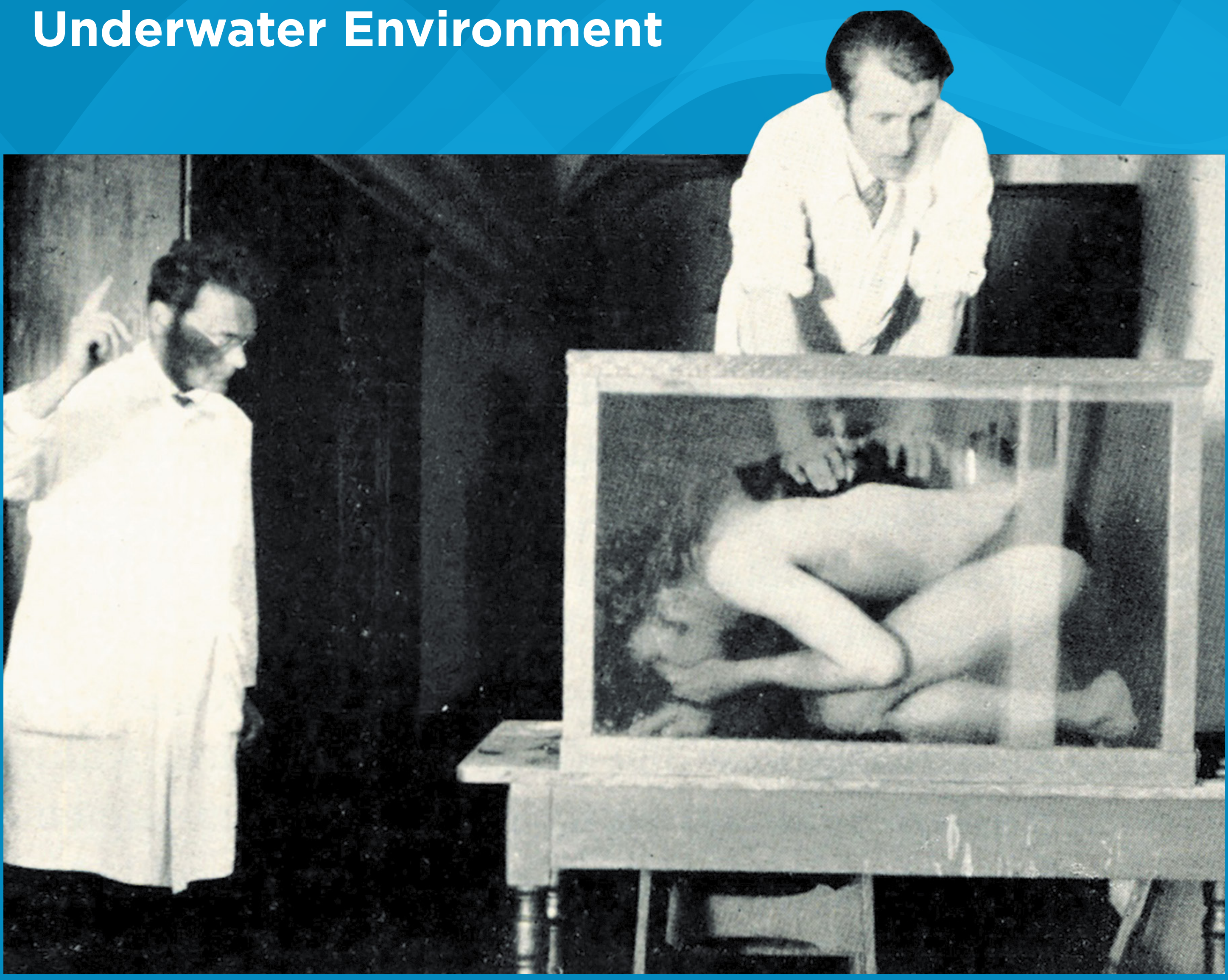
# Acoustics Today

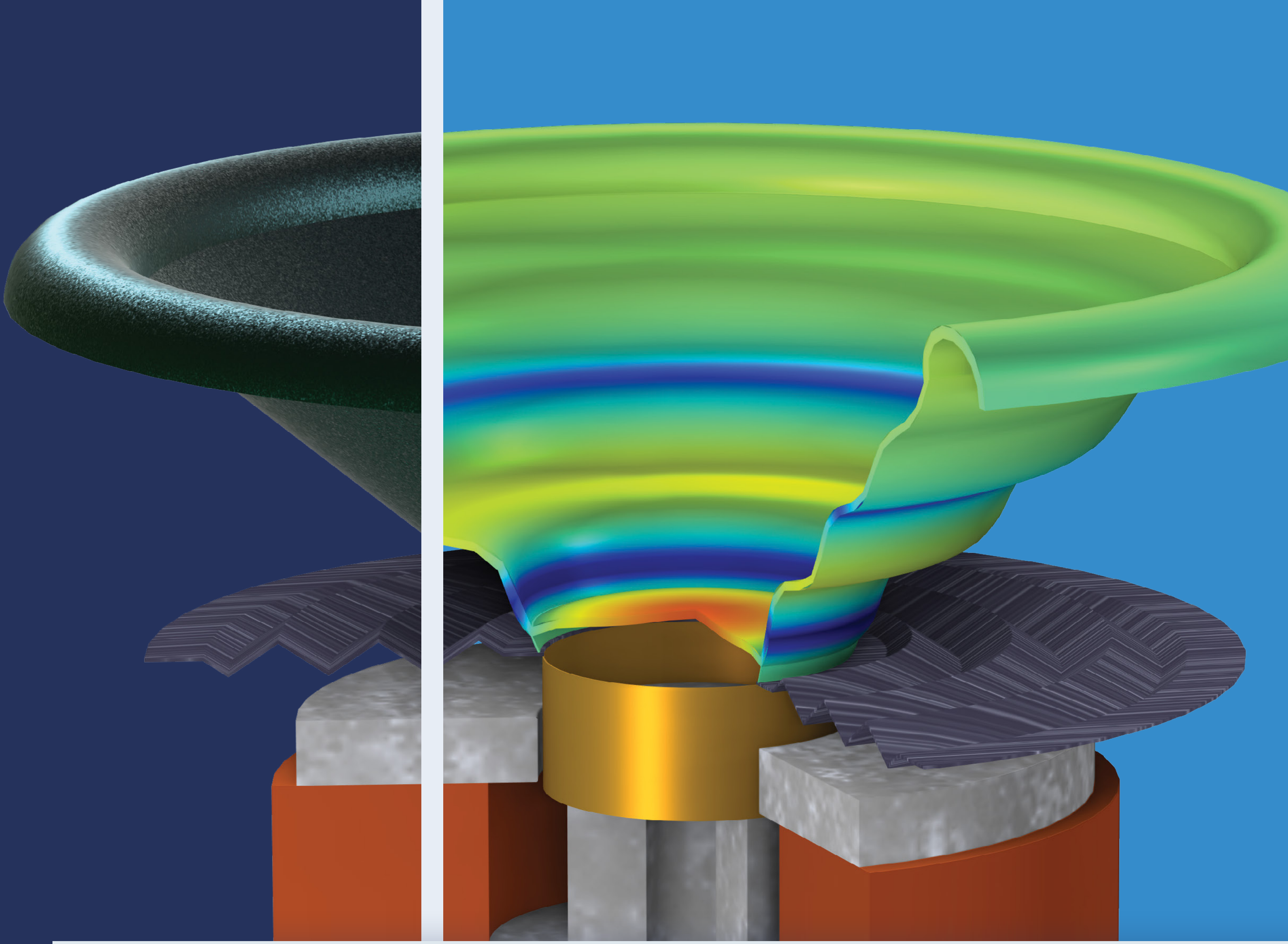
**Spring 2022** Volume 18, Issue 1

ASA ACOUSTICAL SOCIETY OF AMERICA

An Acoustical Society of America publication

## Human Hearing in the Underwater Environment





# Acoustics Today

ASA
An Acoustical Society of America publication

Spring 2022 Volume 18, Issue 1



## Featured Articles

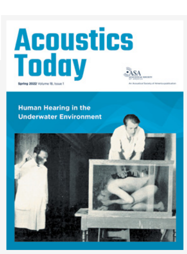


## Sound Perspectives



## Departments






### About the Cover

Cover image from article by Casper and Babina (page 23). Figure reprinted by permission from Springer Nature: Springer Nature, *Zeitschrift für vergleichende Physiologie*, "Untersuchungen über den Gehörsinn der Fische, besonders von *Phoxinds laevis* L. und *Amiurus nebulosus* Raf," H. Stetter, ©1929.

Visit the online edition of *Acoustics Today* at AcousticsToday.org


ASA ACOUSTICAL SOCIETY OF AMERICA

*Publications Office*
P.O. Box 809, Mashpee, MA 02649
**(508) 293-1794**


Follow us on Twitter @acousticsorg

Please see important *Acoustics Today* disclaimer at www.acousticstoday.org/disclaimer.

## Acoustical Society of America

The Acoustical Society of America was founded in 1929 "to generate, disseminate, and promote the knowledge and practical applications of acoustics." Information about the Society can be found on the website:
www.acousticalsociety.org

Membership includes a variety of benefits, a list of which can be found at the website:
www.acousticalsociety.org/asa-membership


Acoustics Today (ISSN 1557-0215, coden ATCODK) Spring 2022, volume 18, issue 1, is published quarterly by the Acoustical Society of America, 1305 Walt Whitman Rd., Suite 110, Melville, NY 11747-4300. Periodicals Postage rates are paid at Huntington Station, NY, and additional mailing offices. POSTMASTER: Send address changes to Acoustics Today, Acoustical Society of America, 1305 Walt Whitman Rd., Suite 110, Melville, NY 11747-4300.

SysCheck2™

# Validation of your calibration with a single click

## The first intelligent acoustic sensor system with built-in self-verification

GRAS 246A0
with SysCheck2™
functionality


- Remote verification of the entire measurement chain (microphones, cables, input module and software)
- Easy setup—no physical interaction required
- Rapid validation of all channels with a single click
- Provides monitoring of environmental data
- Use of standard CCP-compatible cabling
- Connected to a CCP power module with TEDS support

SysCheck2™ is a GRAS patented technology for verifying measurement chain integrity. SysCheck2™ allows remote, simultaneous health check on microphones, channel gain and cable integrity.

GRAS 246A0 with SysCheck2™ is a high-precision microphone ideal for testing hearing aids, earphones, headphones, and headsets in ear simulators/couplers and test fixtures/chambers.

SysCheck2™ helps reduce human errors and increase data reliability both in the lab and on the production line. It is particularly convenient when microphones are placed where physical verification is a risk - for example, if the microphones are hidden in complex mechanical structures, such as couplers or specialized test boxes on production lines.

SysCheck2™ functionality can be accessed with many data acquisition systems but offers seamless plug-and-play operation with Audio Precision's APx data acquisition software.


**Read more at:** www.grasacoustics.com/syscheck2

GRAS Sound & Vibration
*An Axiometrix Solutions Brand*

# From the Editor

*Arthur N. Popper*

Readers of *Acoustics Today* (*AT*) have often commented that the magazine covers are attractive and well designed. In fact, we take great pride in our covers and seek to make them interesting and eye-catching. In deciding on our cover image, our first choice is to have a figure from an article, but if there is no figure that excites our cover selection team, we invite an artist to create an illustration that represents one of the articles.

Once we have selected the figure, we turn it over to our production group, Opus Design. Opus creates several cover designs that are reviewed by the cover selection team who then makes the final decision on a cover. Once the issue is published, we offer the author of the cover article a high-resolution image of the cover. But if anyone would like a copy of a cover, email me (apopper@umd.edu) and we will try and provide you with a high-resolution PDF or JPEG.

The first article in this issue, by H. Timothy Bunnell, addresses the science behind the artificial voices such as Siri and Alexa that we deal with every day. Tim talks about the different approaches that have been used in speech synthesis and how they have changed over the years.

This is followed by an article by Brandon M. Casper and Matthew A. Babina who discuss human hearing underwater. A major takeaway from the article is that humans are not adapted to hearing underwater and we don't do nearly as well as fishes or marine mammals. In trying to focus on this comparison, the authors found a wonderful photo from a 1929 paper that is the cover of this issue. As an aside, the person at the left in the picture is one of the early experts on fish hearing, Karl von Frisch who won the 1973 Nobel Prize in Physiology or Medicine.

In the third article, Patrick O. Kanold writes about how the "wiring" of the auditory region of the brain during fetal development is influenced by early auditory experiences. In doing this, Patrick considers the potential influence of the mother singing or speaking on brain development.

In the next article, Robert A. Lutfi and Christopher A. Brown write about the fascinating scholarly contributions of former Acoustical Society of America (ASA) President William A. (Bill) Yost. Bill has made lasting contributions to our understanding of how humans perceive sound, and the article not only talks about these contributions but also puts them into a more general perspective of human hearing. You can learn more about human hearing at https://bit.ly/AT-psychoacoustics.

I found the fifth article by Christina J. Naify, Kathryn H. Matlack, and Michael R. Haberman of particular interest because it covers a topic I've heard about but know nothing about, additive manufacturing. Additive manufacturing uses three-dimensional printers to "build" objects that can range in size from tiny circuits to houses. The article provides a wonderful overview and introduction to the topic and focuses the application of the technique to diverse areas of acoustics.

The final article by Sean E. Olive is one that most ASA members will easily relate to, the quality of headphones that many of us use on a daily basis. Sean takes the reader through how headphone quality is evaluated and describes how manufacturers determine what listeners prefer in headphones. Sean does not make any recommendations as to what headphones to get, but I suspect that many of us will start to use his information when we consider purchasing new headphones.

This issue has three "Sound Perspectives" essays. The first one is the last essay in our "Ask an Acoustician" series. The series will be replaced by a new series of essays, again to be developed and "organized" by *AT* Associate Editor Micheal Dent. Be sure and look for that series in the next *AT* issue.

In deciding that this would be our last essay in this series, Micheal decided that I should be the subject. I will admit that I was rather reluctant to do this at first, but I was persuaded to agree. I found that answering the questions that Micheal has been posing to our colleagues for the past four years to be interesting and instructive. Indeed, although most readers will not have been subjects of this series, I invite everyone to look at the questions and think about how they would answer them; you might learn something about yourself! And if you want to look at past pieces, they are all posted at https://bit.ly/3FjTCeL.

I also want to thank Micheal for her development, organization, and cowriting all the "Ask an Acoustician" essays. Micheal is a great partner to work with on *AT*, and I am very grateful for her collaboration in so many aspects of the magazine, including her "eagle eye" in doing the final editorial review of most articles.

The second essay is by Tyrone Porter. Tyrone writes a very powerful piece about his experiences as a Black acoustician. Personally, I found the piece very moving and have already encouraged Tyrone to write more about his experiences for future issues of *AT*. I strongly recommend that every member of the ASA read and think about Tyrone's essay and his experiences.

I also invite other members of the ASA with important and interesting stories that can teach others about issues of diversity to share their stories through our "Sound Perspectives" essays. If you would like to consider doing this, please email me and let me know what you have in mind.

Our third essay is actually a repeat of one we had in the winter 2021 issue about Spanish-Speaking acousticians in the Americas, but in Spanish. When we originally did this essay, I invited the authors to provide a Spanish translation for the *AT* web page. I then realized that many ASA members might value seeing the essay in Spanish, and, with the agreement of the authors, we publish the piece in this issue. Indeed, there are other Spanish-language articles from *AT* on our web pages — please look at http://bit.ly/AT-Spanish.

Finally, it is a delight to welcome the newest member of the *AT* family, Bennett Easthon Setzer. Bennett is the son of *AT* Editorial Associate Kat Setzer and her wife Lindsey Easthon. We look forward to Ben being an *AT* intern in perhaps 20 years!

# From the President

*Maureen Stone*

My goals for this column are to promote participation in the leadership and increase awareness of the finances of the Acoustical Society of America (ASA). We are a member-driven Society and you, the members, are better served when this information is easily accessible. I hope you will enjoy participating at all levels of the Society, especially the highest.

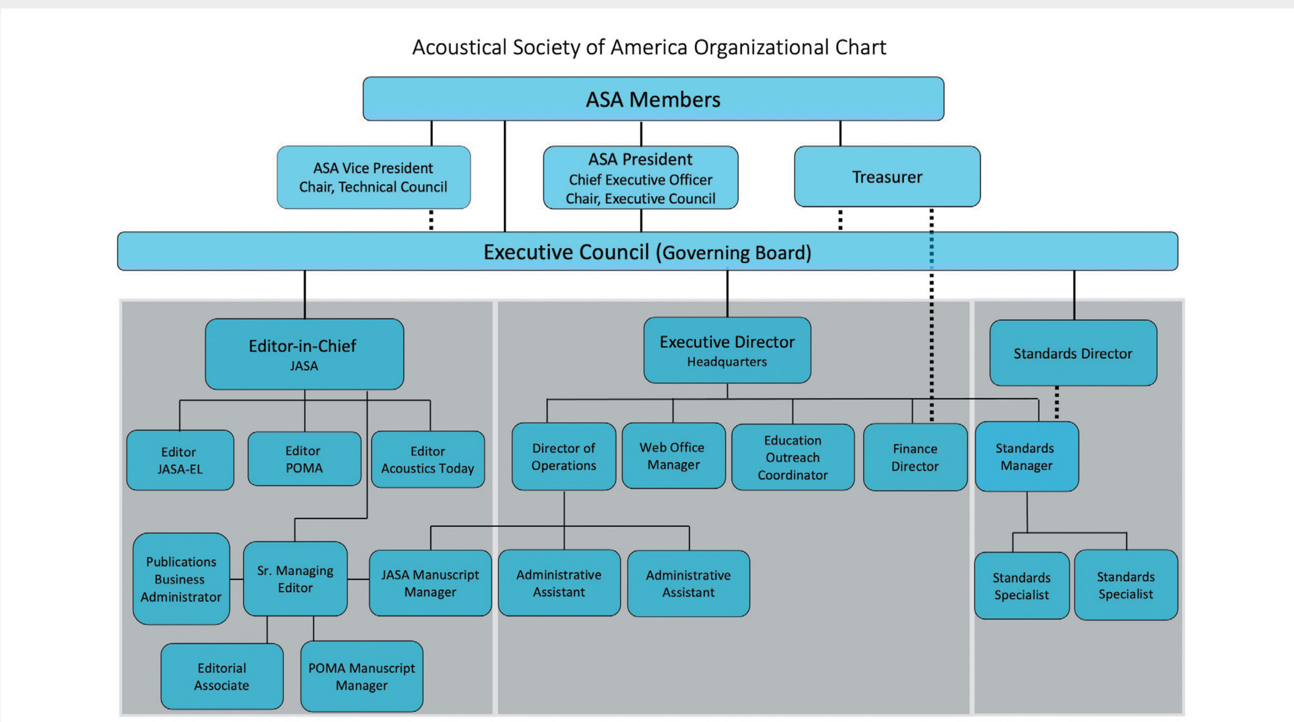
## Participate in Acoustical Society of America Leadership

In my fall 2021 column (see https://bit.ly/From_President), I discussed ways to become involved in the ASA. Today, I want to build on that by discussing how to move into ASA leadership roles. The ASA governing structure is completely captured in two organizational charts. **Figure 1** illustrates the basic organization chart.

