution of the components from these two locations varies systematically with stimulus parameters (e.g., Fahey and Allen, 1997; Knight and Kemp, 2001; Shera and Guinan, 2006). In particular, upper-sideband DPOAEs and lower-sideband DPOAEs measured with $f_2/f_1$ ratios close to 1 are generally dominated by emissions from the distortion-product place, whose characteristics are very similar to SFOAEs. Indeed, Kalluri and Shera (2001b) showed by direct comparison that the DPOAE component originating from the DP place closely matched the SFOAE evoked at the same frequency. Previous results thus establish that (1) CEOAEs resemble the DP-place component of DPOAEs and (2) the DP-place component of DPOAEs matches SFOAEs. Taken together, these results are consistent with the equivalence reported here between CEOAE and SFOAE transfer functions.
3.5.4 Implications for emission mechanisms

Our results contradict two proposed models of CEOAE generation, both of which suggest that CEOAEs originate primarily by nonlinear mechanisms within the cochlea. Nobili and colleagues argue that CEOAEs arise from spatially complex, nonlinear “residual oscillations” of the basilar membrane that trace their origin to spectral irregularities in middle-ear transmission (Nobili, 2000; Nobili et al., 2003a,b). Based on model simulations, Nobili et al. conclude that when found in the absence of spontaneous emissions, transient evoked OAEs are attributable to the characteristics of forward middle-ear filtering. In this view, CEOAEs result from mechanisms that are both inherently nonlinear and fundamentally different from those responsible for generating SFOAEs. We note, for example, that Nobili et al.’s proposed middle-ear filtering mechanism for generating CEOAEs cannot produce SFOAEs at any level of stimulation: Although CEOAEs are evoked by transient stimuli containing many frequency components, and are therefore potentially sensitive to frequency variations in middle-ear transmission as proposed, SFOAEs are evoked by pure (single-frequency) tones and, ipso facto, cannot originate via any mechanism that operates across frequency. We show here that the characteristics of CEOAEs and SFOAEs are nearly identical (in ears both with and without SSOAEs), in clear contradiction to Nobili et al.’s model predictions.

Our findings also contradict the notion that CEOAEs arise via nonlinear interactions among the frequency components of the stimulus. Based on measurements in guinea pig in which they evoked CEOAEs using high-pass filtered clicks and identified significant OAE energy outside the stimulus passband, Yates and Withnell (1999b) proposed that CEOAEs result primarily from intermodulation distortion within the cochlea. CEOAEs, they suggest, are “predominantly composed of intermodulation distortion energy; each component frequency of a click stimulus presumably interacts with every other component frequency to produce a range of intermodulation distortion products” (Withnell et al., 2000). Our finding that CEOAE and SFOAE transfer functions are almost identical argues against this interpretation, at least in humans. Although the contribution of nonlinear intermodulation distortion mechanisms to human CEOAEs appears small at low and moderate levels, our use of the windowing technique to measure CEOAEs may have eliminated short-latency distortion components present in the response (e.g., Knight and Kemp, 1999; Withnell and McKinley, 2005). Because of a stimulus artifact due to earphone nonlinearities we were unable to accurately quantify the size of any short-latency physiological component using the nonlinear residual method. Nevertheless we can report that any such short-latency component is small enough to be indistinguishable from the distortion measured in a test cavity of similar impedance. Any short-latency nonlinear component in human ears is therefore small relative to the long-latency linear response. Similar conclusions
apply also to human SFOAEs (Shera and Zweig, 1993a).

Although the observed equivalence between CEOAEs and SFOAEs contradicts these inherently nonlinear models of CEOAE generation, the equivalence is entirely consistent with predictions of the coherent-reflection model (e.g., Zweig and Shera, 1995; Talmadge et al., 1998; Shera and Guinan, 2006). In this model, OAEs are generated by a process equivalent to wave scattering by pre-existing (place-fixed) micromechanical perturbations in the organ of Corti. Not only does the coherent-reflection model predict the empirical equivalence between $T_{CE}(f; A)$ and $T_{SF}(f; A)$, the model also predicts the observed spectral characteristics of the transfer functions across frequency (e.g., their slowly varying amplitudes punctuated by sharp notches and their rapidly rotating phases).

Because different stimuli are used to evoke them, CEOAEs and SFOAEs are conventionally classified as different OAE types. Our results establish, however, that at low and moderate stimulus intensities these two OAE “types” are really the same emission evoked in different ways—CEOAEs and SFOAEs are better understood as members of the same emission family. Our findings thus support the mechanism-based classification scheme proposed elsewhere (Shera and Guinan, 1999; Shera, 2004).

**Acknowledgments**

This work was supported by grant R01 DC003687 from the NIDCD, National Institutes of Health. We thank John Guinan and Chris Bergevin for helpful comments on the manuscript.
Chapter 4

Distortion-Product Source Unmixing

4.1 Introduction

Mammalian otoacoustic emissions (OAEs) have generally been regarded as originating through nonlinear electromechanical distortion (e.g., Kemp, 1978, 1997, 1998; Probst et al., 1991; Allen and Neely, 1992; Allen and Lonsbury-Martin, 1993; Patuzzi, 1996). Shera and Guinan (1999), however, argue that OAEs arise by at least two fundamentally different mechanisms within the cochlea. These differences in mechanism, they suggest, can profitably be used to define an “OAE family tree.” The mechanism-based taxonomy groups emissions into two basic types: distortion-source emissions, which arise by nonlinear distortion induced by the traveling wave, and reflection-source emissions, which arise via linear reflection off pre-existing micromechanical impedance perturbations (Shera and Zweig, 1993b; Zweig and Shera, 1995). This distinction between distortion- and reflection-source emissions differs from the “wave-” and “place-fixed” dichotomy maintained by Kemp and Brown (1983) in that the latter was introduced and developed within an integrated framework that views all OAEs as manifestations of cochlear mechanical nonlinearity. The mechanism-based taxonomy, by contrast, emphasizes the fundamental differences between linear (reflection-source) and nonlinear (distortion-source) emission mechanisms.

The analysis underlying the taxonomy predicts that the two types of OAEs mix to form the evoked emissions measured in the ear canal. In any given measurement, the different emission types contribute in degrees dependent on species, stimulus parameters, and cochlear state. As an example of the process, Shera and Guinan suggest that the generation of lower-side-band distortion-product otoacoustic emissions (DPOAEs) can be understood in terms of the mixing of the two OAE types. Much of DPOAE
fine structure apparently arises through the interference of emissions originating from two distinct cochlear locations (e.g., Kim, 1980; Kemp and Brown, 1983). Although the “two-place model” for DPOAEs now appears well established (e.g., Gaskill and Brown, 1990; Brown et al., 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Heitmann et al., 1998; Fahey and Allen, 1997; Siegel et al., 1998), the taxonomy identifies the two interfering emission components as arising not simply from two distinct locations, but, more importantly, via two different mechanisms.

The proposed generation process is illustrated in Fig. 4-1. The primary traveling waves, at frequencies \( f_1 \) and \( f_2 \) (with \( f_2 > f_1 \)), interact to produce a region of nonlinear distortion (D), located near the peak of the \( f_2 \) wave, in which distortion creates energy at distortion-product frequencies. In particular, traveling waves at the frequency \( f_{dp} = 2f_1 - f_2 \) are generated that travel in both directions. The backward-traveling wave propagates to the ear canal, where it appears as a distortion source emission. The forward-traveling wave propagates to its characteristic place, where it undergoes partial reflection (R) near the peak of its wave envelope, generating a second backward-traveling wave that propagates to the ear canal (a reflection-source emission). The two types of emission mix in the ear canal.\(^1\)

The proposed model thus predicts that the two components originate not simply from two different regions of the cochlea but—more significantly—by two fundamentally different mechanisms. Similar predictions emerge from recent modeling studies (e.g., Talmadge et al., 1998, 1999; Mauermann et al., 1999a,b). Based on nonlinear cochlear models that meet the requirements detailed by the theory of coherent reflection filtering for the generation of realistic reflection emissions (Shera and Zweig, 1993b; Zweig and Shera, 1995),\(^2\) these studies incorporate both classes of emission-generating mechanisms (i.e., nonlinear distortion and linear coherent reflection). The primary goal of the experiments reported here was to test this two-mechanism model for DPOAE generation.

According to the analysis underlying the taxonomy, distortion- and reflection-source emissions manifest very different frequency-dependencies in their phase. In a nutshell, the argument runs roughly as follows:

**Distortion-source OAEs**: Nonlinear distortion depends upon the interaction between the two primary traveling waves. When produced using frequency-

---

\(^1\)Note that for brevity this simple synopsis neglects contributions to the total reflection-source emission arising from multiple internal reflection within the cochlea (i.e., from multiple cycles of partial reflection at the stapes and linear coherent reflection within the R region).

\(^2\)In a nutshell, the theory says that given “almost any” arrangement of micromechanical impedance perturbations (i.e., an arrangement with the appropriate spatial-frequency content, such as perturbations that are randomly and densely distributed), a model will produce realistic reflection emissions whenever the peak region of the traveling wave has a slowly varying wavelength and an envelope that is simultaneously both tall and broad.
Figure 4-1: Schematic diagram of the two-mechanism model. The figure illustrates the generation of SFOAEs (top) and DPOAEs (bottom) at low sound levels. Each panel shows phase lag relative to stimulus phase (lag increasing downward) of forward- and backward-traveling waves versus cochlear location. At low stimulus levels, SFOAEs ($P_{sfe}$) result from coherent reflection ($R$) in the region near the peak of the traveling-wave envelope. For DPOAEs, the primary traveling waves produce a region of nonlinear distortion ($D$), located near the peak of the $f_2$ wave (at $x_2$), where nonlinear distortion generates traveling waves at the frequency $f_{dp}$ that travel in both directions (shown here for the case $f_{dp} = 2f_1 - f_2$, where $f_{dp}$ equals the SFOAE frequency shown in the top panel). The backward-traveling wave propagates to the ear canal (where it appears as the distortion-source emission, $P_{dp}^D$). The forward-traveling wave propagates to its characteristic place (at $x_{dp}$), where it undergoes partial reflection ($R$) near the peak of its wave envelope, generating a second backward-traveling wave that propagates to the ear canal (the reflection-source emission, $P_{dp}^R$). The two types of emission combine to produce the DPOAE measured in the ear canal ($P_{dp} = P_{dp}^D + P_{dp}^R$). For simplicity, phase shifts due to propagation through the middle ear and reflection by the stapes are not shown. Adapted, with permission, from Shera and Guinan (1999).
scaled stimuli, the combined excitation pattern of the primary traveling waves simply translates along the cochlear partition as the stimulus frequencies are varied. This approximate translation invariance (or "shift-similarity") follows from local scaling symmetry and the logarithmic form of the cochlear frequency-position map. Approximate shift similarity ensures that the amplitudes and phases of the primary waves—and hence any nonlinear interactions between them—remain nearly invariant in a coordinate system that moves with the spatial envelope of the $f_2$ traveling wave as the frequencies are swept (Fig. 4-2, left). OAEs generated by frequency-scaled nonlinear distortion therefore manifest a nearly constant phase.