The EC is the governing board of the ASA and is chaired by the president. It has 13 voting members including the 3 presidents (president-elect, current president, immediate past president), the 3 vice presidents (VP) (VP-elect, current VP, immediate past VP), the treasurer, and 6 additional elected members. The nonvoting members include the editor in chief (EIC), the standards director, and the executive director (ED). The EC ensures that the ASA follows its mission, strategizes and plans for the future of the Society, and signs off on financial decisions and actionable items from various committees including medals and awards. It also acts on reports and appointments from the president, the EIC, the ED, the standards director, the VP (as chair of the Technical Council (TC)), and the treasurer.

**Figure 1** shows the employees and contractors who work at the ASA (gray squares). You can see that fewer than 20 staff members run all the operations of the ASA, which

**Figure 1.** *Organization chart of the Acoustical Society of America. The ASA members elect the officers and Executive Council (EC) at the spring elections (top). The EC supervises Publications, Headquarters and Standards (gray blocks).*



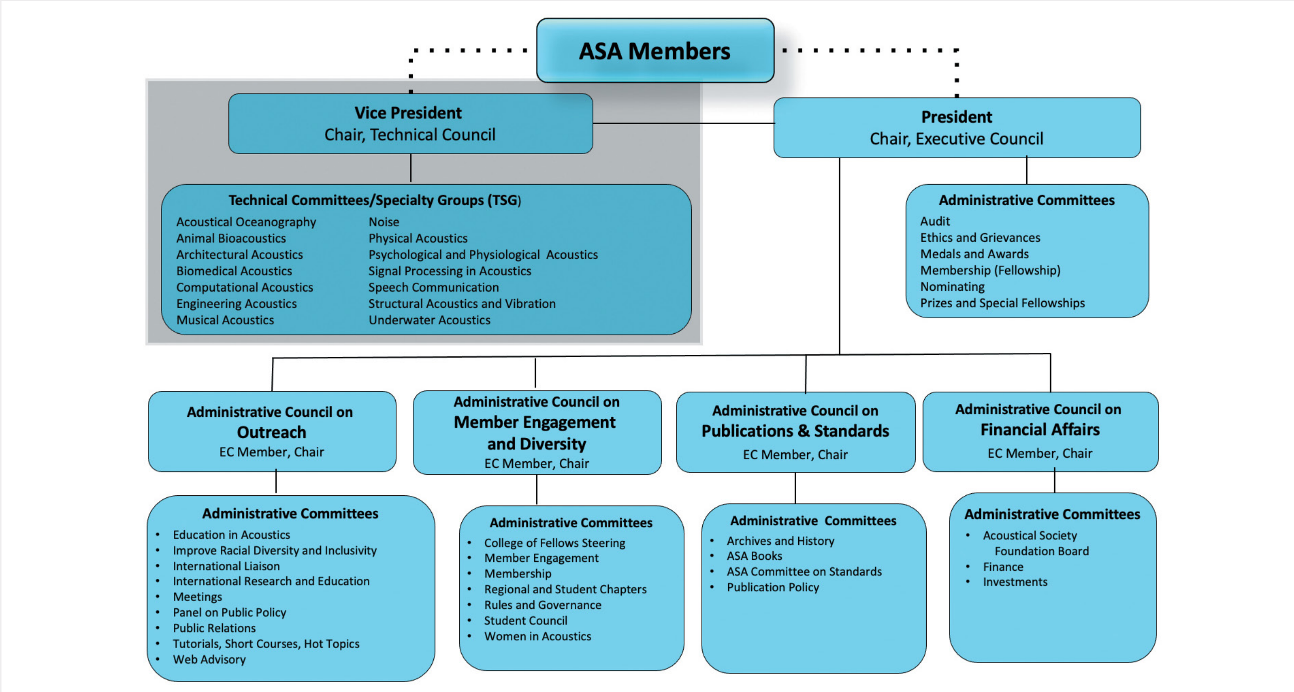



**Figure 2.** *Organization chart of Acoustical Society of America committees.*

has almost 7,000 members! This is only possible because of the army of volunteers who carry out the mission of the Society by working on the various ASA committees and in leadership. This volunteer army is shown in the committee organizational chart (**Figure 2**).

The TC is chaired by the VP and is composed of the 13 Technical Committee chairs who provide direct representation of the Society membership through the open Technical Committee meetings that take place at each ASA meeting. You can find a list of the technical committees and their roles at https://bit.ly/ASA_committees (see also **Figure 2**). The term TC is used at the ASA to represent both the Technical Council, and the individual Technical Committees, e.g.: TCUW. In this column I will use it to represent the council and will write out technical committee in full, but this is not usually done, so keep the two meanings in mind in future when you hear or read the acronym.

The ASA also has administrative committees that support the mission of the Society and report to the EC (**Figure 2**). Membership on some administrative committees requires specialized knowledge or ASA experience. Administrative committees are also listed at https://bit.ly/ASA_committees.

As a member-driven organization, member input is solicited on many ASA issues at the Society meetings. The ASA accomplishes this by organizing its meetings into a sandwich structure. On Monday morning, the EC meets to consider and discuss Society business including many items that require membership input. These are discussed by the technical committee chairs (i.e., the TC) on Monday afternoon. The technical committee chairs discuss the ideas at the open technical committee meetings later in the week where they solicit feedback, new ideas, issues, and concerns from their members. Everyone is welcome to attend these meetings. The ideas generated at the open meetings are discussed further by the technical committees at their Friday morning meeting and then brought to the EC by the VP Friday afternoon. Along with their other responsibilities, the EC acts on the new and continuing technical committee items as appropriate.

It is not uncommon for chairs or members of technical committees and administrative committees to later be a candidate for ASA elected office. These committee experiences are not required for officer positions, but the expertise is very useful. To be a candidate for elected office, you must be a full member or Fellow (see https://bit.ly/3r01Ovt for an online application to transfer to full membership).

Every year, a new nominating committee proposes candidates for president-elect, VP-elect, and two of the six members of the EC. In addition, any ASA member may propose names of potential candidates, including themself, to the nominating committee. Self-proposal of candidates is intended to widen the pool of candidates to as many members as possible.

Now let's turn to the running of the Society and, in particular, ensuring our fiscal health.

## Understanding Acoustical Society of America Finances

ASA finances are managed with an eye toward Society growth and sustainability. We must optimize our current financial position (balance sheet) and configure our annual operating budgets for the long run. ASA finances are overseen by an elected treasurer (Judy Dubno), the ED, and the Finance Director and supported by the ASA headquarters (HQ) staff, all of whom recommend policy to the EC. Four administrative committees (Audit, Finance, Acoustical Society Foundation Board [ASFB], and Investment) provide additional oversight to ASA finances and also recommend policy to the EC. These committees are comprised of ASA members and supported by ASA HQ staff.

The treasurer is the chief financial policy advisor to the EC and is responsible for oversight of the long-term strategic financial matters of the Society and short-term financial policy implementation. The treasurer works closely with the four financial administrative committees, the EC, and the HQ staff on all things financial. The ED provides oversight to the ASA HQ staff to manage day-to-day financial activities, and the finance director provides expertise and guidance to committees on all financial and accounting related issues.

The Audit Committee, composed of three EC members, assists the EC by working with our external auditors to monitor the appropriateness and integrity of the Society's financial reporting, accounting policies, and internal controls.

The Finance Committee assists the EC on long-term financial viability of the ASA by making recommendations on long-range financial decisions. The Finance Committee looked deeply into ASA finances and submitted a report to the EC in December 2020. As a result of that report, it was decided to hire a consultant to review the ASA entire financial structure so that we can move toward a fiscally sustainable model to support operations and programs.

The ASFB manages the Acoustical Society Foundation Fund (ASFF) for the Society, and acquires, maintains, and ensures the correct expenditure of endowment funds and restricted and unrestricted donations. Their monies go entirely toward ASA prizes, awards, grants, fellowships, scholarships, and programs (for more details, see https://bit.ly/3ta8x90).

The Investment Committee reviews the financial investments of the ASA and, with EC approval, develops investment and spending policies and procedures for the Society. The Investment Committee works with the ASFB and two independent investment advisors to plan the optimal methods for increasing earnings from our endowment, dividends, and interest while managing risk.

For more information, about these and all the ASA administrative committees see the Rules at https://bit.ly/3ASARules and the committee organizational chart in **Figure 2**.

Another crucial piece of information needed to understand the ASA financial structure is the relationship between our sources of revenue and our expenditures. To make this straightforward, I am presenting this as four financial groups: HQ staff, meetings, standards, and publications, although some overlap may occur.

The first group is the ASA HQ staff. The ED and her staff (see **Figure 1**) manage the day-to-day operations of the ASA and ASA outreach programs, with collaboration from many administrative committees. HQ salaries and other expenses are funded in part by the revenue from annual membership dues.

The second group contains the semiannual Society meetings, whether in person or virtual. These are also managed by the ED and her staff, with collaboration from local teams and many administrative committees. Sources of revenue include meeting registration fees, exhibitor fees, and the newly developed sponsorship fees. Expenses include everything related to the meeting, including staff.

The third group contains the standards director and standards office (see **Figures 1** and **2**), also supported by five standards committees. The standards director works with the standards office to support the development of and maintain the national and international standards in acoustics. The revenue from standards royalties and fees pays for a portion of the standards related salaries and expenses.

The fourth group is publications. The EIC is supported by three other editors, several staff, and administrative committees, who manage all aspects of the various ASA publications (see **Figures 1** and **2**). Publications is our largest source of revenue, and the only group of the four that routinely realizes an annual surplus; the other groups have had recurring annual deficits.

The financial picture of the ASA is currently experiencing some challenges, like many other scientific societies, but we are taking major steps to address these. Among many reasons for the annual operating deficits is inflation in salaries and hiring additional staff throughout the ASA. In addition, increasing expenses of operations and programs managed by the HQ staff exceeds revenue from membership dues. The growth in expenses to support our wonderful meetings is rising faster than the revenue, with some meeting deficits exceeding $100,000. We are just learning about revenue and expenses for virtual meetings, and they are not necessarily revenue positive. Standards royalties and fees are not constant from year to year and tend not to cover expenses for the standards group. In the past, the annual surplus from ASA publications has been used to make up the shortfalls in the other three groups. However, more recently, the publications surplus has not kept pace with the deficits, making this practice unsustainable in the long run.

Although we have substantial fiscal issues, we are working very hard to turn the situation around. Indeed, investment reserves and earnings are very strong, and we are currently working on strategies to reduce expenses and increase revenue over the next few years to provide a more sustainable long-term fiscal model. We have been working with our financial consultants to gain a better understanding of the annual operating budget and define better strategic uses of the investment reserves. We now have a budget dashboard that will provide timely monthly reports of revenues and expenses. This will allow us to better project and control future expenditures and help us align resources to our operating budget. The Meetings Reimagined Ad Hoc Committee is focusing on every aspect of meetings to enhance meeting programs while increasing revenue and reducing expenses wherever possible. We also are creating a Revenue Reimagined Ad Hoc Committee, which will formalize our continuous exploration of new revenue streams to support ASA activities.

On the expense side, the EC is reviewing ASA programs, committee expenses, and other expenditures to ensure they continue to be mission critical and to determine what can be modified to reduce expenses. This will continue over the next few years as we continue to enhance ASA fiscal policies.

## In Summary

This is an exciting time for the ASA. Every member can be an important part of the process of running and steering this ship. I encourage you to make your opinion count by preparing for and joining an administrative committee, attending open TC meetings, and developing your skills at organizing and leading. You will learn about the mechanisms of running a large society, determine what interests you in directing the organization, and work with the ASA team of dedicated, smart, valuable contributors. It is very enjoyable, and you will be making a substantial contribution to the strong and vital Society that is so important to the careers of all of us!

# Speech Synthesis: Toward a "Voice" for All

*H. Timothy Bunnell*

Text to speech (TTS) has become so much a part of our everyday lives thanks to Alexa, Google, Siri, and many others that we have come to know (if not always love) that it is difficult to recall a time when it was not so. Synthetic voices like those for Siri and others fill multiple roles today. They deliver announcements of important information over public address systems in noisy places like airports where high intelligibility of the speech in noise is crucial to ensure the information they carry is heard correctly. A synthetic voice may be the first entity a customer interacts with when contacting a company and it is important for the voice, as a representative of the company, to present a natural and pleasing voice quality that is representative of the company's image. Synthetic voices serve as the *only* voice for individuals whose own voice is lost due to injury or a progressive neurological disease like amyotrophic lateral sclerosis (ALS; also called Lou Gehrig's disease or motor neuron disease [MND]) or who have a congenital dysarthria due to a condition such as cerebral palsy. And TTS voices allow blind or nonliterate users to read content from news stories, books, and computer screens while giving busy people an opportunity to "read" email even when driving their car.

## A Framework and Baseline for Text to Speech

These current use cases for TTS voices provide insight into the successes of the underlying technology and also highlight areas where work remains. The need for intelligibility, naturalness, and ability to convey an individual's vocal identity are obvious from these examples. Less obvious but no less important is the expressiveness of the synthetic speech: the ability to express through intonation or "tone of voice" (Pullin and Hennig, 2015) the intent underlying the words of an utterance.

In this article, I trace how we arrived at the current state of the science for TTS, showing how the technology improved with the adoption of newer approaches and improved numerical techniques. A natural start is with the work of Klatt (1980) who provided Fortran software for implementing a cascade/parallel formant synthesizer. Klatt (1987) provided a history of TTS conversion, which was remarkable for the inclusion of a collection of audio examples for many of the synthesizers he discussed (see Ramsay, 2019, for an interesting review of early mechanical synthesizers).

Crucially, the period around the publication of these two articles by Klatt (1980, 1987) marked an important era in the TTS field. From a purely commercial perspective, it was arguably during this time that TTS systems became commercially mainstream, largely through improvements in the intelligibility of the speech that they generated.

Second, during this period, TTS technology started to be adopted by nonvocal persons to enhance their ability to communicate with others. One of Klatt's visions for Digital Equipment Corporation's DECtalk system, which emerged directly from his work at MIT, Cambridge, Massachusetts, was its application in augmentative and alternative communication (AAC) devices for communication by individuals who are nonvocal. Until that time, augmented communicators depended mainly on mechanical communication boards that required communicants to point to words or letters to express themselves. Recently, the field has come to refer to these speech-enabled communication devices as speech-generating devices (SGDs), the term I use in this article.

In this article, I present a framework that captures the structure and function of the TTS advances. Throughout, a goal is to focus on the implications for SGD users' communication.

**Figure 1** provides a unified framework for discussing modern TTS systems. Each block or component in the



https://doi.org/10.1121/AT.2022.18.1.14

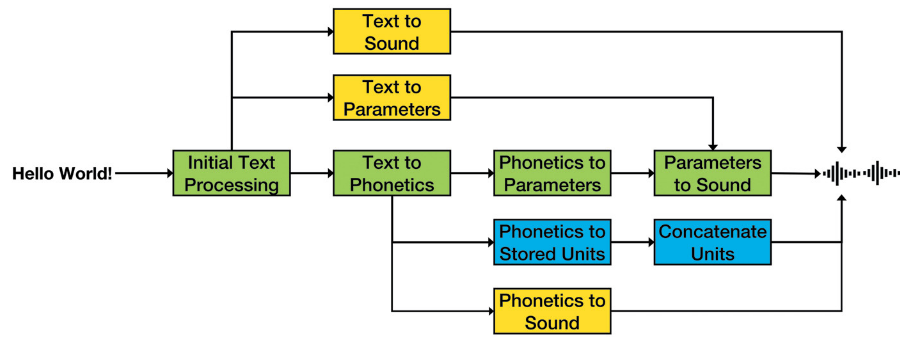



**Figure 1.** *Unified schematic covering current text to speech (TTS) system designs.* ***Colors*** *highlight components for different types of TTS systems.* ***Green components*** *are shared by many types of TTS systems. See* ***Figure 2***, ***green*** *and* ***blue***, *and* ***5***, ***green*** *and* ***yellow***, *for specific pathways.*

figure represents a *logical* element of the TTS process as it is usually conceived. I start with a description of a generic rule-based formant synthesizer like DECtalk (**Figure 1**, *green*). I focus on this pipeline to set the baseline to show the types of changes that have been made over time to improve the technology.

## Formant Synthesis from Rules

Formant synthesis systems (and virtually all other TTS systems I discuss) require some form of *initial text processing* (**Figure 1**, *green*). Typically, this involves tokenizing the input text stream into distinct words or tokens and text normalization to convert nonword tokens such as numbers and abbreviations into the words one would speak when reading the tokens aloud. Thus, consider the text input "Dr. Smith lives at 1702 S. Park Drive and can be reached by phone at 555-456-7890." The first instance of "Dr." must be converted to the word "doctor," while the second instance should be replaced with the word "drive." Given that 1702 S. Park Drive appears to be an address, a likely rendering would be "seventeen oh two south park drive." The final phone number would be replaced with the words "five five five, four five six, seven eight nine oh," with commas or other textual markers to indicate the appropriate phrasing for a phone number. Of course, the challenge for text normalization is to derive enough information of the textual input to make accurate guesses about things like phone numbers or addresses.

A related problem for text normalization is disambiguating the pronunciation of homophonous words. Often, context can provide helpful clues; if someone is "playing a bass," they are more likely to be a musician than an actor impersonating a fish. But sometimes disambiguation requires much deeper semantic/pragmatic knowledge that can easily be guessed from context alone. Is a shiny white bow a holiday decoration or the front of a boat?

The tokenized and normalized input text, along with any additional meta information related to prosodic properties (the intonation and timing properties) derived from the initial text processing, is next passed to the *text to phonetics* component (**Figure 1**, *green*), which produces a symbolic phonetic representation. In the original rule-based formant synthesis systems like DECtalk, this representation consisted of little more than a string of phoneme symbols along with some formal boundary and intonation symbols. Boundary symbols indicate the degree of acoustic/phonetic separation between two adjacent phonemes. For example, the boundaries between words are often marked by distinct acoustic features; consider the distinction between "gray day" and "grade A." Moreover, the boundaries between phrases of different types are also marked by phonetic duration differences, pauses, and intonational features such as the rising pitch at the end of many questions or the falling pitch at the end of a declarative sentence.

The intonation symbols express the relative locations and types of pitch accents or "tones" relative to the phonetic symbols. Over time, a standardized system has developed based on the concepts of "tones and break indices" (ToBI; e.g., Silverman et al., 1992) that describes the intonational structure of English and other languages in terms of a discrete set of tones corresponding to a relative maximum or minimum in fundamental frequency (perceived as voice pitch) that aligns to a specific syllable within an utterance. Similarly, break indices are single-digit integers that indicate the relative separation between two elements in an utterance. ToBI-like symbol sets are often used for the boundary and intonation symbols in current TTS systems.

Next, the *phonetics to parameters* components (**Figure 1**, *green*) maps the symbolic phonetic description of the input text to a numerical representation suitable for input to a vocoder or parametric synthesizer to generate a speech waveform from the numerical parameter values. Whereas the phonetic symbols imply a sequence of related acoustic events, there are no time units at the symbolic level. In a rule-based formant synthesizer like DECtalk, the phonetics to parameters component is responsible for laying out the parameters as a dynamic time-varying sequence with defined temporal coordinates. Typically, parameters are updated every few milliseconds at a constant prespecified rate, for example, every five milliseconds.

Finally, the *parameters to sound* component (**Figure 1**, *green*), often referred to as a "vocoder," accepts the parametric representation of speech and generates audio output. In many parametric systems, a source/filter model of speech is adopted wherein a source signal consisting of either a periodic impulse train or white noise is passed through a digital filter representing the human vocal tract.

## Application of Text to Speech to Speech-Generating Devices

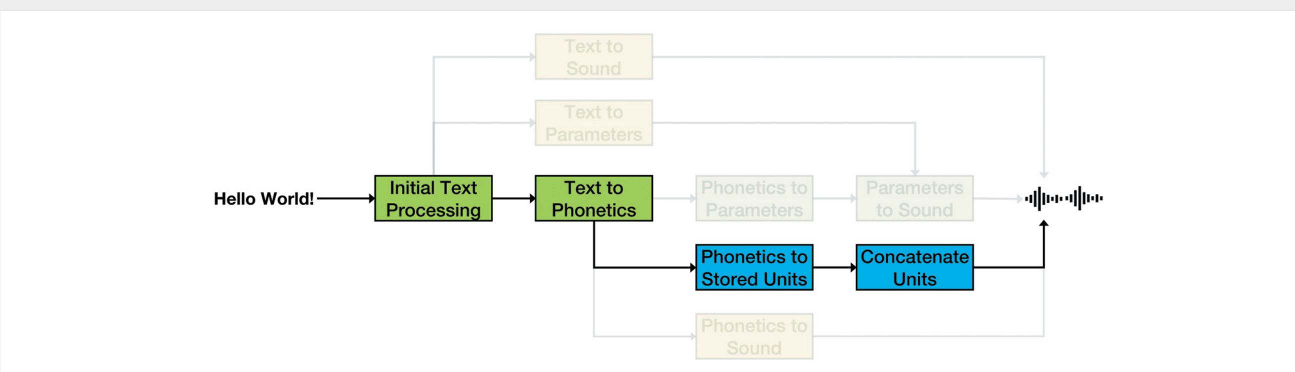
Formant-based TTS systems were intelligible enough to become widely adopted by assisted communicators in the late 1980s and 1990s, with DECtalk being the most commonly used system in the SGDs of the time (see https://bit.ly/31E9A54). Perfect Paul, which was demonstrably the most intelligible of the DECtalk voices (Green et al., 1986), was the voice of choice for many AAC users of the time. Even women would often choose to use the male Perfect Paul voice because it was more easily understood by others. Imagine attending a meeting in a conference room with multiple people using SGDs all tuned to Perfect Paul and not being entirely certain whose device had just emitted an important comment! So, although many nonvocal persons now had a voice, they did not have their *own* voice for communication.

In addition to not providing every AAC user with a unique voice, the formant synthesis systems of the time did not sound particularly human. As I discuss in **Diphone Synthesis**, a technique called diphone synthesis emerged as one possible way to generate more human-sounding and identity-bearing synthetic speech. But neither formant synthesis nor diphone synthesis addressed another shortcoming, a lack of expressiveness. Attempts were made to create a more expressive output for DECtalk by modifying the synthesis parameters to convey emotional states such as boredom or sadness (Murray and Arnott, 1993), but they were not widely implemented.

## Diphone Synthesis

Diphone systems represented an important bifurcation in TTS technology: the distinction between knowledge-based systems and data-based systems. This distinction can also be described as between rule-based systems where a human expert must design the rules and corpus-based systems where a corpus of speech data provides the

**Figure 2.** *Component pipeline for diphone and other concatenative synthesis methods from **Figure 1**.*



information that would otherwise need to be expanded from rules. Or, as seen in **Statistical Parametric Speech Synthesis**, the corpus can be used to automatically discover the rules through machine-learning algorithms so that no expert is needed. Thus, the rules needed for the phonetics to parameters component of a formant synthesis system required expert knowledge of acoustic phonetics and a lot of hard work. However, corpus-based systems were able to replace much of that work by simply storing the data that would otherwise need to be developed from rules.

As illustrated in **Figure 2**, diphone synthesis (and related "concatenative" methods) follows a slightly different path within our overall TTS model.

A diphone is the region of speech spanning roughly the middle of one phoneme to the middle of the next phoneme. **Figure 3** illustrates this using the word "bob." The initial and final /b/ segments are relatively stable as is the /a/ vowel near its center. However, the acoustic structure changes rapidly around the borders between the consonants and the vowel. As long as the phoneme centers are reasonably similar across different phonetic contexts (they really are not, but we are assuming that they are close enough!), then cutting speech up into diphone-sized units ought to allow one to concatenate the diphones in novel ways to produce nearly any utterance. For example, take the [ba] from [bab] and the [at] from "cot" [kat] to create "bought" [bat]. This was the insight that led Dixon and Maxey (1968) to develop a formant diphone synthesizer (see #18 at https://bit.ly/3qxs3uL) that used stored formant synthesis parameters rather that a rule system to generate the parameters prior to synthesis.

Formant synthesis parameters are an interesting choice for the diphone storage because they have several useful properties. (1) They do not require a large amount of storage (a factor that was especially important in 1968!). (2) They are orthogonal, that is, it is possible to change any one parameter value without impacting the values of other parameters. (3) Interpolation between values for any parameter will yield another valid parameter value.

However, formant synthesis parameter values have not been the most common format for storing diphone units. More commonly, diphones have been stored as linear predictive coding (LPC) coefficients (e.g., see #34

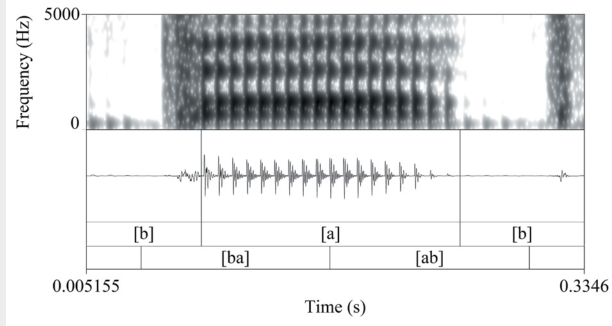



**Figure 3.** *Illustration of phonemes versus diphones.* ***Top****, spectrogram of the word bob.* ***Dark bands****, regions of high energy, corresponding to formants.* ***Middle****, acoustic waveform.* ***Bar*** *below waveform, phoneme locations ([b], [a], and [b]).* ***Bottom bar****, locations of the two diphone regions ([ba] and [ab]).*

at https://bit.ly/30n0V6V) or as waveform data stored in a format amenable to the fundamental frequency (F0) and duration modification using an algorithm like Pitch Synchronous OverLap Add (PSOLA; Moulines and Charpentier, 1990).

As is often true with speech processing, the most natural sounding of these formats in terms of voice quality would be waveform data because that is the least processed. LPC coding preserves much of the speaker identity information, but some voice quality may be lost in processing. Formant synthesis generally produces the least natural-sounding audio. Unfortunately, waveform data are the least compact storage format and also the most difficult to work with in that they afford little opportunity to adjust for discontinuities at diphone boundaries.

The *phonetics to stored units* (**Figure 2**, *blue*) is the path taken from the text to phonetics component for diphone synthesis. There are a relatively small number of diphones for any language. For example, Dixon and Maxey (1968) based their inventory on a total of 41 phonemes, so a theoretical maximum of $41^2$ = 1,681 possible diphones. Consequently, the conversion from phonetics to stored units amounts to simply looking up the needed sequence of diphone units.

The selected diphone units can then be passed to the *concatenate units* (**Figure 2**, *blue*) component that concatenates the selected units to form the desired output utterance. If the storage format permits, there may be additional adjustments to the units during the concatenating process. This could include adjustments such as smoothing potential discontinuities across diphone boundaries, adjusting diphone duration per a timing model, or even adjusting the F0 per an intonation model. Once the diphones have been assembled and concatenated to form an utterance, additional processing, if any, is applied to map from the diphone storage format to a digital audio waveform.

Diphone synthesis held one particularly intriguing possibility for SGD users, the ability to capture an individual's vocal identity. Because only a small amount of recorded speech is needed to create a diphone inventory, it would be possible to inexpensively mass produce unique diphone voices as long as the process of selecting diphones from recordings could be automated. People using SGDs could have a unique personal voice by selecting a suitable voice donor to do the recording. Moreover, people diagnosed with a condition such as ALS that threatens the loss of their voice could do the recording themselves and thus "bank" their voice for later use as a synthetic voice in an AAC device. In the mid-1990s, my laboratory at the Nemours Children's Hospital, Delaware, began experimenting with an extension of diphone synthesis (e.g., Bunnell et al., 1998) that would allow ALS patients to bank their voice in this way, a process referred to as "voice banking."

Diphone TTS voices, although a promising technology, did not generally gain much traction among AAC device manufacturers or SGD users. The small memory footprint for rule-based formant synthesis was certainly an important factor in favor of the formant-based TTS voices for AAC manufacturers. Furthermore, diphone TTS voices did capture the vocal identity of the person who recorded the diphone inventory but did not permit expressiveness, particularly for systems that used waveform concatenation, and despite capturing voice quality well, diphone synthesis tended not to flow in a natural manner. Moreover, many of the inexpensive diphone TTS systems available in the 1980s and later were less pleasing to listen to than the DECtalk voices that were provided with most AAC devices (e.g., see #29 at https://bit.ly/30n0V6V). That changed, however, with the emergence of unit selection TTS systems in the 1990s.

## Unit Selection Text-to-Speech Voices

One of the greatest difficulties with diphone synthesis was the impossibility of selecting a collection of diphones that did not suffer from sometimes jarring discontinuities at concatenation boundaries. This was less of an issue for diphones stored, as per Dixon and Maxey (1968), in a format that was amenable to substantial adjustments to smooth over or entirely eliminate disjunction by interpolating smoother parameter trajectories at segment boundaries. However, the highest voice quality obtainable from diphone synthesis was for diphones stored as waveform data or equivalently prewindowed PSOLA epochs. Unfortunately, with waveform concatenation and other issues, notably jarring differences in spectral features, F0, and amplitude at diphone boundaries were common.

These issues with waveform concatenation were largely addressed by an extended approach called "unit selection" (e.g., Zen et al., 2009) wherein a large amount of speech from a single individual is recorded and segmented into units that could be diphone size or smaller. This approach is illustrated in **Figure 4** using the word *two* as the target utterance and assuming each unit is roughly half of a phoneme. The units are stored along with additional features describing the linguistic details of the phoneme or waveform region from which they were drawn, such as the type of word (function vs. content word), syllable stress, syllable location, phrase location, presence and type of pitch accent on the associated syllable, and boundary level for the associated syllable. Because a unit selection database may contain a large number of candidates for each possible unit, there is a much greater chance of finding one or more units that exactly or nearly match the intended output context along all of the coded linguistic dimensions. Moreover, in the process of selecting units for concatenation, it is possible to select the specific candidates that will also minimize spectral discontinuities or sudden jumps in F0 or other factors that cannot be indexed as specific linguistic features.

Unit selection voices came to dominate the commercial TTS voice market in the late 1990s and 2000s because they are much more natural-sounding and intelligible than other commercially available TTS voices. Sometime in the 2000s, most SGD manufacturers included at least a few unit selection voices in their products. Moreover,

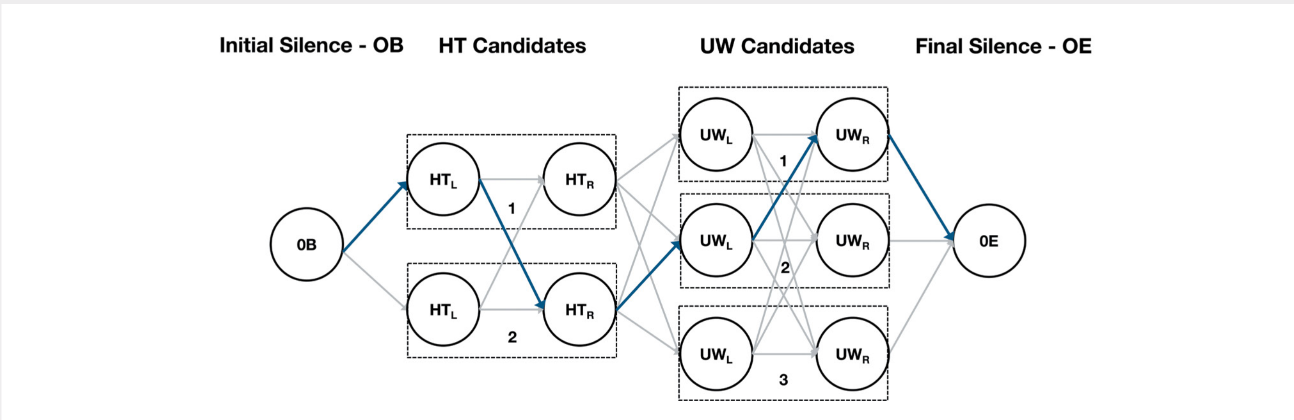



**Figure 4.** *Unit selection search process for the word "two." Two phonemes are required: /t/ (HT) and /u/ (UW) along with initial and final silence pseudo phonemes (0B and 0E). Multiple instances of each phoneme (**numbers** in boxes) are selected, each of which has two subphonemic "units" (e.g., $HT_L$ and $HT_R$). Each unit receives a target cost based on linguistic appropriateness and joined costs are assigned between units based on the acoustic continuity (**gray arrows**). The search locates the specific candidate units that minimize the combined target and joined costs over the utterance (paths shown with **blue arrows**).*

most SGDs transitioned from proprietary hardware to being software running on embedded Microsoft Windows systems. Because of this, most SGDs were also able to include voices provided by Microsoft or third-party voices written to published Microsoft standards.

My laboratory moved to a full unit selection system for voice bankers based on 1,600 utterances of various lengths and composition, comprising roughly one hour of running speech at normal speaking rates. With funding from the National Institute for Disability and Rehabilitation Research and later from the National Institutes of Health (NIH), I was able to offer a free experimental voice-banking service and provided a small number of voices to participants throughout most of the 2000s. Voices built in the laboratory could be incorporated with any Windows-based SGD. I formally began referring to the service as the ModelTalker project (Bunnell et al., 2005). Although the ModelTalker service was the first such service regularly used by ALS patients for voice banking, there are now excellent voice-banking services offered by a variety of commercial TTS companies, notably Acapela.com and Cereproc.com, who also offer voices for languages other than English. I have live example voices on the ModelTalker.org website (see https://bit.ly/3C57WpT; it might be slow when the website is busy).