Reflection-source OAEs: According to the theory of coherent reflection filtering (Zweig and Shera, 1995), reflection-source emissions are generated when a forward-traveling wave reflects off "random" perturbations in the mechanics of the cochlea. The phase of each scattered wavelet depends on the phase of the forward-traveling wave at the location of scattering. Since the micromechanical impedance perturbations are fixed in space (unlike sources of nonlinear distortion, which move with the excitation pattern as the frequency changes), the phase of the incident wave at each perturbation changes as the frequency of the forward-traveling wave is varied (Fig. 4-2, right). Consequently, OAEs generated by linear reflection manifest a phase that rotates rapidly with frequency.

In this paper we apply this reasoning to test the two principal predictions of the two-mechanism model, as suggested by the taxonomy and framework presented in Fig. 4-1. Specifically, we test the predictions that

#1 The total distortion-product emission, $P_{dp}$, represents the sum of distortion- and reflection-source components, $P_{dp}^D$ and $P_{dp}^R$:

$$P_{dp} = P_{dp}^D + P_{dp}^R,$$

where the components $P_{dp}^D$ and $P_{dp}^R$ manifest frequency dependencies in their phase consistent with their presumed mechanisms of generation. Specifically, the model predicts that when $P_{dp}$ is evoked using frequency-scaled stimuli (e.g., with the ratio $f_2/f_1$ fixed), the phase of $P_{dp}^D$ should be essentially independent of frequency whereas the phase of $P_{dp}^R$ should rotate rapidly.

#2 The putative reflection-source component, $P_{dp}^R$, of the total DPOAE closely matches the reflection emission measured at the same frequency under similar conditions. According to the taxonomy, stimulus-frequency emissions (SFOAEs) evoked at low stimulus levels are nearly pure reflection emissions (see Fig. 4-1).
Nonlinear Distortion

Linear Reflection

Figure 4-2: Schematic diagram illustrating the consequences of scaling for the phase of distortion- and reflection-source emissions. The left-hand panel shows a snapshot of the $f_2$ traveling wave at two different frequencies (top) and the corresponding phase lags (bottom) versus cochlear location. The two frequencies are denoted by black and gray lines, respectively. For simplicity, the $f_1$ traveling waves are not shown. Distortion sources result from nonlinear interaction between the primary traveling waves. The sources illustrated here ($\times$) are idealized as points at the peak of the $f_2$ traveling wave. When the frequency ratio $f_2/f_1$ is held fixed during the measurement sweep, the primary traveling waves (and thus the resulting distortion sources) simply shift along the cochlear partition, maintaining a nearly constant relative phase relationship as the stimulus frequencies are varied. Note, for example, that the phases of the primary traveling waves at the distortion source remain constant as the frequency is increased and the wave pattern shifts ($\leftarrow$) along the partition. As a result, the phases of all resulting distortion products are essentially independent of frequency. The right-hand panel shows a similar diagram for a reflection source (e.g., a perturbation in the mechanics of the cochlea). Since the perturbation ($\ast$) is fixed in space the phase of the wave scattered by the perturbation changes considerably ($\downarrow$) as the stimulus frequency is varied. Consequently, the phases of OAEs generated by linear reflection vary rapidly with frequency.
We thus test the prediction that

\[ P_{dp}^R \approx P_{sfe} \tag{4.2} \]

where \( P_{sfe} \) is the SFOAE at the same frequency. Once stimulus parameters have been adjusted to yield comparable overall emission levels, the predicted match includes the frequency dependence of both the amplitude (or spectral shape) and the phase.\(^3\)

Testing these predictions requires a technique for unmixing the total DPOAE into putative distortion- and reflection-source components. Initially, we adopt an experimental approach based on selective suppression that exploits the spatial separation of the presumed distortion- and reflection-source regions within the cochlea. To explore the sensitivity of our results to variations in the methodology of unmixing, we compare our results obtained using suppression to an alternative unmixing procedure based on spectral smoothing or time-domain windowing. A preliminary account of this work has been presented elsewhere (Kalluri and Shera, 2000).

### 4.2 Unmixing Via Selective Suppression

Reference to Fig. 4-1 suggests that one can separate the two components, \( P_{dp}^D \) and \( P_{dp}^R \), of the total DPOAE pressure if the reflection-source emission originating from the \( R \) region can be eliminated. The unmixing procedure would then be to (1) measure the total emission, \( P_{dp} \), using frequency-scaled stimuli; (2) eliminate the \( R \) component and remeasure the DPOAE to obtain the pure distortion-source component, \( P_{dp}^D \); and (3) compute the reflection-source component, \( P_{dp}^R \), by subtraction, \( P_{dp}^R = P_{dp} - P_{dp}^D \).

The spatial separation of the two source regions within the cochlea suggests trying to eliminate the \( R \) component by introducing a third, suppressor tone at a nearby frequency. The suppressor would act by reducing the amplitude of the wavelets incident upon and/or scattered back from the \( R \) region. Suppression techniques for separating OAE sources originating at different spatial locations in the cochlea were pioneered by Kemp and Brown (1983) and later refined by others (e.g., Heitmann et al., 1998; Siegel et al., 1998). The selective suppression strategy for unmixing yields the following estimates of \( P_{dp} \) and its components:

\[ P_{dp} = P_{ec}(f_{dp}) \text{ (measured at fixed } f_2/f_1) \tag{4.3} \]

\(^3\)Some differences (e.g., in phase) between \( P_{dp}^R \) and \( P_{sfe} \) are, of course, expected because the initial sources of forward-traveling cochlear waves at the emission frequency are at different spatial locations in the two cases (i.e., at the distortion-source region, \( D \), for \( P_{dp}^R \), and at the stapes for \( P_{sfe} \)).
In these expressions, $P_{ec}(f)$ denotes the complex ear-canal pressure at frequency $f$ resulting from stimulation at primary frequencies $f_1$ and $f_2$.

### 4.2.1 Measurement methods

We measured emissions from one ear of each of four normal hearing humans. Treatment of human subjects was in accordance with protocols approved by the Human-Studies Committee at the Massachusetts Eye and Ear Infirmary. All measurements were performed with subjects reclining comfortably in a sound-proofed, vibration-isolated chamber (Ver et al., 1975). Stimulus waveforms were generated and responses acquired and averaged digitally using a custom-built data-acquisition system. Acoustic signals were transduced using a Etymotic Research ER-10c DPOAE probe system supplemented with an ER-3A earphone whose sound-delivery tube was threaded through the ER-10c foam ear-tip. In situ earphone calibrations were performed at regular intervals throughout all measurement sessions. The calibrations were used to guarantee that the stimulus tones had constant level and zero starting phase in the ear canal at all frequencies. Real-time artifact rejection was implemented by comparing the time waveforms in successive data buffers before adding them to the averaged responses. In these and other respects, the methods and equipment used to obtain both SFOAEs and DPOAEs are generally similar to those described elsewhere (Shera and Guinan, 1999). We briefly summarize relevant differences here and provide detailed descriptions in the Appendix.

#### Measurement of DPOAEs

We measured distortion-product emissions at the frequency $2f_1 - f_2$ using frequency-scaled stimuli (i.e., using frequency sweeps performed with the primary-frequency ratio, $f_2/f_1$, held constant). The measurements reported here were obtained using primary levels of $\{L_1, L_2\} = \{60, 45\}$ dB SPL at the frequency ratio $f_2/f_1 = 1.2$. To ensure that our ability to maintain a constant $f_2/f_1$ ratio during the sweep was not systematically compromised by the frequency quantization imposed by digital stimulus generation, we modified our data-acquisition system to allow the sampling frequency to vary between measurement points. This flexibility enabled us to choose $f_1$ and $f_2$ so that the ratio $f_2/f_1$ varied by less than a thousandth of a percent between measurements (at our typical frequency spacing of about 15 Hz). The resulting sampling frequencies varied by less than 3% about the nominal value (59.94 kHz).
To allow any multiple reflections that might be present within the cochlea to settle into an approximately steady-state response, we measured $P_{dp}$ and $P_{dp|\text{suppressed}}$ over time intervals ($\approx 136$ ms) much longer than the estimated round-trip travel time for cochlear waves ($\approx 10-15$ ms). To guard against possible systematic variations in emission amplitude over time that might invalidate the unmixing procedure (e.g., due to efferent feedback or to changes in earphone calibration caused by subject movement or by temperature variations), we interleaved measurements of $P_{dp}$ and $P_{dp|\text{suppressed}}$ in time and averaged multiple repetitions (typically $n = 64$). We set the suppressor frequency approximately 44 Hz below the distortion-product frequency (e.g., Siegel et al., 1998; Dreisbach and Siegel, 1999). The suppressor level was adjusted (separately for each subject) to minimize the DPOAE fine structure while leaving the mean DPOAE amplitude (as averaged over several fine-structure periods) largely unchanged (cf., Heitmann et al., 1998). The suppressor level chosen in this way was generally in the range 50–55 dB SPL.

### Measurement of SFOAEs

We measured stimulus-frequency emissions using the suppression method (e.g., Guinan, 1990; Kemp et al., 1990; Brass and Kemp, 1991, 1993; Souter, 1995; Shera and Guinan, 1999). In this method, the emission is obtained as the complex (or vector) difference between the ear-canal pressure at the probe frequency measured first with the probe tone alone and then in the presence of a stronger suppressor tone at a nearby frequency. Thus, the SFOAE pressure at the probe frequency, $P_{sfe}(f_p)$, is defined as

$$P_{sfe}(f_p) \equiv P_{ec}(f_p) - P_{ec}(f_p)_{\text{with suppressor at } f_s \neq f_p}.$$  \hspace{1cm} (4.6)

In the measurements reported here, the suppressor frequency was approximately 44 Hz below the probe ($f_s \approx f_p - 44$ Hz). To prevent contamination from the considerable cross-talk between output channels of the ER-10c, we generated the suppressor tone using a separate ER-3A earphone whose sound-delivery tube was threaded through the foam eartip. Unless otherwise noted, the probe and suppressor levels, $L_p$ and $L_s$, were 40 and 55 dB SPL, respectively (Shera and Guinan, 1999). Exploratory measurements at other nearby probe levels indicate that the spectral shape and phase of $P_{sfe}$ are not strong functions of intensity at these levels. We found that probe levels of 40 dB SPL gave emission levels generally comparable to those of $P_{dp}$, especially after introduction of the primary-mimicking tone described below. As with the DPOAE measurements, we interleaved measurements of $P_{ec}(f_p)$ and $P_{ec}(f_p)_{\text{suppressed}}$ in time to minimize artifacts that might contaminate the difference. The measurement frequency resolution, approximately 15 Hz between points, was always sufficient.
to prevent ambiguities in phase unwrapping.

In some experiments, we measured SFOAEs in the presence of an additional continuous tone. The idea was to measure $P_{sfe}$ under conditions matching as closely as possible those present during the measurement of $P_{dp}$. Thus, we introduced the additional tone at a frequency and level corresponding to the $f_1$ primary used in the measurement of DPOAEs. In terms of the probe frequency, the frequency of this additional tone (the ”$f_1$-primary mimicker”) was therefore given by $f_1 = f_p/(2 - r)$, where $r$ denotes the $f_2/f_1$ ratio we wished to mimic. We denote the SFOAE measured in the presence of the $f_1$-primary mimicking tone by $P_{sfe}^1$:

$$P_{sfe}^1 \equiv P_{sfe} \bigg|_{\text{with } f_1 \text{ mimicker}}.$$

### 4.2.2 Detailed Measurement Methods

This section describes in more detail the methods used to obtain the emission measurements reported here. A general reader can skip ahead to the results without a loss in continuity.

**Measurement of DPOAEs**

Distortion-product otoacoustic emissions at the frequency $2f_1 - f_2$ were measured using frequency-scaled stimuli (i.e., with the ratio $f_2/f_1$ held constant). At each measurement frequency the acoustic stimulus had the form

$$\text{stimulus} = <XX\cdots X>_{#\geq M}$$

where $X$ represents a periodic $(5 \times 4096)$-sample (≈ 342 ms$^4$) segment consisting of three components:

$$X = \begin{cases} 
\pi_1^1\Pi_2^1\Pi_3^1\pi_4^1\pi_5^1\Pi_6^1\Pi_7^1\pi_8^1 & \text{(primary earphone \#1)} \\
\pi_1^2\Pi_2^2\Pi_3^2\pi_4^2\pi_5^2\Pi_6^2\Pi_7^2\pi_8^2 & \text{(primary earphone \#2)} \\
o_1O_2O_3<_{\sigma_5}\Sigma_6\Sigma_7>_{>8} & \text{(suppressor earphone)}
\end{cases} \text{ (4.9)}$$

Each component consisted of four long intervals (uppercase) and four short intervals (lowercase and angled brackets). The long intervals were each 4096-samples (≈ 68 ms) in duration. The primary segments, $\Pi_1^1$ and $\Pi_1^2$, contained an integral number of

---

$^4$Because we varied our sampling rate between measurement points, corresponding stimulus durations varied by up to ±3%.
periods of the primary frequencies, $f_1$ and $f_2$, respectively. The suppressor segments, $\Sigma_i$, contained an integral number of cycles of the suppressor frequency, $f_s$. The zero segments, $O_i$, were identically zero throughout. Waveform phases were adjusted, using information from the calibration procedure, so that each stimulus had zero (cosine) phase in the ear canal at the beginning of segment $\Pi_2$. The short intervals were one fourth the duration of the long intervals (i.e., 1024-samples or $\approx 17$ ms) and did not, in general, contain an integral number of periods of the corresponding waveform. The short intervals $\pi_i$, $\sigma_i$, and $\sigma_1$ allowed for response settling time and contained segments of the primary, suppressor, or zero waveforms, respectively. The short intervals $\{<4, >8\}$ were used to ramp the suppressor tone {on,off} using the {first,second} half of the Blackman window. The three components of $X$ were synchronized and presented simultaneously through three separate earphones. Note that whereas the primary tones played continuously during the measurement, the suppressor tone cycled on and off repeatedly due to alternation of the zero and suppressor waveforms. Interleaving the measurements of $P_{dp}$ and $P_{dp}$ |suppressed in this way helps to minimize possible artifacts due to systematic variations over time (e.g., due to subject movement, drifts in earphone calibration, efferent feedback, etc). Unless otherwise noted, the primary levels $\{L_1, L_2\}$ were $\{60, 45\}$ dB SPL, respectively. Primary levels were chosen in approximate accordance with the formula $L_1 \approx 0.4L_2 + 39$ dB SPL, which tracks the “ridge” in the $L_1$-$L_2$ plane that maximizes the $2f_1 - f_2$ emission for $f_2/f_1 \approx 1.2$ (Kummer et al., 1998).

Measurements were made versus probe frequency by sweeping the primaries and suppressor from high frequencies to low, with $f_s = f_{dp} + \Delta f_s$ and $\Delta f_s = -44$ Hz. The periodic segments $X$ were played repeatedly until a total of $M$ corresponding artifact-free responses were collected. In these measurements, $M$ was typically 64 so that at each frequency the total stimulus duration was therefore $\gtrsim 64 \times 342$ ms $\approx 22$ s. To reduce unwanted transients the probe waveform was ramped on and off by pre- and postpending two additional segments [indicated by the angled brackets $<$ and $>$ in Eq. (4.8)] with envelopes of half Blackman windows with 2.5-ms rise and fall times. After digitizing the resulting ear-canal pressure, responses to all primary-alone segments (i.e., all segments $\Pi_2$ and $\Pi_3$) were averaged to form $Y_p$; similarly, the responses to all probe+suppressor segments (i.e., all segments $\Pi_6$ and $\Pi_7$) were averaged to form $Y_{p+s}$. From these averaged response waveforms, the complex amplitudes of the $f_{dp}$ components of the ear-canal pressure, denoted $P_{dp} = P_{ec}(f_{dp})$ and $P_{dp} = P_{ec}(f_{dp})e^{-2\pi i\Delta N\Delta T}f_{dp}$ |suppressed, were extracted using Fourier analysis. The complex exponential compensates for the phase shift in the probe due to the time interval, $\Delta N\Delta T$, between the primary-alone and primary+suppressor segments. Here, $\Delta T$ is the sampling interval (reciprocal of the sampling rate), and $\Delta N$ represents the
total number of these intervals that separate the two segments:

$$\Delta N = \# \text{samples}(\Pi_2 \Pi_3 \pi_4 \pi_5) = 2^{1/2} \times 4096 = 10240 . \quad (4.10)$$

Note that when the two segments are separated by an integral number of periods of the \( f_{dp} \) waveform, the phase shift modulo \( 2\pi \) is zero. The complex quantity \( P_{dp}^{R}(f_{dp}) \) was then obtained as

$$P_{dp}^{R} = P_{dp} - P_{dp}^{D} . \quad (4.11)$$

**Measurement of SFOAEs**

Stimulus-frequency emissions were measured using the suppression method detailed elsewhere (Shera and Guinan, 1999). In some experiments, we measured SFOAEs in the presence of an additional continuous tone (the “\( f_1 \)-primary mimicker”) at a frequency and level corresponding to the \( f_1 \) primary in the measurement of DPOAEs detailed above.

At each measurement frequency the acoustic stimulus had the form given by Eq. 4.8, with \( X \) representing a periodic \((5 \times 4096)\)-sample (\( \approx 342 \text{ms} \)) segment consisting of three components:

$$X = \begin{cases} 
\pi_1 \Pi_2 \Pi_3 \pi_4 \pi_5 \Pi_6 \Pi_7 \pi_8 & \text{(probe earphone)} \\
o_1 O_2 O_3 <4> \sigma_5 \Sigma_5 \Sigma_7 > & \text{(suppressor earphone)} \\
\mu_1 M_2 \mu_3 \mu_4 \mu_5 M_6 M_7 \mu_8 & \text{(primary-mimicker earphone)} 
\end{cases} . \quad (4.12)$$

Each component consisted of four long (uppercase) and four short (lowercase and angled brackets) intervals. The long intervals were each 4096-samples (\( \approx 68 \text{ms} \)) in duration and contained an integral number of periods of the probe (\( \Pi_i \)), suppressor (\( \Sigma_i \)), zero (\( O_i \)), or primary mimicker (\( M_i \)) waveforms, respectively. The phase of the probe waveform was adjusted, using information from the calibration procedure, so that the stimulus had zero (cosine) phase in the ear canal at the beginning of segment \( \Pi_2 \). The short intervals were one fourth the duration of the long intervals (i.e., 1024-samples or \( \approx 17 \text{ms} \)) and did not, in general, contain an integral number of periods of the corresponding waveform. The short intervals \( \pi_i, \sigma_i, o_1, \) and \( \mu_i \) allowed for response settling time and contained segments of the probe, suppressor, zero, and mimicker waveforms, respectively. The short intervals \( \{<4, >\} \) were used to ramp the suppressor tone \{on,off\} using the \{first,second\} half of the Blackman window. The three components of \( X \) were synchronized and presented simultaneously through three separate earphones. Note that whereas the probe and primary mimicker tones played continuously during the measurement, the suppressor tone cycled on and off repeatedly due to alternation of the zero and suppressor waveforms. The probe and
suppressor levels \( \{L_p, L_s\} \) were generally \( \{40, 55\} \) dB SPL. The primary mimicker was presented at a frequency and level corresponding to the \( f_1 \) primary in the measurement of DPOAEs (i.e., at a frequency equal to \( f_1 = f_p/(2 - r) \), where \( r \) denotes the \( f_2/f_1 \) ratio we wished to mimic, and a typical level of 60 dB SPL).

Other features of the stimulus paradigm and the subsequent data analysis used to compute \( P_{sfe}^1 \) are analogous to the measurement of DPOAEs detailed above and have been described elsewhere (Shera and Guinan, 1999).

### 4.2.3 Results: Unmixing via selective suppression

**\( P_{dp} \) and its components, \( P_{dp}^D \) and \( P_{dp}^R \)**

Typical measurements of the total DPOAE and its components estimated using suppression are shown in Fig. 4-3. To illustrate the variation across subjects, we show results for three of our four subjects (those for whom the most data are available); similar results were obtained in the fourth subject. In each case, the putative distortion-source component, \( P_{dp}^D \), is essentially a smoother version of the total DPOAE in which much of the quasi-periodic fine structure apparent in both the amplitude and phase of \( P_{dp} \) has been eliminated. In agreement with Prediction #1, the phase of \( P_{dp}^D \) is nearly constant, varying by less than a period. By contrast, the phase of the reflection-source component, \( P_{dp}^R \), falls through many cycles (typically eight or more) over the same frequency range. These different frequency dependencies imply generation by fundamentally different mechanisms: The nearly constant phase of \( P_{dp}^D \) is consistent with generation by frequency-scaled nonlinear distortion and the rapidly rotating phase of \( P_{dp}^R \) with generation by linear coherent reflection (Shera and Guinan, 1999).

The fine structure manifest in the total DPOAE arises because of alternating constructive and destructive interference between the two components, \( P_{dp}^D \) and \( P_{dp}^R \), caused by the systematic rotation of their relative phase, a consequence of the very different slopes of their phase vs. frequency functions. Thus, the components \( P_{dp}^D \) and \( P_{dp}^R \) “beat” against each other, producing an oscillatory interference pattern. In other words, DPOAE fine structure arises, fundamentally, because DPOAEs are mixtures of emissions with distinctly different properties that reflect their different mechanisms of generation.