By the late 2000s, unit selection was considered the best available TTS technology. The major voices for services like Siri and Alexa were built on unit selection technology as were enterprise-grade voices for large business call centers. However, the amount of recorded speech from voice talent needed to create the highest quality general-use voices exceeded tens of hours of running speech and many more hours of studio time. Even then, it is fairly easy to find examples of words that did not sound entirely natural within some specific context. There is no way to anticipate and record all of the possible acoustic phonetic variation within any language, even if factors like vocal effort, voice quality (breathy, hoarse, modal, fry, pressed), speaking rate, articulatory precision, and so forth are held constant. Moreover, for a truly natural-sounding and expressive TTS voice, one would not want to hold those factors constant!

The massive increase in memory density and decrease in memory cost over several decades made it feasible to work with unit selection voices despite their rapidly growing data footprint. But no amount of memory is really able to overcome the combinatorial ceiling that unit selection voices ultimately must hit. This prompted much interest in the possibility of returning to parametric synthesis, but rather than parametric synthesis with expertly crafted rules to describe dynamic parameter variation, statistical machine-learning techniques could be used to automatically capture the temporal patterning in synthesis parameters. The improvements brought by this effort to synthesize speech are now discussed.

## Statistical Parametric Speech Synthesis

As with unit selection synthesis, statistical parametric speech synthesis (SPSS)(Zen et al., 2009) requires a substantial corpus of speech data to be used in training its parametric phonetic models. Unlike unit selection synthesis, once the training process is completed, however, the original speech waveform data are no longer needed. Instead, the SPSS machine-learning process develops models for the acoustic structure of each phoneme. These models are then able to generate the time-varying parameters values for the parameters to sound component of the TTS system. Thus, fully trained SPSS models replace hand-coded rule systems in the phonetics to parameters component in **Figure 1**. In practice, the SPSS models are commonly sets of hidden Markov models (HMMs), one model for each phoneme, that describe the acoustic structure of the phoneme as a sequence of acoustic states, allowing the time-varying trajectories of parameters to be regenerated from the properties of the state sequence. The parameters the SPSS models learn are typically those describing the time-varying speech source function (voicing or friction) and moment-to-moment spectral features. The parameters to sound or vocoder component then uses the source and spectral parameters to regenerate audio data via digital filtering.

SPSS synthesis has several advantages over both rule-based formant synthesis and unit selection. First, because the SPSS models for parameter generation can be trained on a corpus of speech from a single talker, the output of the SPSS voice sounds recognizably like the talker who recorded the corpus. Moreover, because the training process is largely automatic, building multiple personal voices is not especially difficult or labor intensive. Compared with unit selection based on a similar-size speech corpus, particularly for smaller corpora (those having less than four hours of running speech), SPSS voices are not prone to discontinuities at segment boundaries and tend to have more natural-sounding prosodic structure. And because SPSS voices use parametric synthesis, it has the potential for changing characteristics of the voice quality or introducing expressiveness, but this potential is not yet realized.

There are, however, two main drawbacks to SPSS voices. First, the naturalness of the resulting synthetic voice is limited by the ability of the vocoder to reproduce natural-sounding voice quality. Some vocoder output sounds "buzzy" or "mechanical" when compared with unit selection voice quality. Second, in SPSS, each phonetic model represents an average of the acoustic patterns seen for all instances of the same contextually similar phonetic segment. This averaging tends to obscure some of the natural variability in human speech, leading to more monotonous sounding speech. Often, SPSS systems attempt to compensate for this averaging effect by exaggerating or boosting the variability of parameters over time. However, once the natural variability is lost due to averaging, it is not really possible to restore it.

Despite these two drawbacks, ACC users of ModelTalker voices have generally had favorable reactions to SPSS voices and the best of the SPSS laboratory TTS systems have been able to produce speech with audio quality closely approaching that of unit selection systems. Any long-term debate about the relative merits of unit selection versus SPSS voices, however, appears to rapidly becoming moot, particularly as it applies to large commercial grade TTS voices. This is due to the emergence of new deep-learning models.

## Deep Neural Network Speech Synthesis

In the past decade, deep neural networks (DNNs) and deep learning have revolutionized machine learning and led to large-scale improvements in several application areas. Large improvements have been observed in areas as diverse as speech recognition, machine translation between languages, natural language processing, text summarization, and speech synthesis. Explaining, even grossly, how DNNs function is beyond the scope of this article, but a few examples and consideration of how some models are changing the flow within the TTS system framework shown in **Figure 5** may give a reasonable sense of the emerging changes.

In **Figure 5**, the path from text to phonetics through phonetics to sound is a good place to start because this is the path used by WaveNet (van den Oord et al., 2016), which was one of the first "end-to-end" neural TTS systems. The authors have created an excellent website that describes their work and provides audio examples (see https://bit.ly/3qtNrkm). Training for WaveNet required about 25 hours of speech from a single female speaker and required days of CPU and GPU processing on Google's servers.

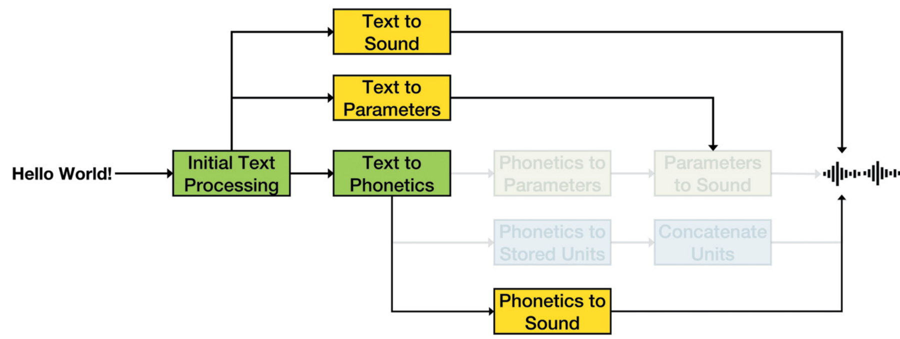



**Figure 5.** *Deep neural network (DNN) TTS pipelines emerging in current research efforts from* ***Figure 1***.

A large number of current end-to-end neural TTS systems follow the path from initial text processing through text to parameters and thereafter to a parameters to sound component. In some cases, "text" is taken somewhat broadly to refer to both literal words or characters, or to a form in which standard word spellings are replaced with something like International Phonetic Alphabet (IPA) characters to resolve letter to sound ambiguity. This is particularly helpful for languages like English that have borrowed words from many other languages and also helps when building multitalker and multilanguage systems. Most systems on this path generate Mel-scaled spectrograms as the output of the text to parameters component, relying on one of several vocoder methods (e.g., Griffin and Lim, 1984) or DNN-based vocoders, for generating audio output from the Mel-scaled spectrograms without explicitly applying a source/filter model. (Note: the Mel scale is a perceptually motivated transformation of linear frequency to a scale with approximately equal pitch steps; see Stevens et al., 1937.) However, a few systems may also generate parameters for alternative vocoders such as the *WORLD* vocoder (Morise et al., 2016). Although no systems are presently doing this, output in terms of formant synthesis parameters is also conceivable, with the final parameters to sound component being a formant synthesis vocoder.

Finally, as the ultimate end-to-end DNN TTS approach there is the path from initial text processing through TTS directly to audio output. This is a system referred to as end-to-end adversarial TTS (EATS) by Donahue et al. (2020; see https://bit.ly/3wpQBGR for audio examples). There is nothing before the audio generation but a light text-processing stage to handle tokenization and text normalization, perhaps with an additional substitution of IPA word spellings instead of standard word spellings. The system is complex and requires a very large data corpus and much computer time to train, but their examples illustrate output that is virtually indistinguishable from human speech. Unfortunately, expressiveness remains a challenge for this technology. Neural TTS systems can learn to express anything that is present in their training data but generalizing beyond seen expressive modes is an area of active ongoing research (e.g., Skerry-Ryan et al., 2018; see examples at https://bit.ly/30epgeW).

Neural TTS systems come at substantial expense both in terms of the amount of data that is needed and in the computational resources to train the models. Many are currently so resource heavy that they are only usable by well-equipped industry or university laboratories. However, there are elements of this work that are already having an impact, notably the neural vocoder programs, which produce highly natural-sounding speech output given the correct input. It may take a very large amount of data and heavy server load to train these vocoders, but once trained, they can be used with Mel spectrograms generated by many other applications and are able to run in real time on desktop-grade computers.

## Conclusions

The path from rule-based formant synthesis in the 1980s to the DNN voices being studied in research laboratories today represents significant growth in TTS technology. This growth has been followed through the lens of how the improvements impact one of the potentially most exciting applications of TTS technology: its potential to provide unique personal voices for people who are unable to communicate vocally without assistance. A notable subset of the potential users of TTS technology are those whose speech is at risk of being lost due to disease or injury. For those users, the ability to bank their existing speech for its use later in as a personal TTS voice of the quality now emerging form the laboratory is a highly promising prospect.

We initially identified four features that seem to be of greatest importance to users for assistive voice technology: intelligibility, naturalness, identity, and expressivity. Of these four, the first three are essentially solved problems, at least for laboratory-grade neural TTS systems. Given the rate of progress with the technology, it seems likely that for these three features, medical and consumer applications will not be long in coming. Expressivity, however, remains the largest unsolved issue for TTS systems. Parametric synthesis affords the ability to control features known to relate to expressive modes of speaking, and it will be fascinating to see how natural language processing (NLP) may end up helping users quickly find the right emotion to convey along with their text when it is spoken aloud.

### References

Bunnell, H. T., Hoskins, S., and Yarrington, D. (1998). A biphone constrained concatenation method for diphone synthesis. *Proceedings of the Third International Workshop on Speech Synthesis*, Jenolan Caves, Blue Mountains, NSW, Australia, November 26-29, 1998, pp. 171-176.

Bunnell, H. T., Pennington, C., Yarrington, D., and Gray, J. (2005). Automatic personal synthetic voice construction. *Proceedings of the Eurospeech 2005*, Lisbon, Portugal, September 4-8, pp. 89-92.

Donahue, J., Dieleman, S., Bińkowski, M., Elsen, E., and Simonyan, K. (2020). End-to-end adversarial text-to-speech. Available at https://bit.ly/3C7jVmI. Accessed November 12, 2021.

Greene, B. G., Logan, J. S., and Pisoni, D. B. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. *Behavior Research Methods, Instruments, & Computers* 18(2), 100-107.

Griffin, D., and Lim J. (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing* 32(2), 236-243. https://doi.org/10.1109/TASSP.1984.1164317.

Klatt, D. H. (1980). Software for a cascade/parallel synthesizer. *The Journal of the Acoustical Society of America* 67, 971. https://doi.org/10.1121/1.38940.

Klatt, D. H. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America* 82, 737-793.

Morise, M., Yokomori, F., and Ozawa, K. (2016). WORLD: A vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Transactions on Information and Systems* E99-D(7), 1877-1884.

Moulines, E., and Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9, 453-467.

Murray, I. R., and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America* 93(2), 1097-1108.

Pullin, G., and Hennig, S. (2015). 17 ways to say yes: Toward nuanced tone of voice in AAC and speech technology. *Augmentative and Alternative Communication* 31(2), 170-180.

Ramsay, G. (2019). Mechanical speech synthesis in early talking automata. *Acoustics Today* 15(2), 11-19.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the 2nd International Conference Spoken Language Processing*, Banff, AB, Canada, October 13-16, 1992, pp. 867-870.

Skerry-Ryan, R. J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., Weiss, R., Clark, R., and Saurous, R. A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. *Proceedings of the International Conference on Machine Learning*, Stockholm, Sweden, July 10-15, 2018. Available at https://bit.ly/3CgXvPU. Accessed November 12, 2021, pp. 7471-7480.

Stevens, S. S., Volkmann, J., and Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America* 8(3), 185-190.

van den Oord, A. Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*. Available at https://bit.ly/3qtNrkm. Accessed November 12, 2021.

Zen, H., Tokuda, K., and Black, A. W. (2009). Statistical parametric speech synthesis. *Speech Communication* 51(11), 1039-1064.

### About the Author

**H. Timothy Bunnell**
tim.bunnell@nemours.org

*Nemours Children's Hospital, Delaware Center for Pediatric Auditory and Speech Sciences*
*1701 Rockland Road*
*Wilmington, Delaware 19803, USA*

**H. Timothy Bunnell** is the director of the Center for Pediatric Auditory and Speech Sciences (CPASS) at the Nemours Children's Hospital, Delaware, Wilmington; head of the Speech Research Lab in the CPASS; and an adjunct professor of Computer and Information Sciences at the University of Delaware, Newark. He received his PhD in experimental psychology in 1983 from The Pennsylvania State University, University Park; served as research scientist at Gallaudet University, Washington, DC, from 1983 to 1989; and joined Nemours Children's Health to found the Speech Research Laboratory in 1989. His research has focused on the applications of speech technology for children with hearing and speech disorders.

# Human Hearing in the Underwater Environment

*Brandon M. Casper and Matthew A. Babina*

Hearing is a key sense that informs us about our environment. The cues we obtain from sounds grab our attention, allow us to communicate, and warn us of danger. Human hearing has evolved to detect sounds in air. As a result, anyone who has tried snorkeling or Scuba diving or have put their head underwater in a bathtub has noticed that the world sounds very different. With ears underwater, sounds seem quieter, as though the listener has cotton stuffed in their ears. Moreover, in air, when one hears a sound, one can usually tell if it is coming from the left or right and, to a lesser degree, if it is from the front or back. Underwater, although a diver can hear a boat's engine, identifying where the sound is coming from is challenging. This is because early terrestrial vertebrates evolved to hear well in air, and these adaptations are not the same as those needed for the underwater hearing abilities possessed by aquatic ancestors.

It makes evolutionary sense that human in-air hearing is better than their underwater hearing. Nevertheless, human underwater hearing may not be quite as bad as you think. The goal of this article is to introduce the field of human underwater hearing and to touch on several aspects of the topic that investigators have explored over the last century. It includes discussions of the mechanism of hearing underwater, underwater hearing thresholds, sound localization, and concerns about noise exposure and potential hearing damage. This article presents a broad overview of peer-reviewed literature and government technical reports.

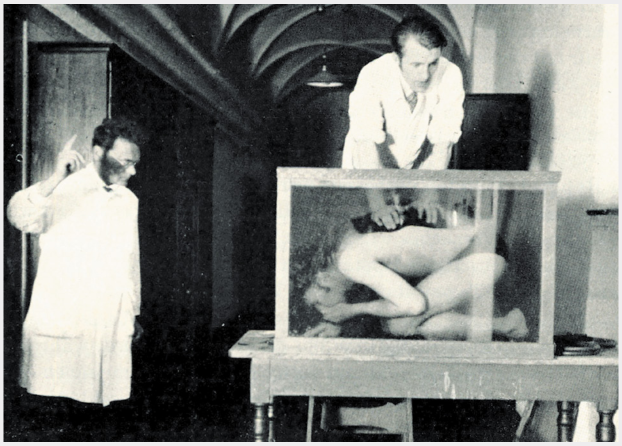
## Of Minnows and Men

Stetter (1929), a well-known German investigator of fish hearing, published a famous image of a research subject submerged in a clear-sided tank (**Figure 1**). Stetter's experiment compared the underwater hearing ability of humans with that of the common minnow (*Phoxinus laevis* L.) to a 662-Hz tone produced by a whistle. To adjust the level of the signal, another experimenter walked up and down a hall outside of the room blowing the whistle. The subject would move their finger if they could hear the sound. Details on the minnow testing were not as clear in the paper, but the conclusion reached by the investigators was that the minnow's hearing was much more sensitive than the human's. This is possibly the first article on human underwater hearing.

## How Do We Hear Underwater?

From this early study, it was established that minnows and many other fishes have more sensitive underwater hearing than humans. Why is this? Keep in mind that vertebrate hearing evolved in the earliest fishes to function in water (Fay and Popper, 2000). Once early vertebrates came onto land, they could not hear unless

**Figure 1.** *Classic image depicting possibly the first experiment in human underwater hearing. While Stetter is keeping the subject submerged (**right**), another scientist in a different room is blowing a whistle while moving closer and further from the subject. Karl Von Frisch (**left**), later winning the Nobel Prize, is observing the subject's responses. From Stetter (1929), used with permission.*



https://doi.org/10.1121/AT.2022.18.1.23

they adapted to the terrestrial environment. Let's "dive" deeper into what is going on between these two different environments and their effects on these auditory systems.

The most important differences between air and water in this context are their relative density and compressibility that, when combined, define the acoustic impedance of these two fluids. The acoustic impedance of the human head is very similar to that of water, which is unsurprising because most human soft tissues are close to 80% water. When surrounded by air, the high acoustic impedance of our heads reflects most sound energy, whereas underwater, sound travels through our heads instead of being reflected off them. Unfortunately, this removes the ability of the outer ear to "catch" and focus sound onto the tympanic membrane (eardrum).

Furthermore, the tympanic membrane and ossicles (middle ear bones) normally match the acoustic impedance of air-conducted sound and transmit the vibration to the fluid-filled cochlea. When stimulated via this sound path, the ossicular vibration produces a displacement wave in the fluid of the cochlea. Underwater, this traditional pathway is ineffective because sound energy transmission would have to travel from water (ear canal) to air (middle ear) and back to fluid (inner ear). Instead, sound energy is conducted through the skull directly to the ossicles and cochlea.

Like the human head underwater, the minnow's body is also "acoustically transparent." Fish ears have dense otoliths in contact with the sensory hair cells of the auditory region of the ear. As sound travels through the minnow's body, there is a relative lag between the motion of the dense otolith and surrounding tissues. This results in the ciliary bundles of the sensory cells being "bent" and therefore stimulated, allowing the minnow to hear the sound.