**Comparison between \( P_{dp}^R \) and \( P_{sfe} \)**

According to Prediction #2, the reflection-source component of the total DPOAE, \( P_{dp}^R \) (Fig. 1, lower panel), should closely match other reflection-source emissions measured under comparable conditions (e.g., SFOAEs at low stimulus levels as in Fig. 1, upper panel). Figure 4-4 tests this prediction by comparing \( P_{dp}^R \) and \( P_{sfe} \) measured in the
Figure 4-3: The DPOAE $P_{dp}$ and its estimated distortion- and reflection-source components, $P^D_{dp}$ and $P^R_{dp}$, obtained using suppression. The figure shows typical measurements of the amplitude (top) and phase (bottom) of the $2f_1 - f_2$ DPOAE and its components measured using a frequency-scaled stimulus (i.e., the primary frequencies $f_1$ and $f_2$ were swept with their ratio held constant at the value $f_2/f_1 = 1.2$). Left to right, the panels show data from subjects #1, #2, and #3, respectively; similar results were obtained in the fourth subject. In each case, the total DPOAE (solid line) was unmixed using a suppressor tone near the distortion-product frequency, $f_{dp}$. Although the phases of the estimated $P^D_{dp}$ components (dotted lines) vary less than a period, the phases of the estimated $P^R_{dp}$ components (dashed lines) fall through many cycles over the same frequency range, in agreement with Prediction #1. The measurement noise floor was approximately $-25$ dB SPL and the frequency resolution was always sufficient to prevent ambiguities in phase unwrapping. Stimulus levels for subject #1: $\{L_1, L_2, L_s\} = \{60, 45, 50\}$ dB SPL. Stimulus levels for subject #2: $\{L_1, L_2, L_s\} = \{60, 45, 50\}$ dB SPL. Stimulus levels for subject #3: $\{L_1, L_2, L_s\} = \{60, 45, 55\}$ dB SPL.
same ear. In agreement with predictions, the phase slopes of $P_{sfe}$ and $P_{dp}^R$ are nearly identical. In addition, both $P_{sfe}$ and $P_{dp}^R$ have similar amplitude features (e.g., a deep notch near 1.4 kHz). These similarities support the idea that $P_{sfe}$ and $P_{dp}^R$ are generated by a similar mechanism. Note that deep spectral notches such as that apparent near 1.4 kHz are predicted by the theory of coherent reflection filtering (cf. Fig. 11 of Zweig and Shera, 1995). In the model, such notches arise from random spatial fluctuations in the irregularities that scatter the wave. At some frequencies, wavelets scattered from different locations within the scattering region combine nearly out of phase and cause near cancellation of the net reflected wave.

Although the overall match between $P_{dp}^R$ and $P_{sfe}$ is good—especially when one considers the substantial differences in the way that the two emissions are evoked and measured—details of the spectral shape (e.g., the precise location of the notch) do not match perfectly. Do these discrepancies suggest important differences between $P_{dp}^R$ and $P_{sfe}$ and their mechanisms of generation? Or do they reflect differences in measurement conditions that influence the magnitude and/or phase of the traveling-wave energy scattered back from R? For example, the primaries present during the DPOAE measurement may suppress the traveling wave near the $f_{dp}$ place, thereby affecting the frequency dependence of $P_{dp}^R$.

**Mimicking suppression by the primaries**

To address these questions, we modified our $P_{sfe}$-measurement paradigm to better mimic the intracochlear conditions under which $P_{dp}^R$ originated. Specifically, we measured $P_{sfe}$ in the presence of an additional tone whose frequency and level were chosen to match those of the $f_1$ primary used during the measurement of $P_{dp}^R$ (see Methods above). We mimic the $f_1$ primary because we expect it to have the greater effect; the $f_1$ primary is both closer in frequency to $f_{dp}$ and higher in level than the $f_2$ primary. We define $P_{sfe}^1$ as the value of $P_{sfe}$ measured in the presence of the $f_1$-primary mimicker.

Measurements of $P_{sfe}^1$ are shown and compared to those of $P_{dp}^R$ in Fig. 4-5. The match between the two putative reflection-source emissions is now much closer. This result is consistent with the idea that the differences in Fig. 4-4 reflect differing intracochlear stimulus conditions; differences in the mechanisms of emission generation are not required. Thus, the similarity in both magnitude and phase between $P_{dp}^R$ and $P_{sfe}^1$ is in agreement with Prediction #2 and provides strong support for the model. Note that the changes in the overall amplitude and spectral shape of $P_{sfe}$ caused by the $f_1$-primary mimicker suggest that the primaries have a significant effect on the reflection-source component of the DPOAE (presumably via suppression of the wave incident upon and/or reflected back from the R region).
**Figure 4-4**: Comparison between the estimated reflection-source component, $P_{dp}^R$, and the SFOAE, $P_{sfe}$. The figure shows the amplitude (top) and phase (bottom) of $P_{dp}^R$ (dashed line), the reflection-source component of the total DPOAE obtained in Fig. 4-3 for subject #1. Shown for comparison is $P_{sfe}$ (solid line), the SFOAE measured in the same subject at a probe level of 40 dB SPL. Note the considerable agreement in both amplitude and phase (e.g., $P_{dp}^R$ and $P_{sfe}$ have similar amplitude notches and phase slopes). SFOAE stimulus levels: $\{L_p, L_s\} = \{40, 50\}$ dB SPL.
Figure 4-5: Comparison between the estimated reflection-source component, $P_{dp}^R$, and the $f_1$-mimicked SFOAE, $P_{sfe}^1$. The figure compares the amplitude (top) and phase (bottom) of $P_{dp}^R$ (dashed line, from Fig. 4-3) and $P_{sfe}^1$ (black solid line), the value of $P_{sfe}$ measured in the presence of an additional tone at the frequency and level of the $f_1$ primary present during the measurement of $P_{dp}^R$. Left to right, the three panels show data for subjects #1, #2, and #3, respectively. Shown for comparison is $P_{sfe}$ (gray line). The match between the amplitude and phase of $P_{dp}$ and $P_{sfe}^1$ is generally excellent, in agreement with Prediction #2. The differences between $P_{sfe}$ and $P_{sfe}^1$ caused by the mimicker suggest that the primaries have a significant effect on the reflection-source component of the DPOAE. SFOAE stimulus levels: $\{L_p, L_2, L_1\} = \{40, 55, 60\}$ dB SPL.
4.3 Unmixing Via Spectral Smoothing

A potential difficulty with suppression-based unmixing is that the suppressor tone, introduced with the intent of selectively suppressing the reflection-source component, may inadvertently modify the response in other ways. For example, the suppressor tone may also suppress the distortion-source component (either directly, or through its effects on the primaries) or “catalyze” the generation of additional distortion-sources at the frequency $f_{dp}$ (Fahey et al., 2000). As a test of these possibilities, and to investigate the sensitivity of our conclusions to the method of unmixing, we repeated our analysis using a completely different method. This method—spectral smoothing (or its equivalent, time windowing)—was suggested by the correspondence, in a linear system, between phase slope in the frequency domain and latency in the time domain (e.g., Papoulis, 1962). As unmixed by suppression, the two components $P^D_{dp}$ and $P^R_{dp}$ of $P_{dp}$ have very different phase slopes, evidently reflecting fundamental differences in their mechanisms of generation. Consequently, if we apply Fourier analysis to our frequency-domain measurements of $P_{dp}$, we expect to see two components of very different latencies in the corresponding “latency-domain response”: namely, a short-latency component corresponding to $P^D_{dp}$ and a long-latency component corresponding to $P^R_{dp}$. Thus, our suppression results suggest that the two components of $P_{dp}$ should be separable using signal-processing strategies based on appropriate windowing in the latency domain. Techniques for analyzing OAEs in this way were introduced by Shera and Zweig (1993a; Zweig and Shera, 1995), who applied them to the study of SFOAEs; similar methods have since been applied to other emissions (e.g., L.J. Stover et al., 1996; Brown et al., 1996; Fahey and Allen, 1997; Knight and Kemp, 2000b; Ren et al., 2000).

Multiplication by a window in the latency domain corresponds to convolution with a smoothing function in the frequency domain. Although the two approaches are entirely equivalent, we refer to the technique as “spectral smoothing” rather than “time windowing” because viewing the process in the frequency domain yields equations for the components that are more directly analogous to those of the suppression method (cf., Eqs. 4.3–4.5). The spectral smoothing strategy for unmixing thus yields the following estimates of $P_{dp}$ and its components:

\[ P_{dp} = P_{ec}(f_{dp}) \quad \text{(measured at fixed } f_2/f_1) \]; \hspace{1cm} (4.13)

\[ P^D_{dp} \approx P_{dp} \bigg|_{\text{smoothed}} \quad \text{(convolved with smoothing filter)} \]; \hspace{1cm} (4.14)

---

\[ \text{We put “latency-domain response” in quotes because the signal we obtain by Fourier transforming the frequency response does not correspond with the time-domain impulse response of the system.} \]
4.3.1 Analysis methods

Measurements of transient-evoked and stimulus-frequency emissions indicate that reflection-emission latency varies with frequency (e.g., Kemp, 1978; Wilson, 1980; Norton and Neely, 1987; Neely et al., 1988; Shera and Guinan, 2000a). This frequency dispersion tends to smear out the reflection-source component in time, making it more difficult to separate by windowing. To help compensate for this dispersion, it proves helpful to work in the log-frequency domain. Consequently, we perform Fourier transforms with respect to the dimensionless frequency variable

\[ \nu \equiv -\log(f/f_{\text{ref}}), \]  \hspace{1cm} (4.16)

where \( f_{\text{ref}} \) is a reference frequency taken, for convenience, as the maximum frequency of hearing. Fourier transformation with respect to a log-frequency variable, suggested by the approximate local scaling symmetry of cochlear mechanics, results in sharper, more well-defined peaks in the Fourier-conjugate latency domain (Zweig and Shera, 1995; Knight and Kemp, 2000b). 7 The conjugate dimensionless latency variable, here denoted \( \tau \), represents emission latency expressed in periods of the emission frequency (Zweig and Shera, 1995). 8

Unmixing by smoothing involves convolving \( P_{dp} \) with a smoothing function, \( S \), of

\[ P_{dp}^{R} \cong P_{dp} - P_{dp}^{D}. \]  \hspace{1cm} (4.15)
finite bandwidth (e.g., a Gaussian): 9

\[ P_{dp} \bigg|_{\text{smoothed}} \equiv S \ast P_{dp}, \quad (4.17) \]

where * denotes the operation of convolution. The convolution is equivalent to a multiplication (or windowing) in the \( \tau \) domain. Thus,

\[ P_{dp} \bigg|_{\text{smoothed}} = F^{-1}\{ \hat{S} \times F\{P_{dp}\} \}, \quad (4.18) \]

where \( F\{\cdot\} \) represents the operation of Fourier transformation (with respect to \( \nu \)), \( F^{-1}\{\cdot\} \) the inverse transformation (with respect to \( \tau \)), and the window, \( \hat{S}(\tau) \), is the Fourier transform of \( S \):

\[ \hat{S} \equiv F\{S\}. \quad (4.19) \]

Separation of \( P_{dp} \) into meaningful components requires choosing the smoothing function (or, equivalently, the shape and duration of the latency window) appropriately. Ideally, the window \( \hat{S}(\tau) \) should have a sharp cutoff in the \( \tau \) domain—to cleanly separate emission components of different latencies—but avoid extensive spreading (or "ringing") in the frequency response (smoothing function). To approximate these desired characteristics we employ one of a class of “recursive exponential filters” (Shera and Zweig, 1993a). 11 The recursive-exponential filters are entire functions and have

9 Note that unlike the more familiar case of time-domain filtering, the oscillatory function to be removed occurs here in the frequency response. In its reversal of the roles usually played by time and frequency, the technique used here is similar to cepstral analysis (Bogert et al., 1963), although we work with a log-frequency variable, \( \nu \), and analyze \( P_{dp} \) rather than \( \log(P_{dp}) \). [In cepstral analysis, one takes the logarithm of the frequency response in order to decompose a presumed product of spectra into a sum. In our application, the pressure \( P_{dp} \) is represented directly as a sum of components (Prediction #1); taking the logarithm is therefore both unnecessary and undesirable.]

10 To perform our transforms numerically, we resampled our measurements of \( P_{dp} \) at equal intervals in log frequency using cubic spline interpolation. Because our sampling rate was variable, our measurements of \( P_{dp} \) were not equally spaced in linear frequency.

11 The \( n \)-th order recursive-exponential filtering window is defined by (Shera and Zweig, 1993a):

\[ \hat{S}_n(\tau; \tau_{cut}) \equiv 1/\Gamma_n(\lambda_n \tau/\tau_{cut}), \]

where the parameter \( \tau_{cut} \) is the cutoff time (length of the window) and the function \( \Gamma_n(\tau) \) is defined recursively:

\[ \Gamma_{n+1}(\tau) = e^{\Gamma_n(\tau)-1}, \quad \text{with} \quad \Gamma_1(\tau) = e^{\tau^2}. \]

The window \( \hat{S}_n(\tau; \tau_{cut}) \) has a maximum value of 1 at \( \tau = 0 \); the scale factor \( \lambda_n \) is chosen so that the window falls to the value \( 1/e \) at \( \tau = \tau_{cut} \):

\[ \lambda_n = \sqrt{\gamma_n}, \quad \text{where} \quad \gamma_{n+1} = \ln(\gamma_n + 1) \quad \text{with} \quad \gamma_1 = 1. \]
no poles, discontinuities, or other undesirable features in the complex plane to contribute large oscillations to the smoothing function.

In practice, measurements are only available over a finite frequency range, and the smoothing operation is complicated by end effects. Throughout this paper, the analyzed frequency range was chosen to include an approximately integral number of spectral cycles and smoothing was performed using periodic boundary conditions (the data were effectively wrapped around a cylinder). When necessary, a linear ramp was subtracted, and subsequently restored after smoothing, to remove any discontinuity at the “seam.” The estimate of \( P_{dp}^R \) so obtained was then discarded at each end over a frequency interval equal to the approximate bandwidth of the smoothing function.\(^{12}\)

**Determining the window duration**

Unmixing via windowing (spectral smoothing) requires specification of the duration of the time window (bandwidth of the smoothing function) used to separate components with different latencies. The suppression studies reported above indicate that the long-latency component, \( P_{dp}^R \), closely matches the characteristics of reflection emissions measured under comparable conditions (e.g., \( P_{sfe}^1 \)). Consequently, an estimate of the appropriate window duration can be obtained from measurements of SFOAEs evoked at low stimulus levels. Analysis of such measurements indicates that in the 1–2 kHz range, reflection emissions are delayed by an average of about 15 periods of the stimulus frequency with a spread of roughly ±35% (Zweig and Shera, 1995; Shera and Guinan, 2000a). Multiplication by a window of duration \( \tau_{cut} = 8–9 \) periods might therefore be expected to cleanly remove reflection-source components in this frequency range.

Figure 4-6 corroborates this analysis using our measurements of \( P_{dp} \) and \( P_{sfe}^1 \). Both short- and long-latency components are clearly apparent in the Fourier transform \( F\{P_{dp}\} \), the latency-domain representation of \( P_{dp} \). [By contrast, the long-latency component is almost entirely absent in the Fourier transform of \( P_{dp}^R \) obtained by suppression (not shown).] As expected, the long-latency component in \( F\{P_{dp}\} \), centered at a latency of about 15 cycles, coincides with the peak in \( F\{P_{sfe}^1\} \). The 10th-order recursive exponential filter, \( \hat{S}_{10}(\tau; \tau_{cut}) \), with a cutoff latency of \( \tau_{cut} = 9 \) periods is

\[ \Delta \nu = \Delta f/f \geq 1/\pi \tau_{cut} \]

Note that the 1st-order filtering window is a simple Gaussian; in the limit \( n \to \infty \), \( \hat{S}_n \) approaches a rectangular (or boxcar) window. For intermediate \( n \) (e.g., the value \( n = 10 \) used here), \( \hat{S}_n \) has a much sharper cutoff than standard windowing functions (e.g., the Hamming, Blackman, etc.) and considerably less “ringing” in the smoothing function than the simple boxcar.

\(^{12}\) The smoothing function has approximate width (Shera and Zweig, 1993a)
shown for comparison. In subsequent analysis, we use \( \hat{S}_{10}(\tau; \tau_{\text{cut}}=9) \) to separate the short- and long-latency components of \( P_{dp} \).

### 4.3.2 Results: Unmixing via spectral smoothing

**\( P_{dp} \) and its components, \( P_{dp}^D \) and \( P_{dp}^R \), revisited**

Typical measurements of \( P_{dp} \) and its components unmixed by spectral smoothing are shown with the components obtained by suppression in Fig. 4-7. Qualitatively, the two methods unmix \( P_{dp} \) into similar components. For example, the estimates of \( P_{dp}^R \) obtained by the two methods have nearly identical phases and manifest similar frequency dependence in their amplitude curves. There are, of course, differences in the details. For example, the distortion-source components, \( P_{dp}^D \), obtained by suppression unmixing have larger fine structure than the same components obtained by smoothing. We examine this issue further in the next section. Despite differences in detail, the qualitative agreement between the estimated components indicates that our tests of the two-mechanism model are not especially sensitive to the method of unmixing.

### 4.4 Errors due to Incomplete Unmixing

We explore in Fig. 4-8 the effects of varying key parameters in each of our two unmixing paradigms. For unmixing by suppression, the top panels show how estimates of \( P_{dp}^D \) and \( P_{dp}^R \) depend on suppressor level, \( L_s \); for unmixing by smoothing (time windowing), the bottom panels show the dependence on the duration of the latency window, \( \tau_{\text{cut}} \) (or, equivalently, the bandwidth of the smoothing function). Note how the fine-structure oscillations in \( P_{dp}^D \) (left) increase toward the bottom of each plot (i.e., at lower values of \( L_s \) or longer \( \tau_{\text{cut}} \)). By contrast, the fine-structure oscillations in \( P_{dp}^R \) (right) increase towards the top (i.e., at higher values of \( L_s \) or shorter \( \tau_{\text{cut}} \)).

These systematic trends can be understood using a simple model of the unmixing process. Let the model pressure \( P_{dp} \) be the sum of two components, \( D \) and \( R \), with very different phase slopes. As a consequence of this difference, \( D \) and \( R \) beat against each other, producing an oscillatory interference pattern in the amplitude and phase of \( P_{dp} \). Imagine now that we attempt to unmix the components experimentally; let our estimates of the two components be denoted \( P_{dp}^D \) and \( P_{dp}^R \), respectively. Perfect unmixing would yield \( P_{dp}^D = D \) and \( P_{dp}^R = R \). In general, however, unmixing is incomplete, and the estimates contain contributions from both \( D \) and \( R \):

\[
\begin{pmatrix}
P_{dp}^D \\
P_{dp}^R
\end{pmatrix}
= 
\begin{pmatrix}
1 - \delta & \rho \\
\delta & 1 - \rho
\end{pmatrix}
\begin{pmatrix}
D \\
R
\end{pmatrix},
\tag{4.20}
\]
Figure 4-6: The smoothing function and corresponding latency window. The figure shows both latency-domain (top) and corresponding frequency-domain representations (bottom) of $P_{dp}$ and the matched smoothing function, $S$. The top panel shows the amplitudes of the Fourier transforms $F\{P_{dp}\}$ (solid gray line) and $F\{P^1_{sfe}\}$ (dashed gray line) versus $T$, the emission latency expressed in stimulus periods. The $\tau$-domain representations of both $P_{dp}$ and $P^1_{sfe}$ show a strong peak centered at a latency of about 15 cycles. Shown for comparison (solid black line) is the 10th-order recursive exponential filter, $S_{10}(\tau; \tau_{cut})$, with a cutoff time of $\tau_{cut} = 9$ used in subsequent analysis to separate short- and long-latency components of $P_{dp}$. The bottom panel shows frequency-domain representations of $P_{dp}$ and $S$. Note that the frequency axis is logarithmic (linear in $\nu$). The real and imaginary parts of $P_{dp}$ are shown with the solid and dotted gray lines, respectively. A linear ramp has been subtracted from $P_{dp}$ to render the function periodic on a cylinder. The black line shows the smoothing function $S_{10}(\nu; 9)$ which, when convolved with $P_{dp}$, yields our estimate of $P_{dp}^D$. Note that the vertical scale for $S$, dependent on the number of points in our numerical Fourier transform, is not especially illuminating and has been left unspecified. The measurements of $P_{dp}$ and $P^1_{sfe}$ are from Figs. 4-3 and 4-5, respectively (subject #1).
Figure 4-7: The DPOAE $P_{dp}$ and its estimated distortion- and reflection-source components, $P_{dp}^D$ and $P_{dp}^R$, obtained using spectral smoothing (time windowing). The figure shows the amplitude (top) and phase (bottom) of $P_{dp}$ from Fig. 4-3. Left to right, the three panels show data from subjects #1, #2, and #3, respectively. In each case, the total DPOAE (solid black line) was unmixed as described in the text using the 10th-order recursive exponential filter, $S_{10}(\tau; 9)$. The two components, $P_{dp}^D$ (dotted black line) and $P_{dp}^R$ (dashed black line), are qualitatively similar to those obtained using suppression (gray lines).
Figure 4-8: Changes in estimates of $|P^D_{dp}|$ and $|P^R_{dp}|$ with variations in the parameters of the unmixing process. The figure shows the levels $L^D_{dp}$ (left) and $L^R_{dp}$ (right) obtained when the parameters for the suppression- and smoothing-based unmixing are varied systematically. The top panels show the results obtained by varying the level of the suppressor tone, $L_s$. The bottom panels show the results of varying the duration of the latency window, $\tau_{cut}$. The original, unsmoothed measurement of $L_{dp}$ is shown for comparison in the bottom left ($\tau_{cut} = \infty$). Note that because the estimates of $P^D_{dp}$ and $P^R_{dp}$ were discarded at each end over a frequency interval equal to the bandwidth of the smoothing function (see Sec. 4.3.1), the estimates cover a more limited frequency range at smaller values of $\tau_{cut}$. In all panels, the different curves have been offset from one another for clarity. Unmixing parameters used earlier in the paper ($L_s = 55$ dB SPL and $\tau_{cut} = 9$ periods) are marked with an asterisk. Data from subject #3 with stimulus levels of $\{L_1, L_2\} = \{60, 45\}$ dB SPL.
where the complex, frequency-dependent coefficients $\delta$ and $\rho$ quantify the unmixing errors. Note that the coefficients satisfy the constraint $P_{dp}^D + P_{dp}^R = D + R$. Although perfect unmixing requires $\delta = \rho = 0$, acceptable results occur with $|\delta| \ll 1$ and $|\rho| \ll 1$.

The unmixing errors $\delta$ and $\rho$ depend on unmixing parameters such as the level of the suppressor and the duration of the latency window. To explicate the trends in Fig. 4-8, we consider three special cases of incomplete unmixing:

1. **Case $\delta = 0$ and $\rho \neq 0$** so that
   \[
   P_{dp}^D = D + \rho R ; \\
   P_{dp}^R = (1 - \rho)R .
   \]
   For suppression-based unmixing, this case results from using a weak suppressor that leaves $D$ unchanged but only incompletely removes the $R$ component from the mix (i.e., under-suppression); in the smoothing case, it corresponds to under-smoothing (i.e., to using too narrow a smoothing function or too long a latency window). Since the resulting estimate of $P_{dp}^D$ appears contaminated by the $R$ component, the magnitude $|P_{dp}^D|$ should oscillate with frequency. These features are found in the figure: At smaller values of $L_s$ and longer values of $\tau_{cut}$, the estimates $L_{dp}^D$ manifest considerable fine structure.