Humans do not have otoliths. Without the sound energy being transmitted through the traditional lower impedance pathway, the displacements produced in the cochlea of the human inner ear are much smaller than the sensory organ had evolved to detect. Smaller displacements mean less stimulation of the sensory hair cells and reduced hearing sensitivity, as discussed in **Underwater Hearing Thresholds**.

Preliminary evidence for this underwater acoustic pathway came from studies by Wainwright (1958) who had divers plug up their ears with their fingers. The divers were still able to detect sounds, although it was later pointed out that the bones in the fingers could still be transmitting the sound to the cochlea and the tissue of the finger would also be acoustically transparent (Smith, 1969).

Later, Hollien and Brandt (1969) had divers wear ear plugs underwater. Interestingly, the investigators had the divers put the ear plugs in prior to submersion, thereby trapping the air within the ear canal. In theory, this would eliminate the impedance mismatch around the tympanic membrane, which it did, but it ultimately just moved the mismatch of the air/water interface to the location of the ear plugs. Regardless, hearing thresholds were no different between tests with and without ear plugs, supporting the direct inner ear stimulation hypothesis.

Further evidence for direct inner ear stimulation comes from a study by Smith (1969). Smith tested underwater hearing thresholds in divers with known impaired in-air hearing but normal in-air bone conduction thresholds. In-air bone conduction hearing bypasses the outer and middle ear, so Smith was comparing whether air-conducted or bone-conducted thresholds better predicted the divers' underwater thresholds. The results from the underwater testing revealed no evidence of raised underwater hearing thresholds regardless of the divers' in-air hearing thresholds.

Hollien and Feinstein (1975) then tested diver hearing with three scenarios: (1) bare headed, (2) wearing a neoprene dive hood, and (3) wearing a neoprene hood with rubber tubes inserted into the ear canal through holes in the hood. As discussed in **Underwater Noise Exposure and Hearing Conservation**, neoprene is an effective blocker of sound transmission, especially at frequencies above 500 Hz. In the Hollien and Feinstein study, the divers' hearing thresholds were significantly higher in scenarios 2 and 3 where the hood reduced direct inner ear stimulation and tubes to the ear canal had no effect on hearing thresholds.

In summary, humans can hear underwater but not through the traditional in-air hearing pathway. There is evidence that sound transmission underwater to the cochlea is occurring directly through the skull, but what kind of impact does this have on human hearing sensitivity?

## Underwater Hearing Thresholds

Defining the average underwater hearing threshold is challenging. One issue is that most human testing has been conducted with just a handful of subjects per experiment due to the complexities of testing underwater. Still, human underwater hearing has been tested by at least eight different research teams, building a foundation of information on hearing thresholds in divers. At the same time, the only thing that all the conclusions constructed by these teams have in common is that they are different.

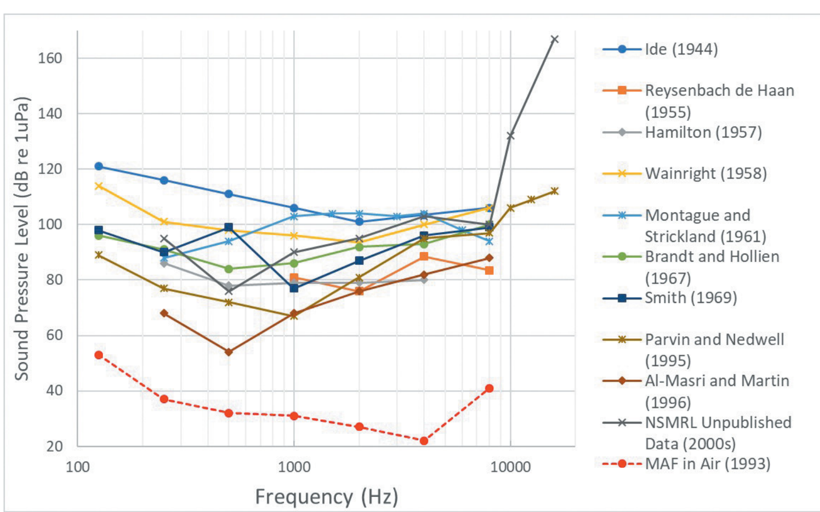
Two methodological approaches are consistent among all the studies. (1) Breath holding by the subjects was done to reduce noise during sound presentations, and (2) underwater hearing was measured as minimum audible field (MAF) audiograms. This means that the subjects are facing a sound projector and detecting and responding to the free sound field to which their heads are exposed. In most cases, this sound field is then calibrated by placing a hydrophone where the location of the head would be to measure the sound level. The calibration procedures are challenging underwater because creating an anechoic (nonreflective) environment is nearly impossible. There is also the problem of removing or limiting environmental noises from the test site.

The first underwater testing of multiple frequencies to measure human hearing was conducted by Ide (1944) of the United States Naval Research Laboratory (**Figure 2**). The methods were not detailed and the background noise as well as the sound measurements at the swimmers' heads were not reported, making these data challenging to interpret.

Several studies were then conducted in the 1950s and early 1960s, with threshold results between experiments varying 10-15 dB at each frequency (**Figure 2**) (Hamilton, 1957; Wainwright, 1958; Montague and Strickland, 1961). Although each team used similar approaches to measure hearing, the likely reasons for these equivocal results are background noise in testing environments, insufficient data on the subjects' in-air air conduction and bone conduction thresholds, and challenges with calibrating the sound fields underwater. Thus, although these studies began to map out human underwater hearing abilities, uncertainty remained about the sensitivity and frequency range. This uncertainty was partially alleviated by two research groups that emerged in the late 1960s and early 1970s, both of which advanced understanding of how humans hear underwater.

The first group formed at the Communication Sciences Laboratory at the University of Florida, Gainesville, and was led by Harry Hollien and included John Brandt and Stephen Feinstein among others. In 1967, the group built their Diver Communication Research System (DICORS) to conduct standardized, calibrated testing of human hearing and communication under water (Brandt and

**Figure 2.** *Human underwater hearing thresholds. NSMRL, Naval Submarine Medical Research Laboratory; MAF, minimum audible field. MAF data from International Standards Organization (1993). Figure modified from Al-Masri and Martin (1996).*



Hollien, 1967). This system included a seat to keep a weighted diver in place, transducers (speakers) at known distances, calibrated hydrophones at the location of the diver, and mechanisms for the diver and researchers on the surface to communicate.

The team conducted much of their research at the Underwater Sound Reference Division of the Naval Research Laboratory, Orlando, Florida, test facility at Bugg Spring, which was an extremely quiet, nonreverberant testing environment. Due to the controlled testing environments, the hearing thresholds obtained by this team were long considered one of the gold standards for underwater hearing in humans. Although their data (Brandt and Hollien, 1967) were not that different from the data obtained in previous studies (**Figure 2**), the rigorous testing procedures and quiet location supported the accuracy of their results. They also tested whether water depth affected thresholds but failed to find a significant difference of thresholds for depths ranging between 3.7 m and 32 m (Hollien and Brandt, 1969).

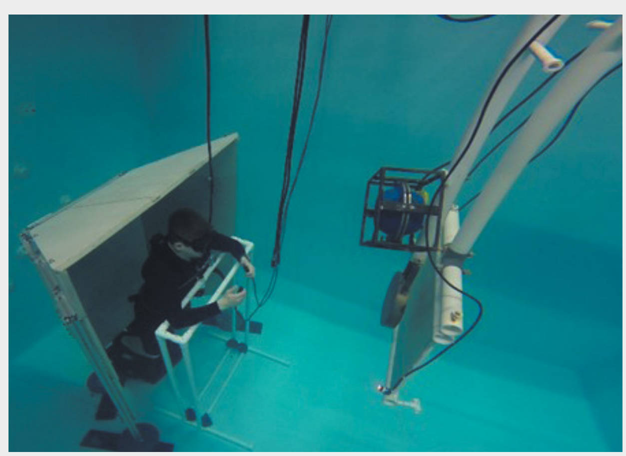
The other key research team that worked on underwater hearing starting in the 1960s and lasting through the 1990s was led by Paul Smith of the Naval Submarine Medical Research Laboratory (NSMRL) at the New London Submarine Base in Groton, Connecticut. Smith's efforts in the underwater realm kicked off research that continues at the NSMRL today, covered everything from underwater hearing thresholds to diver aversion to sound, and also produced early recommendations on hearing conservation.

Smith's (1969) underwater hearing threshold testing was the first to include examination of bone conduction thresholds in air. This was critical because the pathway for bone conduction in air appears to mirror the underwater direct inner ear stimulation. Smith recruited US Navy subjects with normal air conduction and bone conduction hearing as well as some with reduced air conduction and bone conduction responses. The study was done in a deep, quiet pond (75-80 ft) with the subjects in the middle of the pond at a depth of 4.5 m. Like Hollien's team, Smith built a platform that housed the diver, transducer, and reference hydrophones all at fixed locations. The underwater hearing thresholds obtained in this study matched those in Brandt and Hollien (1967) (**Figure 2**), establishing the importance of running experiments in quiet environments. An interesting finding was that the divers with reduced bone conduction thresholds also had reduced underwater hearing thresholds, further supporting the direct inner ear stimulation hypothesis.

Following Hollien and Smith, beginning in the 1990s and extending to today, two other groups entered the underwater hearing field. The initial studies were conducted by Mohammad Al-Masri, University of Portsmouth, Portsmouth, United Kingdom, in 1993 (reviewed by Al-Masri and Martin, 1996) and then carried forward by Parvin and Nedwell (1995) through the rest of the 1990s. These teams built on the lessons learned from previous research, creating as quiet an environment as possible, and reducing the ambient levels in their test tank to ~44 dB re 1 μPa by acoustically isolating the tank from the surrounding laboratory environment (compared with ~60 dB re 1 μPa in Smith's [1969] experiments). In addition, the investigators characterized the sound levels in the tank environment so that the level at the diver's head was as well defined as possible. They also conducted in-air hearing tests on all divers (mix of Navy and recreational) to confirm that they had "normal" hearing. All these efforts resulted in underwater hearing thresholds that were significantly lower (15-20 dB lower at many frequencies) than any measured previously (**Figure 2**).

The second group beginning to work in this field in the 1990s was out of the NSMRL (Fothergill et al., 2002, 2018). This program reinvigorated the underwater hearing research that Smith had started in the 1960s but focused on concerns of US Navy divers being exposed

**Figure 3.** *Diver participating in an underwater hearing test in the NSMRL dive pool.*



to low-frequency sonar. The reimagined team was led by Ed Cudahy and gained momentum in the 2000s. Cudahy and his team often conducted hearing studies at the Naval Undersea Warfare Center Newport Dodge Pond Acoustic Test Facility. This testing environment is known for its low ambient noise and minimal reflection except from the surface. Although the NSMRL underwater thresholds were not quite as sensitive as those in the United Kingdom (**Figure 2**), procedures established by Cudahy's team are still in use today (**Figure 3**) as we continue to expand knowledge of underwater hearing in humans.

Between the two groups, over 100 divers (mix of Navy and recreational) were tested, resulting in the largest sample size of divers measured. The underwater thresholds from the United Kingdom resulted in significantly lower thresholds detected at many frequencies (**Figure 2**) and ultimately became the benchmark for underwater hearing thresholds. These measurements still apply to today's guidance. These increased sensitivities are likely due to the emphasis placed on lowered ambient-noise levels in the testing environments.

Combining the results of the body of research on hearing thresholds, we can draw some conclusions. Overall, it appears that there is around a 30-60 dB increase in sensitivity between equivalent air and water thresholds (**Figure 2**). There is somewhat of a U-shaped threshold curve, with thresholds increasing fairly quickly above 10 kHz. These studies have established average thresholds for a frequency range from 250 Hz to 16 kHz, showing greatest sensitivity between 500 and 1,000 Hz.

In summary, the most important finding from the studies described here is that human hearing underwater is much less sensitive than in air. Many researchers have measured human underwater hearing thresholds. Although the methodologies used were fairly similar, the results vary. Just as in air, measuring hearing in as quiet a place as possible is critical when testing near the threshold of hearing. As researchers learned to minimize ambient noise and refine their techniques, they expanded their knowledge of the range and sensitivity of human underwater hearing.

## Where Is That Sound Coming From?

In air, humans use several cues to identify the direction of a sound. Two of the critical cues are interaural time difference (ITD) and interaural level difference (ILD). The ITD is defined as the time interval between when a sound is perceived by one ear versus the other ear, and the ILD is the difference in loudness between the two ears. Both features take advantage of the acoustic shadowing provided by the head. After reaching one ear, sound must travel around the head before reaching the other ear. The human auditory system is sensitive enough to process these differences in time of arrival and loudness to determine the direction of sounds. This is a simplified explanation of the process; in actuality, humans use additional cues to refine the ability to determine direction (Middlebrooks and Green, 1991).

When submerged, sound travels through the head instead of going around like it does in air. Furthermore, sound travels about 340 m/s in air and 1,480 m/s in water, and so the sound reaches both ears so close in time that the brain cannot differentiate between arrival times. In combination, these differences effectively eliminate directional cues. Without the directional cues of ILD and ITD, sound should appear to be coming from all directions equally. Humans certainly cannot localize sound in water as effectively as they can in air, but with enough practice, they are not completely lost underwater either.

Feinstein (1973) ran a series of studies measuring minimum audible angles (MAAs) for divers to test the ability to discriminate sounds coming from different directions. The MAA is the smallest angular separation at which two sounds are perceived as coming from distinct directions. Once the sounds originate closer to one another than the MAA, the listener perceives the sounds as coming from the same location.

Feinstein (1973) had divers wearing neoprene hoods with holes at the ears sit on a custom-built platform that kept their head in a fixed position. Two speakers were set up in a way that allowed them to be offset from each other by a known angle of separation. The diver would pull one of two ropes to signal if the sound was coming from the left or right speaker. The stimuli were either a 3.5-kHz tone, a 6.5-kHz tone, or white noise. The MAA for each stimulus was 21.5°, 14.5°, and 9.8°, respectively. A second study provided training to the divers by letting them know when they made a mistake. Following the training, the divers improved to 11.3°, 11.5°, and 7.3°, respectively. Feinstein determined that sound localization underwater is on average around three times poorer than comparable studies conducted in air.

More directional hearing studies were conducted by a French team led primarily by Sophie Savel (Savel and Drake, 2014). They found that lower frequency sounds and white noise were easier to discriminate than higher frequencies. Divers were able to identify angles to the left and right successfully but had severe challenges with front/back discrimination. They did find that divers in general were more successful at all localization studies with experience, something that Feinstein (1973) also noticed. This included experience and training with the experiment cues as well as general diving experience (i.e., total number of career dives).

Interestingly, in one of their studies, Savel et al. (2009) had divers wear neoprene hoods with holes cut around the ears. They also plugged the ear canal with homemade neoprene ear plugs. When the ears were plugged with neoprene, the divers' ability to localize sound dropped significantly, suggesting that the ear conduction pathway could play a role in sound localization underwater. The authors postulated that a phase difference at the cochlea between the arriving direct inner ear stimulated sound and ear conducted sound could provide some directional cue. This hypothesis needs further investigation, but Savel et al. are not the first to notice a drop in localization capabilities when the ear canals are blocked by neoprene (Norman et al., 1971).

## Underwater Noise Exposure and Hearing Conservation

There are many sources of underwater biological sounds ranging from marine mammals to fishes and invertebrates, although there has been no record of any of these sounds being of obvious concern to human hearing. Rather, anthropogenic or human-made sounds underwater are the primary sources of concern.

There is one kind of noise that divers cannot avoid: the sound of their own breathing. The bubbles produced during respiration in Scuba and surface-supplied air are quite noisy. Indeed, this is the reason why divers are required to breath-hold during the hearing tests. The bubble noise is not dangerous to humans, but it is not quiet. This was one of the reasons for the development of the rebreathing system.

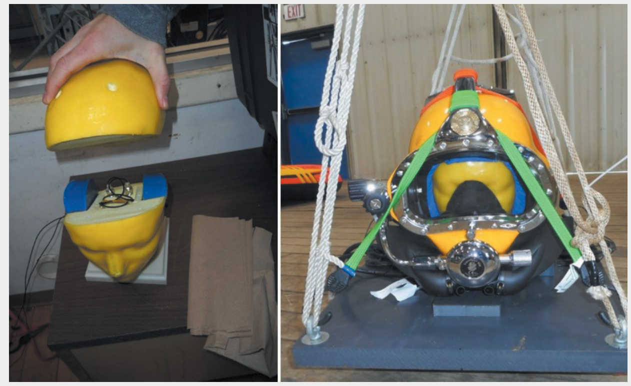


**Figure 4.** *NSMRL head simulant (Ned Stark;* ***left****) used for testing sound transmission in the Kirby Morgan 37 dive helmet (***right***).*

Breathing noise is also a concern in helmeted divers. The NSMRL and others in the US Navy (Curley and Knafelc, 1987) have documented that the sound levels measured during inhalation and exhalation in the standard Kirby Morgan dive helmets (**Figure 4**) often exceed the traditional 85 dBA hearing safety limits. These helmets are the current standard for working divers. Divers also use a valve to blow air into the helmet to defog the faceplate, which greatly exceeds the limits. An additive effect is created by the communication system within the helmet. Divers must often turn the sound level up to effectively hear and communicate with people on the surface.

Looking beyond diver-produced noise, there are many external anthropogenic sources of sound that could potentially impact divers (e.g., underwater explosions, tool noise, pile driving, sonar, or boat noise). Any of these sources could generate high levels of acoustic energy. Knowing that hearing underwater is different than in air, how does one determine what is safe or unsafe in terms of human exposure? This is where the problem gets challenging!

We now assume for the sake of a discussion that the divers have wet ears (i.e., nonhelmeted). The two primary challenges for providing safety guidance underwater are the lack of personal protective equipment and the differences in underwater hearing abilities compared with in-air hearing.