2. **Case $\delta \neq 0$ and $\rho = 0$** so that
   \[
   P_{dp}^D = (1 - \delta)D ; \\
   P_{dp}^R = R + \delta D .
   \]
   Here, the suppressor is strong enough to completely remove the $R$ component, but in so doing it modifies the $D$; for smoothing, this case results from over-smoothing (i.e., using an overly broad smoothing function or too short a latency window). In this case, the estimate $P_{dp}^R$ is contaminated with part of the $D$ component, and its magnitude should therefore oscillate. These features occur in Fig. 4-8: At the largest values of $L_s$ and shortest values of $\tau_{cut}$, the estimates $L_{dp}^R$ show evidence of fine structure.

3. **Case $\delta \neq 0$ and $\rho \neq 0$** so that
   \[
   P_{dp}^D = (1 - \delta)D + \rho R ; \\
   P_{dp}^R = (1 - \rho)R + \delta D .
   \]
   In this more general case, the suppressor is neither strong enough to eliminate
the $R$ component nor weak enough not to affect the $D$ component. For smoothing, this case results from a temporal overlap between the $D$ and $R$ components in the latency domain. In this situation, both $P_{dp}^D$ and $P_{dp}^R$ will show fine structure oscillations, as seen in Fig. 4-8 at certain intermediate values of $L_s$ and $\tau_{cut}$.

Our results with suppression unmixing suggests that there is no “ideal” suppressor level valid over a range of frequencies that simultaneously eliminates the reflection-source component while leaving the distortion-source essentially unaffected. Figure 4-8, for example, shows some residual fine structure in both $P_{dp}^R$ and $P_{dp}^D$ at this subject’s “optimal” suppressor level of $L_s = 55$ dB SPL. The choice of suppressor level involves a tradeoff between minimizing $|\delta|$ and minimizing $|\rho|$, with their sum inevitably finite. With proper choice of the windowing function, the prospects for near ideal unmixing by spectral smoothing (time windowing) appear brighter. Judging by the almost negligible amplitude of the fine structure obtained at intermediate values of $\tau_{cut}$, unmixing by smoothing appears able to effect a cleaner separation between the two components than is possible using suppression.

### 4.4.1 Estimating $\delta$ and $\rho$

We illustrate the tradeoff between $\delta$ and $\rho$ and give a feel for the suppression-based unmixing errors in Fig. 4-9, which shows estimates of $|\delta|$ and $|\rho|$ for three different suppressor levels. Computation of $\delta$ and $\rho$ requires knowledge of $D$ and $R$; the estimates in Fig. 4-9 were computed by substituting for $D$ and $R$ the components obtained by spectral smoothing (with $\tau_{cut} = 9$). Since the two equations represented in matrix form in Eq. (4.20) are not independent (but are related by $P_{dp}^D + P_{dp}^R = D + R$), an additional constraint is necessary to determine $\delta$ and $\rho$ uniquely. Since two parameters are available, a natural choice is to match both $P_{dp}^D$ and its frequency derivative, $P_{dp}^{D'}$. We thus obtain values of $\delta$ and $\rho$ by solving the pair of simultaneous equations

\[
\begin{align*}
P_{dp}^D &= (1-\delta)D + \rho R; \\
P_{dp}^{D'} &= (1-\delta)D' + \rho R'.
\end{align*}
\]

The values of $\delta$ and $\rho$ obtained in this way vary with frequency; at every point, the coefficients $\delta$ and $\rho$ are chosen to match to the curve $P_{dp}^D$, both its value and its derivative, as closely as possible.\footnote{The quantities $\delta$ and $\rho$ are thus analogous to the osculating parameters used in the theory of linear differential equations (e.g., Mathews and Walker, 1964).} Because of the constraint $P_{dp}^D + P_{dp}^R = D + R$, these same coefficients also provide a match to $P_{dp}^R$ and its derivative.
Figure 4-9: Estimates of the unmixing errors $\delta$ and $\rho$ at different suppressor levels. The figure shows magnitudes of the unmixing errors $\delta$ and $\rho$ computed as the solution to Eqs. (4.27) and (4.28). The components $P_{dp}$ and $P_{dp}^{R}$ obtained by spectral smoothing (with $\tau_{cut} = 9$) were used as estimates of $D$ and $R$. Results for three different suppressor levels were computed using the data whose magnitudes are shown in the top panels of Fig. 4-8.
Since the true components $D$ and $R$ are not known, the unmixing errors $\delta$ and $\rho$ shown in Fig. 4-9 were computed relative to the components obtained by spectral smoothing; they therefore provide only a rough guide to the actual errors. The results are, however, generally consistent with expectations based on the three special cases of Eq. (4.20) considered above. Note, for example, that $|\delta|$ and $|\rho|$ vary in opposite directions with changes in suppressor level. At the largest suppressor level, $|\delta|$ is relatively large and $|\rho|$ relatively small (corresponding to the expectations for strong suppression outlined in case #2 above). Similarly, at the smallest suppressor level, the relative magnitudes of $\delta$ and $\rho$ are reversed (weak suppression, as in case #1). At the “optimal” suppressor level, the errors $|\delta|$ and $|\rho|$ are intermediate between these extremes. Not surprisingly, $|\delta|$ and $|\rho|$ can become large in frequency regions where the total DPOAE is itself poorly determined (e.g., near 0.8 kHz where $|P_{dp}|$ is relatively close to the noise floor) and/or where the estimated components $P_{dp}^D$ and $P_{dp}^R$ change rapidly (e.g., near notches of $P_{dp}^R$). Overall, however, the unmixing errors are fairly small for intermediate suppressor levels (typically $|\delta| \sim 0.1$ and $|\rho| \sim 0.2-0.3$). These findings corroborate the qualitative visual impression that the two methods, selective suppression and spectral smoothing, unmix into generally similar components.

4.5 Discussion

In this paper we tested the two key predictions of the two-mechanism model of DPOAE generation by successfully unmixing DPOAEs into components, $P_{dp}^D$ and $P_{dp}^R$, with characteristics indicative of fundamentally different mechanisms of generation (i.e., nonlinear distortion vs. linear reflection). In agreement with Prediction #1, the phase of the putative distortion-source component ($P_{dp}^D$) is nearly constant whereas the phase of the reflection-source component ($P_{dp}^R$) varies rapidly with frequency. These differing phase slopes imply fundamental differences in the respective mechanisms of emission generation. In particular, the two slopes are consistent with generation by nonlinear distortion ($P_{dp}^D$) and linear coherent reflection ($P_{dp}^R$), respectively (Shera and Guinan, 1999). Furthermore, in agreement with Prediction #2, the spectral shape and phase of the reflection-source component closely match those of the SFOAE measured at the same frequency under comparable conditions (i.e., with the addition of an $f_1$-primary mimicker). Changes in the SFOAE caused by the mimicker suggest that the primaries have a significant influence on the reflection-source component of the DPOAE, presumably via suppression. To investigate the robustness of our conclusions, we unmixed DPOAE sources using two completely different methods: (a) selective suppression of the reflection source using a third tone near the distortion-product frequency and (b) spectral smoothing (or, equivalently, time-domain windowing). Although the two methods unmix in very different ways,
explicit analysis of the unmixing errors demonstrates that they yield similar DPOAE components, indicating that our results are not especially sensitive to the method of unmixing.

4.5.1 Source mechanism versus source location

The quasi-periodic fine structure often evident in DPOAE spectra is now generally regarded as resulting from the alternating constructive and destructive interference between backward-traveling waves originating in two separate regions of the cochlea (Kim, 1980). The physics underlying the interference pattern has generally been understood as follows (e.g., Brown et al., 1996; L.J.Stover et al., 1996; Fahey and Allen, 1997): Because the two sources are spatially separated, backward-traveling waves originating at the more apical location must travel further to reach the ear canal than waves originating at the basal location. Consequently, waves from the apical source are delayed relative to the basal source; in the frequency domain, this delay corresponds to a frequency-dependent phase shift. Thus, the relative phase of the two waves rotates with frequency, alternately passing through plus and minus one. This rotation of relative phase creates the interference pattern—known as DPOAE fine structure—when the two waves are combined in the ear canal. Kim (1980) originally referred to the two DPOAE sources as the “primary-place source” and the “characteristic-place source,” and considerable evidence now suggests that the two backward-traveling waves do indeed originate at these locations (e.g., Furst et al., 1988; Gaskill and Brown, 1990; Brown et al., 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Talmadge et al., 1998, 1999; Heitmann et al., 1998; Fahey and Allen, 1997; Siegel et al., 1998; Mauermann et al., 1999a,b).

We demonstrate here, however, that this place-based nomenclature—and the conceptual model that underlies it (e.g., Brown et al., 1996; L.J.Stover et al., 1996; Fahey and Allen, 1997)—although apparently accurate in its specification of the locations of wave origin, fails to capture the critical distinction between the two sources. As suggested by Shera and Guinan (1999), the fundamental distinction between the two sources is evidently not source location, but source mechanism. Indeed, only by incorporating both classes of emission-generating mechanisms (i.e., nonlinear distortion and linear coherent reflection) have models been able to account for the known phenomenology of DPOAE fine structure (e.g., Talmadge et al., 1998, 1999; Mauermann et al., 1999a). Accordingly, our terminology distinguishes the two components not by their place of origin, but by their mechanism of generation (i.e., distortion- vs. reflection-source components).

Our results support the two-mechanism model of DPOAE generation. To illustrate, consider how our experimental results would have differed if both sources in Fig. 4-1 had been distortion sources like D. When probed with the frequency-scaled
stimuli used here, both sources would then have generated backward-traveling waves with phases essentially independent of frequency. Consequently, the relative phase of the waves from the two sources would have been nearly constant, and no oscillatory fine structure would have appeared in the ear-canal pressure spectrum. Note that this constancy of relative phase would have occurred despite the fact that the two waves originate at different spatial locations within the cochlea. In other words, although the reflection-source region at R is further from the stapes than the distortion-source region at D, the difference in phase slope characterizing emissions from these two sources is not due to the differing locations of the D and R regions. Rather, contrary to standard assumption, phase slopes are ultimately determined by mechanisms of emission generation. For example, the theory of coherent reflection filtering (Shera and Zweig, 1993b; Zweig and Shera, 1995) implies that reflection-emission latency is determined not by the distance a wave travels to reach its characteristic place but by the characteristics of cochlear tuning it finds when it gets there (Shera and Guinan, 2000a,b).

4.5.2 Comparison with other work

The experiments reported here were designed specifically to test Predictions #1 and #2 and therefore differ from most other studies of DPOAE components (e.g., Brown et al., 1996; L.J.Stover et al., 1996; Siegel et al., 1998) in their use of frequency-scaled stimuli (i.e., fixed $f_2/f_1$). According to the analysis underlying the model (Shera and Guinan, 1999), distortion and reflection mechanisms yield qualitatively different phase behavior (i.e., nearly constant phase vs. rapidly rotating phase) when emissions are evoked with frequency-scaled stimuli. Similar qualitative differences in phase are not found using other measurement paradigms, and the underlying differences in mechanism can therefore be considerably less transparent. For example, much more rapid phase rotation occurs when distortion emissions are measured using stimulus paradigms (e.g., fixed $f_1$, fixed $f_2$, or fixed $f_{dp}$) for which the cochlear wave pattern is not simply translated along the cochlear partition (e.g., Kimberley et al., 1993; O Mahoney and Kemp, 1995; Shera et al., 2000). Unmixing DPOAEs measured using constant $f_2/f_1$-ratio sweeps, rather than one of the more common alternative paradigms, thus greatly facilitates recognition of the two emission mechanisms. By increasing the difference in phase slope between the distortion- and reflection-source components, our use of frequency-scaled stimuli also facilitates unmixing of the two components using spectral smoothing (time windowing) by maximizing the separation between the two components when the data are transformed into the “latency domain” using Fourier analysis.

Our tests of Prediction #2 contrast sharply with the findings of Brown et al. (1996), who performed a DPOAE unmixing analysis using a smoothing technique.
and compared the resulting “DP residual” (their analogue of $P_{dp}^R$) to measurements of SFOAEs. Although they noted similarities in the phase slopes, they found “little correspondence in the magnitude across frequency of the DP residual and SFOAE.” Their reported discrepancy between emission components conflicts with earlier work (Kemp and Brown, 1983), which found at least qualitative agreement between the SFOAE and the DPOAE component believed to originate at the distortion-product place (as obtained, in this case, using a suppression paradigm). In contrast with these results, we find excellent agreement, both between DPOAE components unmixed via different paradigms and between $P_{dp}^R$ and corresponding SFOAEs. Unfortunately, Brown et al. (1996) do not specify their smoothing algorithm in the detail necessary to enable a direct comparison with our method.\footnote{To smooth their frequency-domain measurements Brown et al. (1996) used a 101-point moving average (evidently tailored to a frequency spacing between points of approximately 1.2 Hz) but fail to specify the shape of their smoothing function. If all points in the moving average were weighted equally (i.e., if the smoothing function were rectangular), the corresponding time-domain window, a sinc function, would have been non-monotonic, oscillating about zero with a period of roughly 8.25 ms.} We note, however, that in our experiments the addition of the $f_1$-primary mimicker often improved the agreement between the magnitudes of $P_{dp}^R$ and $P_{fe}$ considerably (cf. Fig. 4-5). This result indicates that suppressive and other effects of the primaries on $P_{dp}^R$ must be taken into account in any such comparison.

4.5.3 Region of validity of the two-mechanism model

The tests of Predictions #1 and #2 reported here, together with more limited data at other (low to moderate) primary levels and at frequency ratios $f_2/f_1$ in the range 1.1–1.3, establish the validity of the two-mechanism model in humans for the DPOAE measurement parameters in common use [i.e., low to moderate sound-pressure levels with $L_1 \geq L_2$ and primary frequency ratios $f_2/f_1 \approx (f_2/f_1)_{optimal}$]. Knight and Kemp (2000c) provide a test of Prediction #1 over a broad range of frequency ratios $(1.01 \leq f_2/f_1 \leq 1.5)$ in an unmixing analysis of their stunning $\{f_1, f_2\}$-area map (Knight and Kemp, 2000b). Their results, based on time windowing of DPOAEs measured using primary levels $L_1 = L_2 = 70$ dB SPL, are consistent with the two-mechanism model and indicate that the relative amplitudes of the components $P_D^P$ and $P_{dp}^R$ vary systematically with $f_2/f_1$. Whether Prediction #2 also applies over a similarly broad range of parameter values remains an important open question.

Described and tested here in the frequency domain, Prediction #1 of the two-mechanism model evidently also applies in the time domain. Combining phase-rotation averaging (Whitehead et al., 1996) with an elegant pulsed-primary technique, Talmadge et al. (1999) provide strong support for model predictions that amount, in
effect, to time-domain analogues of Prediction #1. Since the responses involved arise in a nonlinear system, this conclusion is nontrivial. Time-domain tests of Prediction #2 await further experiment.

The validity of the model at high intensities also remains to be investigated. For example, at higher levels of intracochlear distortion, the emission evoked by the forward-traveling distortion component may contain, in addition to contributions from coherent reflection, significant energy from distortion-source waves created by nonlinear distortion (e.g., Withnell and Yates, 1998). Furthermore, the two emission sources may also begin to mix in ways more complicated than simple linear summation. For example, the strength of the micromechanical impedance perturbations that scatter the traveling wave may depend on the local amplitude of basilar-membrane vibration.

4.5.4 Methods of unmixing

Our success at unmixing using two completely different methods (suppression and smoothing) demonstrates the robustness of our conclusions to the method of unmixing. The two methods unmix in very different ways, and the systematic errors each introduces are presumably quite different. Whereas the suppression method separates components based on their differential modification by an external tone, the spectral-smoothing (or time-domain windowing) method separates components based on latency in the “time-domain response” obtained using Fourier analysis. Despite these differences, the two methods unmix the total emission into rather similar components (at least for $f_2/f_1 = 1.2$ and low-to-moderate primary levels). Whether the two methods yield similar results at other $f_2/f_1$ ratios and/or at higher stimulus levels remains an important open question. Differences between the methods would not be surprising at $f_2/f_1$ ratios close to one—although the spectral-smoothing method does not depend on spatial separation of source regions in the cochlea, the ability

---

15 We are reminded here of the dialectic described by Levins and Lewontin (1985): "A necessary step in theoretical work is to make distinctions. But whenever we divide something into mutually exclusive and jointly all-encompassing categories, it turns out on further examination that these opposites interpenetrate."

16 To unmix $P_{4p}$ into two components we used a window with a “low-pass” characteristic in the time domain. The technique is easily generalized to the unmixing of multiple components with different latencies (e.g., by using multiple “band-pass” windows centered at different latencies or a succession of “low-pass” windows with different cutoffs).

17 Working with SFOAEs at low stimulus levels, Shera and Zweig (1993a) established a similar equivalence between SFOAEs extracted using the vector-subtraction method (Kemp and Chum, 1980) and the method of spectral smoothing. The vector-subtraction method exploits the nonlinear saturation of the SFOAE—or “self-suppression” of the traveling wave (e.g., Kanis and de Boer, 1993)—at higher stimulus levels.
of the suppression method to selectively eliminate one of the sources presumably deteriorates as the two sources draw closer to one another as $f_2/f_1$ approaches one.

An advantage of the spectral-smoothing method is that it requires measurement of only a single quantity (namely, $P_{dp}$, whereas the suppression method requires both $P_{dp}$ and $P_{dp|suppressed}$). Unlike the suppression method, the smoothing method therefore allows each measurement of $P_{dp}$ to serve as its own control against possible systematic changes (e.g., variations in overall emission level due to efferent effects) that may occur during the course of the measurement. In the suppression studies reported here, we sought to minimize these potential problems by interleaving measurements of $P_{dp}$ and $P_{dp|suppressed}$ in time. Although the spectral smoothing method depends only on $P_{dp}$, it requires knowledge of $P_{dp}$ at multiple frequencies. Indeed, the method works best if applied to measurements that span a relatively wide frequency range (i.e., many periods of the microstructure) with good frequency resolution (i.e., many points per period). In addition, because of uncertainties introduced near the end-points due to incomplete knowledge of $P_{dp}$ outside the measured interval, the smoothing method requires measurements over an interval slightly larger than the desired frequency range. The suppression method, by contrast, imposes no such constraints; suppression unmixing requires measurement of $P_{dp}$ and $P_{dp|suppressed}$ only at the actual frequency (or frequencies) of interest.

4.5.5 Implications of unmixing DPOAEs

Uncontrolled mixing may be a substantial source of subject-dependent variability in DPOAE measurements. Indeed, our results imply that the interpretation of DPOAE responses appears doubly confounded. First, DPOAEs are mixtures of emissions originating from at least two different regions in the cochlea. This "spatial blurring," now widely recognized, compromises the frequency selectivity of DPOAE measurements (e.g., Heitmann et al., 1998). Second, DPOAEs are mixtures of emissions arising by fundamentally different mechanisms. This "mechanistic blurring," established here, compromises the etiological specificity of DPOAE measurements. For although both distortion- and reflection-source emissions share a common dependence on propagation pathways from the cochlea to the ear canal, and are therefore both sensitive to modifications of that pathway (e.g., to middle-ear pathology or to reductions in cochlear amplification caused by damage to outer hair cells), their respective mechanisms of generation—and hence their dependence on underlying parameters of cochlear mechanics—remain fundamentally distinct. For example, whereas distortion-source emissions presumably depend on the form and magnitude of cochlear nonlinearities (e.g., on the effective "operating point" along hair-cell displacement-voltage transduction functions), reflection-source emissions depend strongly on the size and spatial arrangement of micromechanical impedance perturbations (e.g., on variations
in hair-cell number and geometry). Distortion-product unmixing, using techniques such as those employed here, should therefore improve the power and specificity of DPOAEs as noninvasive probes of cochlear function.

Acknowledgments

We thank our experimental subjects for their considerable patience and gratefully acknowledge the efforts of Paul Fahey, John Guinan, William Peake, and two anonymous reviewers, who provided valuable comments on the manuscript. This work was supported by grants R01 DC03687 and T32 DC00038 from the NIDCD, National Institutes of Health.
Chapter 5

Summary and conclusions

5.1 Chapter Overview and Summaries

Shera and Guinan (1999) characterized OAEs as originating by two fundamentally different mechanisms, namely: 1) by linear reflection from pre-existing mechanical perturbations that are arrayed along the length of the cochlea, and 2) by distortions that are induced by the stimulus acting on a nonlinear system. The model suggests that the relative contribution from each mechanism to any emission measurement depends on factors such as the type and intensity of the evoking stimulus. In this thesis we tested the relationships between common OAE measurements in humans and these two proposed mechanisms of OAE generation.

The mechanism-based model predicts that OAEs generated by the two mechanisms can be distinguished by the frequency dependence of their phase. Whereas OAEs generated by wave-induced perturbations are predicted to have a phase that is nearly independent of frequency,\(^1\) OAEs generated by reflection from pre-existing mechanical perturbations are predicted to have a phase that changes rapidly with frequency. Based on these predicted differences in phase behavior,\(^2\) we characterized OAE types by measuring and comparing OAEs evoked with single tones and broad-band clicks, as well as those evoked by two-tone complexes at frequencies not contained in the stimulus, so-called distortion-product emissions.

In the following we summarize the key findings and conclusions from each of the three parts of this thesis.

---

\(^1\)This prediction only applies when the emissions are measured with frequency scaled stimuli. See chapter 4

\(^2\)Although for brevity, we focus on the phase characteristics of the emission to make the distinction between emission types, models make predictions regarding both the spectral form and phase of the emission. Our tests always considered both features to compare and discriminate between emissions types
1. **Tone-evoked otoacoustic emissions are not an artifact of measurement method.**

Before interpreting tone-evoked otoacoustic emissions (also referred to as stimulus-frequency otoacoustic emissions) within the context of the two-mechanism model, we first needed to validate the reliability of existing SFOAE measurement techniques. Because SFOAEs and the pure tones used to evoke them overlap in time and are at the same frequency, they are difficult to separate. To deal with this problem, SFOAEs are measured using techniques that exploit nonlinear features of the cochlear response, such as suppression and the compressive growth of emissions with stimulus intensity. Each of these two techniques suffers from potential problems which in turn make it difficult to interpret the resulting measurement of SFOAEs. For example, the suppressor tone might not only suppress the emission but could also induce SFOAE sources. Or, as suggested by (Patuzzi, 1996), SFOAEs might not be real acoustic emissions originating from the cochlea but might instead be an artifact of nonlinear measurement techniques. To test if the features of SFOAEs are dependent on measurement method, we compared the results of suppression and compression to the results of a signal processing technique that does not rely on cochlear nonlinearities (i.e., “time-domain” windowing or equivalently spectral smoothing).