Let's start by talking about the types of protection that exist. Wearing earplugs or any kind of over-ear sound protection has little value. First, earplugs block the ear canals that can compromise the diver's ability to equalize the pressure in the middle ear as they move up and down in the water. More importantly, if humans detect sound through direct inner ear stimulation, then blocking the traditional air conduction pathway is not an effective means for preventing noise-induced hearing damage.

Currently, the only effective method of hearing protection underwater is the aforementioned neoprene hood. Numerous studies (reviewed by Fothergill et al., 2018) have characterized the effectiveness of a neoprene hood at increasing hearing thresholds in divers. As mentioned in **Underwater Hearing Thresholds**, a hood is effective at attenuating frequencies at 500 Hz and above, with the amount of attenuation increasing with frequency (as much as 20-30 dB of attenuation), although some of these effects are reduced with increases in pressure (Fothergill et al., 2018).

Therefore, divers exposed to certain SONAR systems or any tool that produces a lot of high-frequency energy can be protected from hearing damage by wearing a neoprene hood. However, just about all the noise sources that were mentioned create broadband signals with a lot of low-frequency energy below 500 Hz in addition to high frequencies. Thus, a hood provides little-to-no protection from much of the acoustic energy from many underwater tools, explosives, pile driving, and boat noise.

Moreover, in many operations with underwater tools, divers wear helmets so that they can have dry ears. The NSMRL recently completed a data collection to explore the energy transfer function of the Kirby Morgan dive helmets (**Figure 4**). The measurements show that frequencies down to at least 50 Hz are attenuated by the helmet. Again, attenuation increases with frequency, and in the case of the helmet, there is a dip in attenuation at the resonance frequency of the helmet. Although the helmet provides more attenuation than the hood, the sound is being delivered via the more sensitive airborne mechanism of hearing. Therefore, the net effect at most frequencies is that the recommended exposure limits with a helmet will be lower than with a hood. This is especially true at lower frequencies. Thus, divers have few options for underwater hearing protection. Safety guidance for divers exposed to underwater noise must therefore account for the limitations in personal protection equipment.

The in-air community has a wealth of human and animal studies that determined the upper limit of exposures that would induce hearing damage, such as temporary threshold shifts (TTSs) and permanent threshold shifts (PTSs) in hearing (reviewed by Clark 1991; Melnick 1991). TTS is defined as a temporary loss of hearing sensitivity after exposure to sound. Hearing conservation standards for in-air noise consider the onset of TTS as defining the upper limit of safe noise exposure. PTS is a shift in hearing sensitivity at a frequency or range of frequencies that does not resolve with time.

Unfortunately, data for the underwater onset of the TTS are extremely sparse. Only a few studies have been conducted on this topic, and the results, although incredibly valuable, are challenging to interpret due to the small sample size, high variance among divers, and challenges associated with measuring hearing immediately postdiving (reviewed by Smith et al., 1988). Additional studies later conducted by investigators at the NSMRL and in the United Kingdom attempted to measure diver aversion to low frequencies up to 2,500 Hz (reviewed by Fothergill et al., 2002).

To establish an international hearing conservation limit for divers, the United Kingdom and NSMRL worked together in the early 2000s to merge the extensive United Kingdom hearing threshold data with the underwater TTS and aversion data. Most of these studies and the underwater noise guidance for divers are not publicly available so they cannot be discussed in any detail. However, suffice it to say that this guidance is used consistently and has proven effective in protecting divers.

In addition to providing guidance for noise exposure related to hearing concerns, the NSMRL also works with organizations that are involved with underwater explosives (UNDEX). These communities are typically concerned with injuries to the lungs and other air-filled structures. There is established safe standoff guidance for underwater blasts, and the NSMRL continues to explore how to improve on and expand the guidance. Obviously, investigators cannot knowingly expose divers to the UNDEX to establish injury data, so instead physical model simulants (**Figure 5**) have been developed to better understand the injury mechanisms associated with blast exposure. Another entire article could be written on underwater blast injuries and the research associated with the protection of divers so, for now, we direct readers to Cudahy and Parvin (2001) as an excellent primer on the topic.

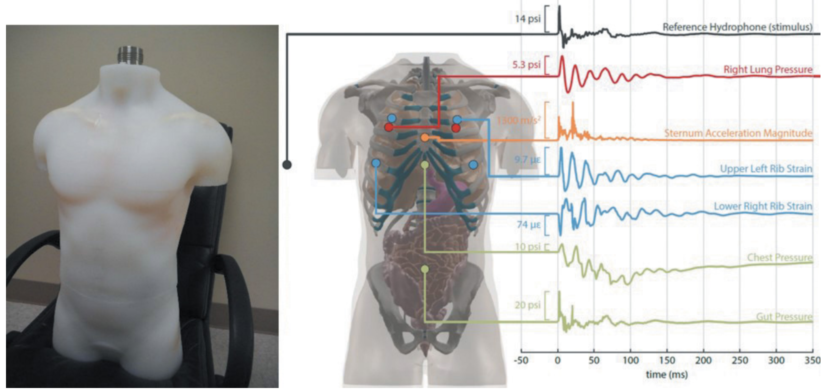



**Figure 5.** *NSMRL human torso simulant (QUantitative Instrumented Torso [QUINT];* ***left****) used to predict impact of high-energy, impulsive sources on internal organs (****right****).*

## Conclusions

We have presented an overview of the field of underwater hearing. Although not every topic could be covered in detail, our goal was to provide a general understanding of how human hearing underwater is different than that in air and what humans' underwater hearing capabilities are. Exposure to loud sounds underwater is a concern for divers, and in many situations, the limited hearing protection that is available is not effective. The NSMRL continues to conduct research to improve safety guidance for Navy divers, but our efforts would not be possible without standing on the shoulders of all the scientists who established the field before us.

## Acknowledgments

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the US Government. We are employees of the US Government. This work was prepared as part of our official duties. Title 17 U.S.C. §105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 U.S.C. §101 defines a US Government work as a work prepared by an employee of the US Government as part of that person's official duties.

### References

Al-Masri, M., and Martin, A. (1996). Underwater hearing and occupational noise exposure. In Axelsson, A., Hellstrom,P., A., Borchgrevink, H. M., Henderson, D., Hamernik, R. P., and Salvi, R. J. (Eds.), *Scientific Basis of Noise-Induced Hearing Loss*. Thieme Medical Publishers, New York, NY, pp. 119-133.

Brandt, J. F., and Hollien, H. (1967). Underwater hearing thresholds in man. *The Journal of the Acoustical Society of America* 42, 966-971.

Clark, W. W. (1991). Recent studies of temporary threshold shift (TTS) and permanent threshold shift (PTS) in animals. *The Journal of the Acoustical Society of America* 90, 155-163.

Cudahy., E., and Parvin, S. (2001). *The Effects of Underwater Blast on Divers.* Technical Report 1218, Naval Submarine Medical Research Laboratory, Groton, CT.

Curley, M. D., and Knafelc, M. E. (1987). Evaluation of noise within the MK 12 SSDS helmet and its effect on divers' hearing. *Undersea Biomedical Research* 14, 187-204.

Fay, R. R., and Popper, A. N. (2000). Evolution of hearing in vertebrates: The inner ears and processing. *Hearing Research* 149, 1-10

Feinstein, S. H. (1973). Acuity of the human sound localization response underwater. *The Journal of the Acoustical Society of America* 53, 393-399.

Fothergill, D. M., Cudahy, E. A., Schwaller, D. W., Townsend, O., and Qin, M. K. (2018). *Psychophysical Measurements of a Neoprene Wetsuit Hoods Sound Attenuation as a Function of Dive Depth and Frequency: Hyperbaric Chamber Trials*. Technical Report 1325, Naval Submarine Medical Research Laboratory, Groton, CT.

Fothergill, D. M., Schwaller, D., Forsythe, S. E., and Cudahy, E. A. (2002). *Recreational Diver Responses to 600-2500 Hz Waterborne Sound*. Technical Report 1223, Naval Submarine Medical Research Laboratory, Groton, CT.

Hamilton, P. M. (1957). Underwater hearing thresholds. *The Journal of the Acoustical Society of America* 29, 792-794.

Hollien, H., and Brandt, J. F. (1969). Effect of air bubbles in the external auditory meatus on underwater hearing thresholds. *The Journal of the Acoustical Society of America* 46, 384-387.

Hollien, H., and Feinstein, S. (1975). Contribution of the external auditory meatus to auditory sensitivity underwater. *The Journal of the Acoustical Society of America* 57, 1488-1492.

Ide, J. (1944). *Signalling and Homing by Underwater Sound; for Small Craft and Commando Swimmers*. Sound Report 19, Naval Research Laboratory, Washington, DC.

International Standards Organization (1993). *Acoustics-Equal Loudness Level Contour for Otologically Normal Listeners, Part 1: Reference Threshold of Hearing Under Free-Field and Diffuse Field Listening Conditions.* ISO/CD 226-2, International Standards Organization, Geneva, Switzerland.

Melnick, W. (1991). Human temporary threshold shift (TTS) and damage risk. *The Journal of the Acoustical Society of America* 90, 147-154.
Middlebrooks, J. C., and Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology* 42, 135-159.
Montague, W. E., and Strickland, J. F. (1961). Sensitivity of the water-immersed ear to high-and low-level tones. *The Journal of the Acoustical Society of America* 33, 1376-1381.
Norman, D. A., Phelps, R., and Wightman, F. (1971). Some observations on underwater hearing. *The Journal of the Acoustical Society of America* 50, 544-548.
Parvin, S. J., and Nedwell, J. R. (1995). Underwater sound perception and the development of an underwater noise weighting scale. *Underwater Technology* 21, 12-19.
Reysenbach de Haan, F. W. (1956) Hearing in whales. *Acta Oto-Laryngologica Suppl.* 134, 1-114.
Savel, S., and Drake, C. (2014). Auditory azimuthal localization performance in water as a function of prior exposure. *Human Factors* 56, 772-783.
Savel, S., Drake, C., and Rabau, G. (2009). Human auditory localisation in a distorted environment: Water. *Acta Acustica united with Acustica* 95, 128-141.
Smith, P. F. (1969). *Underwater Hearing in Man: 1. Sensitivity.* Technical Report 569, Naval Submarine Medical Research Laboratory, Groton, CT.
Smith, P. F., Wojtowicz, J., and Carpenter, S. (1988). *Temporary Auditory-Threshold Shifts Induced by Repeated Ten-Minute Exposures to Continuous Tones in Water*. Technical Report 1122, Naval Submarine Medical Research Laboratory, Groton, CT.
Stetter, H. (1929). Untersuchungen über den Gehörsinn der Fische, besonders von Phoxinds laevis L. und Amiurus nebulosus Raf. *Zeitschrift für vergleichende Physiologie* 9, 339-477.
Wainwright, W. N. (1958). Comparison of hearing thresholds in air and in water. *The Journal of the Acoustical Society of America* 30, 1025-1029.

## About the Authors

**Brandon M. Casper**
brandon.m.casper4.civ@mail.mil

*Warfighter Performance Department*
*Naval Submarine Medical Research Laboratory*
*Naval Submarine Base New London*
*141 Trout Avenue*
*Groton, Connecticut 06349, USA*

**Brandon M. Casper** is a research physiologist and department head of the Warfighter Performance Department in the Naval Submarine Medical Research Laboratory in Groton, Connecticut. His research focuses on the bioeffects of underwater sound and blast in human divers. He and his team collect data and provide guidance to the US Navy to protect divers from harm while also ensuring successful missions. He received his PhD from the University of South Florida, Tampa, where he studied the hearing abilities of sharks, skates, and rays. He was also a postdoc with a reasonably well-known scientist named Arthur Popper.

**Matthew A. Babina**
matthew.a.babina.civ@mail.mil

*Warfighter Performance Department*
*Naval Submarine Medical Research Laboratory*
*Naval Submarine Base New London*
*141 Trout Avenue*
*Groton, Connecticut 06349, USA*

**Matthew A. Babina** is a research engineer at the Naval Submarine Medical Research Laboratory in Groton, Connecticut, where he has worked since 2008. His multidisciplinary research activities have included underwater acoustics, hearing conservation, human performance modeling, psychophysics, and underwater blast pathophysiology. He completed his bachelor's degree in electrical and computer engineering and his master's degree in biomedical engineering at Worcester Polytechnic University, Worcester, Massachusetts. Currently, his work focuses on understanding the bioeffects of underwater sound and blast on humans, providing safety guidance recommendations to the US Navy.

# Listening to Mom: How the Early Auditory Experience Sculpts the Auditory Cortex of the Brain

*Patrick O. Kanold*

## Introduction

Our ability to understand a language is shaped by how we experience speech as a child. However, when auditory experience is important and how auditory experience acts on the different parts of the brain have been unresolved. In particular, our experience with sounds starts before we are born, and many expecting parents wonder if early exposure to music or other stimuli can influence their developing child. Underlying our ability to hear is the precise wiring or circuitry between neurons in the brain. Auditory processing involves many interconnected structures, including the most complex auditory part of the brain, the auditory cortex. This is the region of the brain that is essential for the processing of complex sounds such as speech and music (Wang, 2018). Results from animal studies have started to reveal the influence of early sound exposure on circuits in the auditory cortex (Meng et al., 2021). These studies indicate that early sound experience, which in humans occurs in the womb starting around midgestation, already has the potential to shape auditory cortical circuits.

Thus, early sound experience or lack of sound experience, for example, in deafness, can potentially impact the brain before birth. Moreover, early insults to the developing brain (e.g., due to injury or exposure to drugs) might interfere with the early wiring processes, resulting in altered development. Moreover, these considerations are relevant for the care of prematurely born infants who are suddenly exposed to a different auditory environment in the neonatal intensive care unit (NICU).

## Effects of the Auditory Experience

To show how early experience can shape hearing, I trace the steps occurring in the early development of the auditory system and the influences of an early sensory experience on circuits in the auditory cortex. Hearing, or audition, is central to our ability to communicate. Underlying the ability to identify and distinguish sounds, such as phonemes in languages or the identity of speakers, is the precise wiring between neurons in the auditory system. Hearing is shaped by the experience with sounds, and the effect of this experience is the largest in early childhood. This is illustrated by the ease with which a second language can be learned in childhood versus in adulthood as those of us who learned a second or third language have experienced. Therefore, early exposure to the sounds of a particular language is crucial for perceiving subtle differences between words in that language. From this, it seems to follow that an enhanced auditory experience might be beneficial to the brain. Indeed, playing music or speaking during pregnancy has been popular; however, benefits of such enrichment are unknown. The critical questions to ask are when does the effect of sound experience start and which neurons and which brain circuits are influenced by the experience of sound?

Auditory processing in humans starts in utero, but the effects of fetal sound experience has long been debated. Many parents wonder if playing music or singing to their unborn child will enhance brain function. A variety of studies suggest that a sound experience shapes the fetal brain because fetuses or premature infants can distinguish speech sounds from nonspeech sounds and respond to maternal voices before term (40 gestational weeks [GWs]). For example, 35-GW-old fetuses have been shown to discriminate language (Minai et al., 2017) and newborns have a preference for the voice of their mother (DeCasper and Fifer, 1980) but not the father



https://doi.org/10.1121/AT.2022.18.1.32

(DeCasper and Prescott, 1984). Newborns also have cry melodies reminiscent of their maternal language (Mampe et al., 2009).

Because of their specificity, for example, of the maternal language, all of these abilities are not likely to be hardwired by genetic programs. Consequently, these experiments point to a fundamental effect of the fetus being exposed to the mother's voice and suggest that complex auditory processing is possible in humans before term birth. Moreover, the selectivity of the responses to the maternal voices indicate that a significant amount of circuit plasticity has occurred in the auditory pathway before birth to create neuronal circuits that allow the developing brain to distinguish the mother's voice from other voices. What has been unclear, however, is how such early sound exposure shapes the auditory pathway and which neurons and circuits are being influenced.

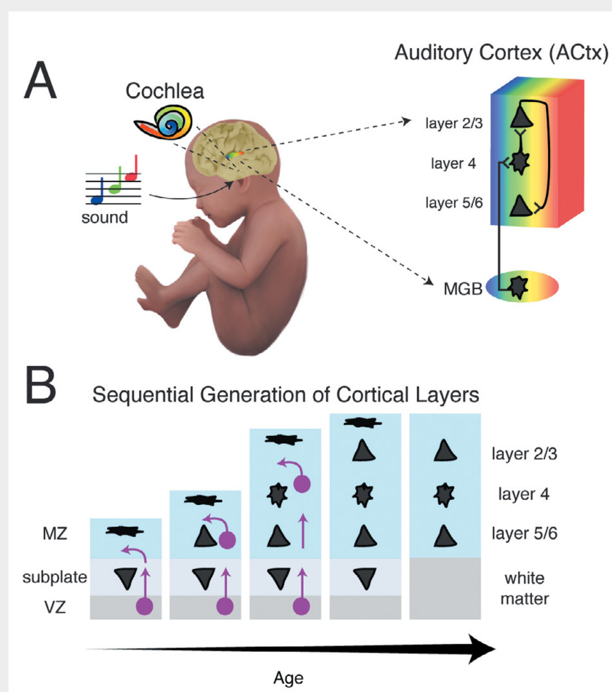
## Functional Organization of the Auditory System

The auditory system is organized in a hierarchical manner starting from the conversion of sound into neural impulses in the ear up to the complex analysis or the evoked neuronal activity patterns in central brain structures. Sound is transmitted through the ear canal and middle ear and then enters the inner ear, the cochlea, where it is converted into neural activity. Sound-evoked neural activity then propagates through a series of different brainstem and midbrain structures before reaching the auditory thalamus (medial geniculate body [MGB]) and finally the auditory cortex (ACtx; **Figure 1A**) (Budinger and Kanold, 2018). The ACtx is a key sound-processing region for many higher order processes such as the processing of complex stimuli like speech and music (Wang, 2018). The ACtx itself is composed of six layers of morphologically different neurons that are highly interconnected and are thought to perform the hierarchical processing of sounds (Budinger and Kanold, 2018).

One of the hallmarks of the functional organization of sensory cortices in the adult is the orderly organization of neurons responding to sensory features such as sound frequency across the cortical surface in that they form "maps" of the sensory space (Kaas, 2000). In the auditory system, cells respond selectively to a particular sound frequency and the orderly organization means that neighboring cells share frequency selectivity and that there is an orderly progression of frequency preference across the cortex (Schreiner and Winer, 2007).

Thus cells preferring low-frequency sounds (**Figure 1A**, *blue*) are located at one end of the ACtx, whereas cells that prefer high-frequency sounds (**Figure 1A**, *red*) are located at the other end of the ACtx, with cells that prefer midfrequency sounds (**Figure 1A**, *green*) in-between. The resulting map of sound frequency is called a "tonotopic map," and the orderly organization is thought to be important for normal brain function (Kaas, 1997). The tonotopic organization of the auditory system originates in the cochlea and requires precisely ordered projections

**Figure 1. *A:*** *hierarchical processing of sound from cochlea to auditory cortex (ACtx). The cochlea transduces sounds into neural impulses that are relayed to the auditory cortex via brainstem nuclei and the auditory thalamus (medial geniculate body [MGB]). Different parts of the cochlea respond selectively to different sound frequencies (**colors**). The orderly frequency map is preserved up to the ACtx. The ACtx contains different interconnected layers. Inputs to the ACtx from the MGB arrive in layer 4.* ***B:*** *sequential generation of cortical layers. Subplate neurons and cells in the marginal zone (MZ) are born before the permanent cortical layers. Newborn neurons in the ventricular zone (VZ;* ***purple****) migrate radially to their target layer and differentiate.*



between the different processing stages (Hackett et al., 2011). For example, cells in the low-frequency region of the MGB (**Figure 1A**, *blue*) project to one end of the ACtx (**Figure 1A**, *blue*) while cells in the high-frequency region of the MGB (**Figure 1A**, *red*) project to the other end of the ACtx (**Figure 1A**, *red*).

## The Development of "Hearing"

The development of the auditory system, and especially the ACtx, is a protracted process starting prenatally. Extensive work in animals has shown that the developmental process requires a complex interplay of genetic programs, spontaneous and sensory-driven neural activity; so both "nature" and "nurture" are heavily involved (Goodrich and Kanold, 2020).

The general sequence of processing stages between the ear and the ACtx is also present in development (Goodrich and Kanold, 2020), with one important exception. In early development, there is an additional specialized population of neurons, subplate neurons, that are present in the ACtx in early development before the MGB is connected to the ACtx (**Figure 1B**). These subplate neurons form early relay circuits connecting the MGB with the input layer of the ACtx (layer 4) and form a specialized developmental structure that provides a functional scaffold for the permanent wiring of the cortex. This review focuses on these specialized circuits, the events that can shape their function, and ways by which these circuits can influence later ACtx function.

In humans, physiological or neural responses to sound emerge around the end of the second trimester. A fundamental concept to define is the onset of hearing. Hearing has both sensory-processing and cognitive components because attention-based mechanisms can amplify or suppress sound-evoked responses, such as when ignoring background noises or attending to a particular instrument in an orchestra. For the purposes of this review, hearing means the onset of auditory processing and does not cover the cognitive aspects.

Auditory-processing development starts with the maturation of the cochlea and requires the neural transmission of sound-evoked neural activity to the brainstem and more central structures such as the ACtx. External sounds can be transmitted to the fetus but are attenuated by the womb, whereas sounds generated by the mother can be enhanced by conduction from the larynx to the body (Richards et al., 1992). Accordingly, the fetal environment is rich in potential auditory stimuli. Fetal movements in response to externally generated low-sound frequencies can be detected midgestation, at about the 19th GW, whereas responses to higher frequencies emerge later (Hepper and Shahidullah, 1994). Consequently, it can be reasoned that the human inner ear and, at least, the brainstem circuits must be functional at these ages, albeit likely not mature.

The more detailed development of auditory processing has been studied in animal models such as mice and ferrets that are born in a more immature state (altricial). Much of the development that happens in the womb in humans happens after birth in altricial animals. Furthermore, altricial animals undergo a major transition in their hearing in that they are born with closed ear canals that attenuate sounds and that open postnatally. Moreover, a major difference between altricial animals and humans is that the early sound environment in humans will be dominated by maternal sounds, whereas maternal sounds will be attenuated in altricial animals.

Indeed, although ear opening in altricial animals is sometimes called the "onset of hearing," data from multiple altricial species such as ferrets (Wess et al., 2017) and mice (Meng et al., 2020) show that auditory responses are present even at the level of ACtx at time periods when the ears are closed. Although sound-evoked responses can be recorded, it should be emphasized that these responses are not mature, and therefore neurons in young animals do encode sensory stimuli as robustly as the adult does. Together, these studies give us rough estimates of when peripheral sounds drive neural activity in the auditory system, but due to experimental limitations, it is possible that even earlier responses exist.

## Formation of the Auditory Cortex and Its Connections

The ACtx consists of six layers of neurons that are distinguished by differences in neuronal cell shape and connectivity (**Figure 1A**) (Budinger and Kanold, 2018). The major group of cortical neurons, excitatory neurons, are generated in the bottom of the cortex, and with each round of cell generation, a different layer is built (**Figure 1B**). Newborn neurons will move past older mature neurons and stack on top of each other; therefore, the cortex is built from the bottom up.

Subplate neurons are the earliest born cortical excitatory neurons and reside at the bottom of the cortex (Kostovic and Rakic, 1980). Moreover, subplate neurons, in contrast to other cortical neurons, mostly disappear during development (Luskin and Shatz, 1985) and form a transient population of deep neurons.

There is also another group of early-generated transient neurons on the outer margin of the cortex; hence, although the adult cortex contains six layers, the developing cortex contains additional largely transient neuronal layers at its deeper and outer margins (Molnár et al., 2020). The sequential generation of neurons is important for understanding the varying effects of developmental insults and injuries. Insults at younger ages will most directly affect early-born deep neurons, whereas insults at later ages will influence both superficial and deep neurons. This means that because deep and superficial neurons perform different functions, the same insults at different times can have distinct functional consequences.

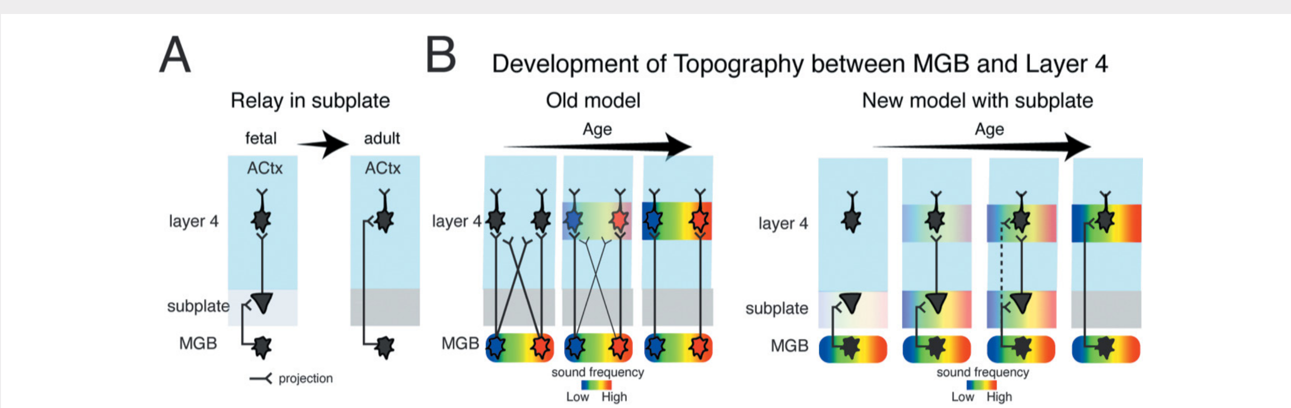
## Subplate Neurons Are the First Cortical Cells to Respond to Sounds and the Substrate of Early Topography

Neurons grow projections, called axons, and communicate via specialized structures called synapses, thereby forming neural circuits. The earliest generated transient neuronal layers are also the site of the early establishment of cortical synapses (Kostovic and Molliver, 1974). In adults, the input layer of the ACtx, layer 4, receives direct synaptic inputs from the MGB (**Figure 2A**) (Budinger and Kanold, 2018). This direct pathway is crucial for transmitting sound-evoked activity from the inner ear to the ACtx and thus is essential for auditory processing. In early development, this direct connection does not exist. Instead, MGB neurons first form synapses with subplate neurons (Kanold and Luhmann, 2010), and MGB axons remain constricted to the subplate for a period before growing to their eventual target in layer 4. Consistent with the early MGB inputs to the subplate, recordings in young animals have shown that subplate neurons respond to sound before ACtx layer 4 neurons (Wess et al., 2017).

During this time period, subplate neurons themselves project to the ACtx layer 4 (Zhao et al., 2009); thus subplate neurons form an essential relay for sound information to reach layer 4 and beyond (**Figure 2A**) (Wess et al., 2017). The direct transmission between the MGB and the ACtx layer 4 and thus the adult-like pattern emerge after ear opening (Barkat et al., 2011) and subplate neurons disappear during subsequent development (Kanold and Luhmann, 2010). These results also suggest that the early sound-evoked responses detected in human babies are due to subplate activation.

Although the ACtx contains the tonotopic map in adults, in early development, there is no map (**Figures 1A** and **2B**).

**Figure 2.** ***A:*** *subplate neurons relay ascending MGB activity to layer 4 in development, whereas in adults, the MGB directly activates layer 4.* ***B:*** *model of topographic mapping of frequency preference (**colors**) by ordered projections from the MGB to ACtx layer 4.* ***Left:*** *old model suggests that the adult pattern emerges from initially unordered and unrefined projections to layer 4.* ***Right:*** *new model suggests that topographic organization emerges first by projections to the subplate and later in layer 4.* ***Dashed line****, maturing connection.*



Classic experiments in the visual system have shown that topographic order emerges during development because of a refinement process of the thalamic projections to cortical layer 4 (LeVay et al., 1978). The development from the MGB to layer 4 of the ACtx is also thought to undergo such refinement (**Figure 2B**, *left*) (Razak et al., 2009). However, recordings in young animals showed a topographic organization of sound-evoked responses in the subplate at ages before layer 4 neurons responded to sound (Wess et al., 2017). Thus, topographic maps emerge in the subplate and not in layer 4 and also earlier than previously appreciated (Wess et al., 2017).

These observations suggested a new model of the development of cortical topographic maps; maps might be established in the subplate and these maps might later be transferred into layer 4 by the projections from the subplate to layer 4 (**Figure 2B**, *right*). Although the development of MGB projections to the subplate and then to layer 4 is sequential, this sequential nature is not a general rule across the auditory-processing hierarchy. Hence, one can speculate that the initial period when MGB fibers are present in the subplate serves a particular developmental purpose, namely, developing a functional scaffold aiding the development of cortical organization. In this model, building an initial sketch, or scaffold, of the topographic map in parallel with the generation of the cortical layers could enable faster development of sensory cortical function than serial layer generation and map development.

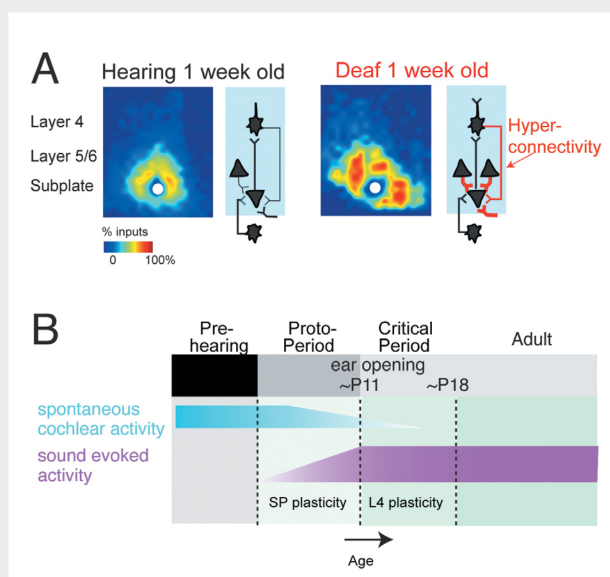
## The Role of Neural Activity and Early Sensory Experience on Cortical Development

Neural activity plays a big role in shaping brain circuits. However, the origin of this activity and its nature change during development. At the earliest stages of development, a large part of the neural activity observed in the brain appears to be overtly "spontaneous," meaning not occurring in response to an external sensory stimulus (Khazipov and Luhmann, 2006). Later in development, neural activity driven by sensory stimuli such as sounds becomes dominant. One important source of early spontaneous activity is the cochlea (Tritsch et al., 2007), which produces spontaneous activity before the onset of sensory transduction and ear opening, although it wanes after ear opening. Work in developing rodents has shown that cochlea-generated spontaneous activity propagates all the way to the ACtx (Babola et al., 2018). The peripheral generated spontaneous activity patterns can be thought of as test patterns that prime the downstream circuits for the onset of the sensory experience and play a key role in sculpting nascent circuits before the onset of sensory functions.

The effects of the sensory experience during the periods where peripheral spontaneous activity is present have only recently been started to be studied. This is because it had been assumed that animals did not hear before ear opening. But because the ACtx can show sound-evoked responses before the ears open (Wess et al., 2017); this indicated the potential for an effect of the sound experience on ACtx circuits.

Subplate neurons also receive inputs from the developing cortical neurons (Viswanathan et al., 2012). The topology of these circuits can be studied in rodent brain slice preparations (**Figure 3A**) and allows for the study of circuit changes after manipulations. Studies in genetically deaf mice, for example, such as those deficient in mechanotransduction (Meng et al., 2021) or synaptic function (Mukherjee et al., 2021), revealed that their subplate neurons receive inputs

**Figure 3.** ***A:*** *subplate neurons receive intracortical inputs, and early deafness leads to hyperconnectivity of subplate neurons. Images show the density of inputs from each cortical location to subplate in mice. In deaf mice, connections to subplate neurons arise from more neurons. Adapted from Meng et al. (2021).* ***B:*** *the 3 phases of early auditory development.* ***Gray****, stages of hearing. SP, subplate; P11 and P18, postnatal days 11 and 18, respectively.*



from more cortical neurons (called hyperconnectivity), even when examined before ear opening. This suggests that the lack of sound inputs even before ear opening had caused circuit changes (**Figure 3A**).

Conversely, raising mice with background sounds before ear opening showed that the presence of sounds even before ear opening reduces connections to the subplate neurons (Meng et al., 2021). These bidirectional changes indicate that even though sound transmission and neural processing is immature at early ages, the auditory environment can already shape auditory cortical circuits. These experiments suggest that a lack of sound input leads to a compensatory increase in connections to subplate neurons and thereby can potentially alter subsequent developmental processes. Importantly, these experiments show that manipulating the sound experience before the onset of the "classic" critical period, which starts at the ear opening period, can alter the development of ACtx circuits (Meng et al., 2021).

The effects of the sound experience at the next stage of development, such as after ear opening, have been well studied, especially on circuits in layer 4 and beyond. This period starts when MGB fibers contact layer 4 neurons; sensory-evoked neural activity during this later stage of development when the eyes and ears function is pivotal for shaping and fine-tuning brain circuits. Hence, it has been called the "critical period" but might be better labeled as the "L4 critical period." Sound exposures during this time, for example, raising rodents in the presence of noise or tones from just before ear opening (in mice around postanal day 11 [P11]), resulted in altered frequency selectivity of ACtx neurons and abolished (Zhang et al., 2002) or altered (Zhang et al., 2001) tonotopic maps in the ACtx. All these sound exposures were effective during a period lasting less than a week following ear opening and therefore show that there is a limited period when L4 circuits seem to be malleable.

These results force us to rethink the early developmental period during which MGB axons are present in the subplate (Kostovic and Rakic, 1990) and when the cochlea is able to transduce sounds. This period is likely highly dynamic in that it involves circuit refinement and emergence of topographic maps (**Figures 2C** and **3B**). Thus, this period represents a "proto-organizational period" in which an outline of cortical organization develops. Accordingly, we can divide the early developmental process into three distinct phases (**Figure 3B**).

(1) *Prehearing Period:* No sensory evoked activity is present. Only spontaneous activity is present.
(2) *Proto-Organizational Period:* Sound-evoked activity is present and can drive plasticity in the subplate. Because of closed ear canals, sound thresholds are high. Peripheral spontaneous activity is also present but decreasing. Layer 4 is not directly activated by MGB.
(3) *Normal-Hearing Period:* Sound-evoked activity is present. The MGB directly activates layer 4 and sound manipulations can cause layer 4 plasticity. The beginning of this period marks the classic critical period. Because of open ear canals, sound thresholds are low.

## Clinical Implications of Early Sensory Effects on Cortical Circuits

Congenital hearing loss is a relatively common condition found in 1 in about 1,000 newborns and is of diverse origin (Chen and Oghalai, 2016). Long-term deafness results in large-scale and fine-scale changes in the ACtx and beyond. For example, adult congenitally deaf cats have a decreased cortical thickness in different auditory cortical regions (Berger et al., 2017), suggesting the atrophy of neurons and/or connections. Similarly, widespread changes in large-scale brain structure are also seen in humans with hearing loss (Manno et al., 2021). However, we now know that deafness already results in brain changes at the younger ages (likely even before birth); thus the adult phenotype might be due to cascading and compounding changes throughout the development of cortical circuits.