Below we summarize key findings from this study:

- SFOAEs measured by suppression, compression, and time-domain windowing are nearly equivalent at low to moderate levels.

- Similarity between suppression and compression applies for suppressors close to the probe frequency and for high-level compressors/suppressors. By taking into account the pressure of the $2f_1 - f_2$ distortion product otoacoustic emission that is present when using suppressors (i.e. the total output of the emission signal), we accounted for a large portion of the difference between suppression and compression for near-probe suppressors.

- Calibration drifts and calibration contamination by OAEs can cause errors in spectral smoothing. Our two-stage calibration procedure addresses these problems and improves the performance of spectral smoothing, particularly at higher stimulus levels.

- The similarity between SFOAEs extracted by the three techniques indicates that near-probe suppressors extract the total SFOAE evoked by the probe. In other words near-probe suppressors do not induce significant SFOAE components, in agreement with modeling studies (Shera et. al. 2004).

Although each of the above three methods exploits a different cochlear phenomenon or signal-processing technique to extract the emission, we found that they all yield SFOAEs with similar spectral structure, magnitude and phase. We conclude that SFOAEs are not an artifact of measurement method. Features of the SFOAE that are generally regarded as characteristic of SFOAEs –
for example, the rapidly varying phase and spectral structure with prominent notches – truly represent the features of an acoustic emission.

2. **SFOAEs and CEOAEs are the same emission generated by linear-reflection mechanisms at low to moderate intensities.**

Although linear-reflection models of OAE generation predict that both click-evoked and stimulus-frequency otoacoustic emissions originate via essentially linear mechanisms, other models imply that differences in the spectrum of the evoking stimulus result in differences in the mechanisms of OAE generation.

The work reported in this section tested for bandwidth dependent differences in mechanisms of otoacoustic emission generation. Click-evoked and stimulus-frequency OAE input/output transfer (or describing) functions were obtained as a function stimulus frequency at various intensities.

Below we summarize the findings from this study:

- At low stimulus intensities, CEOAE and SFOAE transfer functions are independent of level and have nearly identical spectral characteristics. This finding is consistent with the idea that CEOAEs and SFOAEs are generated by the same linear mechanism at low levels (e.g., Zweig and Shera, 1995; Talmadge et al., 1998) but contradicts other models of OAE generation (e.g., Nobili et al., 2003a).

- Although CEOAEs and SFOAEs grow nonlinearly at moderate levels, they continue to share similar spectral and phase characteristics up to the highest intensities measured.

- The close correspondence across a wide range of stimulus levels between responses evoked by tones and clicks is consistent with similar behavior seen in basilar-membrane responses to broadband and narrow band stimuli (e.g., Recio and Rhode, 2000; de Boer and Nuttall, 2002).

- When stimulus intensity is measured in component-equivalent sound pressure level, CEOAE and SFOAE transfer functions have equivalent growth functions at fixed frequency and equivalent spectral characteristics at fixed intensity.

The near-equivalence between CEOAEs and SFOAEs suggests that the OAEs evoked by broad and narrow band stimuli (CEOAEs and SFOAEs) are generated by the same mechanism. The spectral characteristics of CEOAE and SFOAE transfer functions are consistent with the predictions of coherent-reflection models of OAE generation; specifically, a rapidly varying phase with frequency. The near-equivalence between SFOAEs and CEOAEs, provide further confidence that commonly reported SFOAE characteristics are not an artifact of the measurement technique but rather reflect a common underlying mechanism of generation. Our findings indicate that measuring CEOAEs provides a fourth independent way to measure SFOAEs. We conclude that although CEOAEs and SFOAEs are conveniently given different names based on the characteristics of
the stimuli used to evoke them, the two OAEs “types” are better understood as members of the same emission family; namely, linear-reflection.

3. **DPOAEs are a mixture of emissions generated by both linear and nonlinear mechanisms.**

The two-mechanism model predicts that distortion product OAEs (DPOAEs) are comprised of emissions generated by both linear-reflection and nonlinear distortion. We unmixed DPOAEs into components using two completely different methods: (a) selective suppression of the putative reflection source using a third tone near the distortion-product frequency and (b) spectral smoothing (or, equivalently, time-domain windowing). Although the two methods unmix in very different ways, they yield similar DPOAE components.

Below we summarize other findings from this study:

- We tested predictions of the two-mechanism model of DPOAE generation (Shera and Guinan, 1999) by unmixing DPOAEs into components with characteristics indicative of fundamentally different mechanisms of generation (nonlinear distortion vs. linear reflection).
- In agreement with the model, the phase of the putative distortion-source component is nearly independent of frequency and the phase of the reflection-source component rotates rapidly.
- The spectral shape and phase of the reflection-source component match those of the stimulus-frequency emission (SFOAE) measured at the same frequency under comparable conditions (i.e., with the addition of an f1-primary mimicker). Changes in the SFOAE caused by the mimicker suggest that the primaries have a significant influence on the reflection-source component of the DPOAE, presumably via suppression.
- We obtained qualitatively similar results using two completely different methods of unmixing: (1) suppression and (2) spectral smoothing (or its equivalent, time windowing).
- DPOAE fine structure results from the alternating constructive and destructive interference between the distortion- and reflection-source components.

As suggested by Shera and Guinan (1999), distortion product otoacoustic emissions are a mixture of emissions generated by both linear-reflection and induced-distortion sources.

### 5.2 Implications and Future Directions

By restructuring the way we understand emissions to be generated, the two mechanism model suggests a framework by which we can understand apparent inconsistencies that arise from the ‘one mechanism’ view. The one-mechanism model, for
example, suggests that the relative strength of different types of OAEs should be coupled. If all emissions are generated by the same underlying mechanisms, then the factors that create increases in SFOAEs should also create increases in DPOAEs and visa versa. Since distortion and reflection-source emission mechanisms depend on fundamentally different aspects of cochlear mechanics (distortion emissions depend on the form and operating point of cochlear nonlinearities and reflection emissions depend on the size and density of mechanical perturbations), a dissociation between the two types of emissions is now not surprising. For example, SFOAEs and SOAEs are diminished by aspirin whereas DPOAEs are unaffected. In the context of the two-mechanism model, one might explain this effect as aspirin acting preferentially on the reflection-emission generators without effecting the distortion-emission generators.

Measurements that isolate and study each emission type should improve our understanding of the underlying factors that affect each mechanism of OAE generation. Studies that monitor changes in each type of OAE – for example, in response to controlled application of ototoxic drugs, or cochlear injury – are the next logical step toward developing these ideas into useful clinical and research tools. Distortion emissions – since they are evoked by high intensity stimuli and are inherently dependent on the cochlea being driven into its nonlinear region of operation – might be most useful for monitoring cochlear status at high stimulus intensities. On the other hand, reflection-emissions (i.e., SFOAEs and CEOAEs) which are most robust near threshold may be more sensitive to small changes in cochlear status. For example, since active amplification is most effective at low intensities, the effect of ototoxic drugs that might change the gain of the cochlear amplifier will presumably be most apparent at low intensities.

Separating and independently studying OAEs generated by linear-reflection and induced-distortion mechanisms should also improve the utility of existing OAE measurements. For example, uncontrolled mixing between linear-reflection and distortion-induced emissions may be a substantial source of subject-dependent variability in DPOAE responses. The mixing makes interpretation of DPOAE responses doubly confusing: (1) DPOAEs originate at two different spatial locations, thereby compromising their frequency selectivity; and (2) DPOAEs are a mixture of emissions that arise by two fundamentally different mechanisms, thereby compromising their etiological specificity. Unmixing techniques – such as those employed in this thesis - provide a method for separating the emission components.

Models suggest (Shera, 2003) that the sharpness of cochlear tuning, associated with long group delays and large phase slopes of the cochlear traveling wave, might underly the difference in strength between distortion emissions in primates versus non-primates. The argument is based on the observation that SFOAE group delays and sharpness of tuning are greater in humans than in cat or guinea pig (Shera et al., 2002). The synopsis of the argument is that the overlap between the excitation patterns of two spectral components will be smaller in a sharply tuned cochlea, and the rapidly varying phase might create greater phase cancellation between distortion sources. As a consequence the net distortion resulting from the interaction between the two patterns will be smaller.

The variability in OAE amplitudes across different species can be understood as
a difference in the strength of the two mechanisms in each species. In contrast to CEOAEs in guinea pig, which are comprised almost entirely of distortion emissions (Yates and Withnell, 1999b), we found that human CEOAEs are generated by linear-reflection mechanisms. Consistent with this notion, in rodents, DPOAEs are large, whereas SFOAEs are small. In primates, SFOAEs are large but DPOAEs are small. In the context of the two-mechanism model, this finding is consistent with the idea that distortion-source generators are stronger in non-primates than in primates.

Existing measurements of apical mechanics suggest that cochlear responses in the apex are broader than responses in the base. Does this frequency dependent change in sharpness of tuning translate to differences in distortion-emission strength across frequency? How are the emission characteristics affected by these differences? Estimates of SFOAE group delay, for example, deviate from model predictions at frequencies below 1kHz. Since predictions of OAE behavior are based on models of basal cochlear mechanics, deviations from predicted behavior at low frequencies reflects a fundamental gap in our understanding of OAE generating mechanisms and underlying cochlear mechanics at low frequencies.

Although, linear models can account for the observed SFOAE and CEOAE spectral structure and phase at low and moderate levels, does this hold at higher intensities? The emission evoked by tones and clicks at higher levels of intracochlear distortion may contain significant energy from distortion-sources, in addition to contributions from reflection-sources. Furthermore, at higher stimulus intensities, reflection and distortion sources may mix in ways more complicated than simple linear summation. The effect of complex cochlear conditions, including contributions from efferents, in shaping SFOAEs and CEOAEs is also not well understood. It is not clear if and at what stimulus levels these complexities might appear and dominate or if and how these effects vary across subjects.

5.3 Conclusion

For OAEs to be maximally useful as probes of cochlear function one must understand their properties over the entire range of hearing (meaning over both stimulus levels and intensities). In this thesis we focused on one part of this problem, namely understanding how SFOAES, CEOAEs, and DPOAEs are generated over a small range of frequencies (between 1 and 4kHz) and within that frequency range how they change with stimulus intensity. Understanding how these phenomena are frequency dependent over a broad range is an important open question.

Nevertheless our results in this thesis provide strong experimental support for the two mechanism model of OAE generation. At low to moderate stimulus intensities, human SFOAEs and CEOAEs are best categorized as linear-reflection emissions, whereas DPOAEs are a mixture of emissions generated by both linear-reflection and induced-distortion. An integrated two-mechanism model, which incorporates both nonlinearities and micromechanical-impedance perturbations, as required by linear-reflection models, can account for much of the observed features of human OAEs.
Bibliography


Summary & Conclusions


Summary & Conclusions


Summary & Conclusions


Summary & Conclusions