The early susceptibility of subplate neurons to sound is important in the case of babies in the NICU where they are exposed to an abnormal sound environment. Care must be taken to adjust the ambient sensory environment as to not overactivate or deprive cortical circuits. Moreover, these considerations are important in other contexts because in many prelingually deaf humans, cochlear implants (CIs) are fitted within the first years to restore hearing. The programming of these devices must consider that auditory cortical processing might already have been altered at time of implantation and is changing during the initial period of use.

## What Are Subplate Neurons Listening to?

Given that sounds can shape the early established circuits, it seems natural to identify which sounds are likely to influence subplate circuits in both humans and altricial animals. In humans, external sounds will be attenuated and filtered by the womb (Gerhardt et al., 1990) and the mother will be the dominant source of sounds. Sources generated by the mother will include breathing, heartbeat, digestive noises, and vocalizations. A distinguishing feature between these sounds is that the first three are ongoing and have relatively constant spectral content, whereas speech is more rare and variable.

Given that developing synapses show high rates of adaptation and that young neurons do not sustain high firing rates, responses to ongoing stimuli likely adapt quickly. In contrast, natural speech has a varying frequency content, is irregular, and is likely to produce less adaptation. Therefore, it is likely that intermittent speech sounds will cause more subplate activity than background sounds. Similarly, external sounds such as other voices or music can be transmitted to the fetus but will be attenuated and filtered. Thus, rare lower frequency sounds will be most effective in activating subplate neurons (Hepper and Shahidullah, 1994).

In animals, the situation is similar, but because the ear canals are closed, maternal vocalizations are attenuated. Moreover, given that pups are outside the womb, other sources of sound are present. Some major potential sound sources are self-generated vocalizations and vocalizations of conspecifics close by in the nest. Thus, it is intriguing to speculate that self-generated sound stimuli could aid in the development of the auditory system in altricial animals. Such a scenario is not too far-fetched because elegant work in ducklings has shown a role for self-vocalizations in auditory development (Gottlieb, 1971).

These considerations also apply to the auditory environment in the NICU because premature infants are suddenly exposed to a very different auditory environment than they had experienced in the womb. High-frequency sounds are not attenuated outside the womb and can potentially drive neural activity. Therefore, care should be taken to replicate the fetal environment by attenuating such sounds. Furthermore, providing rare, speech-like sounds such as recordings of the mother might be of use.

## Damage of Subplate Neurons Might Cause Developmental Abnormalities and Sensory Dysfunctions

Given their key location and early development, it should be of no surprise that damage of the subplate neurons due to exposure to drugs or injury leads to developmental abnormalities, including those associated with sensory-processing deficits. For example, lesioning subplate neurons prevents the topographic and functional maturation in layer 4 (Kanold and Luhmann, 2010) and leads to altered large-scale activity changes in the brain (Tolner et al., 2012), suggesting that the altered brain activity observed in infants could be indicative of subplate damage or lesions. Moreover, neonatal hypoxia-ischemia, which in humans is linked to a variety of neurodevelopmental disorders, leads to subplate hyperconnectivity (Sheikh et al., 2019). Subplate abnormalities are also seen in rodent models of autism spectrum disorder (ASD) (Nagode et al., 2017). Thus, sensory-processing deficits in multiple neurodevelopmental disorders could be the consequence of early subplate damage that prevents the maturation of cortical sensory processing.

## Outlook

If and how the early sound experience can shape our auditory system has long been debated. Studies of early development of the auditory cortex in animals have shown that sound-evoked activity is present much earlier than previously assumed and that an early sensory experience can leave a long-lasting trace. It remains to be tested if such early exposure can influence the further development of the cortex and could thereby promote language or musical learning at infant ages.

The considerations discussed draw almost exclusively from nonhuman animal studies. The subplate is expanded and more compartmentalized in primates than in rodents (Molnar and Clowry, 2012), indicating that subplate size might scale with brain complexity. It is an open question if humans contain specialized subplate neurons or if human brains are enriched in certain subplate subpopulations.

## Acknowledgments

This work was supported by National Institutes of Health Grant R01-DC-009607. I thank Zara Kanold-Tso for help with Figure 1. Also, I thank both past and present

members of the laboratory as well as all collaborators, colleagues, and mentors.

### References


Babola, T. A., Li, S., Gribizis, A., Lee, B. J., Issa, J. B., Wang, H. C., Crair, M. C., and Bergles, D. E. (2018). Homeostatic control of spontaneous activity in the developing auditory system. *Neuron* 99(3), 511-524.e1-e5. https://doi.org/10.1016/j.neuron.2018.07.004.

Barkat, T. R., Polley, D. B., and Hensch, T. K. (2011). A critical period for auditory thalamocortical connectivity. *Nature Neuroscience* 14(9), 1189-1194. https://doi.org/10.1038/nn.2882.

Berger, C., Kühne, D., Scheper, V., and Kral, A. (2017). Congenital deafness affects deep layers in primary and secondary auditory cortex. *Journal of Comparative Neurology* 525(14), 3110-3125. https://doi.org/10.1002/cne.24267.

Budinger, E., and Kanold, P. O. (2018). Auditory cortex circuits. In Oliver, D. L. Cant, N. B., Fay, R. R., and Popper, A. N. (2018. *Mammalian Auditory Pathways: Synaptic Organization and Microcircuits*. Springer International, Cham, Switzerland, pp. 199-233.

Chen, M. M., and Oghalai, J. S. (2016). Diagnosis and management of congenital sensorineural hearing loss. *Current Treatment Options in Pediatrics* 2(3), 256-265. https://doi.org/10.1007/s40746-016-0056-6.

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science* 208(4448), 1174-1176.

DeCasper, A. J., and Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 17(5), 481-491. https://doi.org/10.1002/dev.420170506.

Gerhardt, K. J., Abrams, R. M., and Oliver, C. C. (1990). Sound environment of the fetal sheep. *American Journal of Obstetrics and Gynecology* 162(1), 282-287. https://doi.org/10.1016/0002-9378(90)90866-6.

Goodrich, L. V., and Kanold, P. O. (2020). Functional circuit development in the auditory system. In Rubenstein, J., Chen, B., Rakic, P., and Kwanm, K. Y. (Eds.), *Neural Circuit and Cognitive Development*. Academic Press, San Diego, CA, pp. 27-55.

Gottlieb, G. (1971). *Development of Species Identification in Birds: An Inquiry into the Prenatal Determinants of Perception*, 1st ed. University of Chicago Press, IL.

Hackett, T. A., Barkat, T. R., O'Brien, B. M. J., Hensch, T. K., and Polley, D. B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *Journal of Neuroscience* 31(8), 2983-2995. https://doi.org/10.1523/JNEUROSCI.5333-10.2011.

Hepper, P. G., and Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood* 71 (Suppl. 3), 8-9. https://doi.org/10.1136/fn.71.2.f81.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin* 44(2), 107-112. https://doi.org/10.1016/s0361-9230(97)00094-4.

Kaas, J. H. (2000). Organizing principles of sensory representations. *Novartis Foundation Symposium* 228, 188-205. https://doi.org/10.1002/0470846631.ch13.

Kanold, P. O., and Luhmann, H. J. (2010). The subplate and early cortical circuits. *Annual Review of Neuroscience* 33(1), 23-48. https://doi.org/10.1146/annurev-neuro-060909-153244.

Khazipov, R., and Luhmann, H. J. (2006). Early patterns of electrical activity in the developing cerebral cortex of humans and rodents. *Trends in Neuroscience* 29(7), 414-418. https://doi.org/10.1016/j.tins.2006.05.007.

Kostovic, I., and Molliver, M. (1974). A new interpretation of the laminar development of cerebral cortex: Synaptogenesis in different layers of neopallum in the human fetus. *Anatomical Record* 178(2), 395.

Kostovic, I., and Rakic, P. (1980). Cytology and time of origin of interstitial neurons in the white matter in infant and adult human and monkey telencephalon. *Journal of Neurocytology* 9(2), 219-242. https://doi.org/10.1007/BF01205159.

Kostovic, I., and Rakic, P. (1990). Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *Journal of Comparative Neurology* 297(3), 441-470. https://doi.org/10.1002/cne.902970309.

LeVay, S., Stryker, M. P., and Shatz, C. J. (1978). Ocular dominance columns and their development in layer IV of the cat's visual cortex: A quantitative study. *Journal of Comparative Neurology* 179(1), 223-244. https://doi.org/10.1002/cne.901790113.

Luskin, M. B., and Shatz, C. J. (1985). Neurogenesis of the cat's primary visual cortex. *Journal of Comparative Neurology* 242(4), 611-631. https://doi.org/10.1002/cne.902420409.

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology* 19(23), 1994-1997. https://doi.org/10.1016/j.cub.2009.09.064.

Manno, F. A. M., Rodriguez-Cruces, R., Kumar, R., Ratnanather, J. T., and Lau, C. (2021). Hearing loss impacts gray and white matter across the lifespan: Systematic review, meta-analysis and meta-regression. *NeuroImage* 231, 117826. https://doi.org/10.1016/j.neuroimage.2021.117826.

Meng, X., Mukherjee, D., Kao, J. P. Y., and Kanold, P. O. (2021). Early peripheral activity alters nascent subplate circuits in the auditory cortex. *Science Advances* 7, eabc9155.

Meng, X., Solarana, K., Bowen, Z., Liu, J., Nagode, D. A., Sheikh, A., Winkowski, D. E., Kao, J. P. Y., and Kanold, P. O. (2020). Transient subgranular hyperconnectivity to L2/3 and enhanced pairwise correlations during the critical period in the mouse auditory cortex. *Cerebral Cortex* 30(3), 1914-1930. https://doi.org/10.1093/cercor/bhz213.

Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., and Sereno, J. (2017). Fetal rhythm-based language discrimination: A biomagnetometry study. *NeuroReport* 28(10), 561-564. https://doi.org/10.1097/WNR.0000000000000794.

Molnar, Z., and Clowry, G. (2012). Cerebral cortical development in rodents and primates. *Progress in Brain Research* 195, 45-70. https://doi.org/10.1016/B978-0-444-53860-4.00003-9.

Molnár, Z., Luhmann, H. J., and Kanold, P. O. (2020). Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* 370(6514), eabb2153. https://doi.org/10.1126/science.abb2153.

Mukherjee, D., Meng, X., Kao, J. P. Y., and Kanold, P. O. (2021). Impaired hearing and altered subplate circuits during the first and second postnatal weeks of otoferlin deficient mice. *Cerebral Cortex*, bhab383, https://doi.org/10.1093/cercor/bhab383.

Nagode, D. A., Meng, X., Winkowski, D. E., Smith, E., Khan-Tareen, H., Kareddy, V., Kao, J. P. Y., and Kanold, P. O. (2017). Abnormal development of the earliest cortical circuits in a mouse model of autism spectrum disorder. *Cell Reports* 18(5), 1100-1108. https://doi.org/10.1016/j.celrep.2017.01.006.

Razak, K. A., Zumsteg, T., and Fuzessery, Z. M. (2009). Development of auditory thalamocortical connections in the pallid bat, *Antrozous pallidus*. *Journal of Comparative Neurology* 515(2), 231-242. https://doi.org/10.1002/cne.22050.

Richards, D. S., Frentzen, B., Gerhardt, K. J., McCann, M. E., and Abrams, R. M. (1992). Sound levels in the human uterus. *Obstetrics and Gynecology* 80(2), 186-190.

Schreiner, C. E., and Winer, J. A. (2007). Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 56(2), 356-365. https://doi.org/10.1016/j.neuron.2007.10.013.

Sheikh, A., Meng, X., Liu, J., Mikhailova, A., Kao, J. P. Y., McQuillen, P. S., and Kanold, P. O. (2019). Neonatal hypoxia-ischemia causes functional circuit changes in subplate neurons. *Cerebral Cortex* 29(2), 765-776. https://doi.org/10.1093/cercor/bhx358.

Tolner, E. A., Sheikh, A., Yukin, A. Y., Kaila, K., and Kanold, P. O. (2012). Subplate neurons promote spindle bursts and thalamocortical patterning in the neonatal rat somatosensory cortex. *Journal of Neuroscience* 32(2), 692-702. https://doi.org/10.1523/JNEUROSCI.1538-11.2012.

Tritsch, N. X., Yi, E., Gale, J. E., Glowatzki, E., and Bergles, D. E. (2007). The origin of spontaneous activity in the developing auditory system. *Nature* 450(7166), 50-55. https://doi.org/10.1038/nature06233.

Viswanathan, S., Bandyopadhyay, S., Kao, J. P. Y., and Kanold, P. O. (2012). Changing microcircuits in the subplate of the developing cortex. *Journal of Neuroscience* 32(5), 1589-1601. https://doi.org/10.1523/JNEUROSCI.4748-11.2012.

Wang, X. (2018). Cortical coding of auditory features. *Annual Review of Neuroscience* 41, 527-552. https://doi.org/10.1146/annurev-neuro-072116-031302.

Wess, J. M., Isaiah, A., Watkins, P. V., and Kanold, P. O. (2017). Subplate neurons are the first cortical neurons to respond to sensory stimuli. *Proceedings of the National Academy of Sciences of the United States of America* 114(47), 12602-12607. https://doi.org/10.1073/pnas.1710793114.

Zhang, L. I., Bao, S., and Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience* 4(11), 1123-1130. https://doi.org/10.1038/nn745.

Zhang, L. I., Bao, S., and Merzenich, M. M. (2002). Disruption of primary auditory cortex by synchronous auditory inputs during a critical period. *Proceedings of the National Academy of Sciences of the United States of America* 99(4), 2309-2314. https://doi.org/10.1073/pnass.261707398.

Zhao, C., Kao, J. P. Y., and Kanold, P. O. (2009). Functional excitatory microcircuits in neonatal cortex connect thalamus and layer 4. *Journal of Neuroscience* 29(49), 15479-15488. https://doi.org/10.1523/JNEUROSCI.4471-09.2009.


## About the Author


**Patrick O. Kanold**
pkanold@jhu.edu

*Department of Biomedical Engineering*
*Johns Hopkins University School of Medicine*
*733 N. Broadway Avenue, Miller 379*
*Baltimore, Maryland 21205, USA*


**Patrick O. Kanold** is currently a professor in the Biomedical Engineering Department, Johns Hopkins University, Baltimore, Maryland. He holds a Diplom-Ingenieur (Dipl.-Ing; MSc) from the Technical University Berlin and a PhD from Johns Hopkins University. His laboratory studies the neural circuits in the developing and adult auditory cortex of animals and how sensory experiences shape these circuits using advanced imaging and physiological methods.

# William A. Yost and the Psychoacoustics of Human Sound Source Perception

*Robert A. Lutfi and Christopher A. Brown*

We think of our eyes as the primary channel through which we perceive the world, "seeing is believing," but, in fact, most of our surroundings at any given moment are out of view. For much of the information about the world around us, we depend on our ears. We hear the approaching bus in the din of traffic and avoid stepping into the street; we hear a familiar voice in the clamor of the crowd and recognize an old friend; we hear music playing, glasses clinking, people laughing, a cocktail party is underway next door. Such seemingly simple acts of recognition are so automatic that we rarely give them any thought, but they are examples of an extraordinary ability to perceive the world through sound, unmatched in accuracy and scale by our most sophisticated machine-recognition systems (Szabo et al., 2016).

William A. (Bill) Yost (shown with his family in **Figure 1**) is a hearing scientist who, for over half a century, has been a leader in the effort to understand this extraordinary ability. As of writing this, Bill has published over 100 peer-reviewed research articles, 6 authored or coauthored books, and 50 book chapters on or related to the topic. The number of major scientific organizations giving their highest form of recognition to Bill's work is too long to list here. You may know Bill from his many years of service to the Acoustical Society of America (ASA). He has held every elected office in the Society including president and was the recipient of the Society's Gold Medal in 2018 (see https://asa.scitation.org/doi/pdf/10.1121/1.5036155). Among other important roles, he has served as president of the Association for Research in Otolaryngology (ARO); program director of the National Science Foundation: Sensory Physiology and Perception; chair of the National Research Council and National Academy of Sciences Committee on Hearing, Bioacoustics and Biomechanics (CHABA); and cochair of the Task Force on Developing the National Strategic Plan for the Establishment of the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health. For more about Bill, see https://asa.scitation.org/doi/pdf/10.1121/1.5036155.

This article provides an overview of Bill's research and, more broadly, the topic of human sound source perception. Readers wishing to delve more deeply into the subject can find review chapters of works by other prominent authors in one of Bill's books, *Auditory Perception of Sound Sources* (Yost et al., 2008).

## Impossible Sound Source Perception

Before talking about Bill's research, it is first necessary to get a sense of why human sound source perception is so remarkable. Bill tells his students the reason is "because it's impossible." As provocative as this answer might seem, it is not far from the truth. We have many examples to choose from; sound source perception can involve something as simple as recognizing the "ping" of a tuning fork or as complex as parsing an entire "auditory scene" (imagine any busy street in downtown Chicago). Let's

**Figure 1.** *Bill Yost on an Alaskan cruise with his family celebrating his 50th anniversary.* ***Left to right:*** *daughter Alyson, Bill, wife Lee, and daughter Kelly.*


https://doi.org/10.1121/AT.2022.18.1.41

start with the tuning fork. Tuning forks we recognize as U-shaped metal bars with a stem. Knowing certain properties of the fork and the way it is held and struck, the prominent modes of vibration, their relative amplitudes, and rate of decay can be estimated from known equations of motion (Russell, 2020). Theoretically, any or all of this information might be used by a listener to correctly perceive the sound as belonging to a tuning fork (Lutfi, 2008). The problem is that, in the real world, the properties of unseen sources are not known beforehand. Instead, they are what we are trying to determine from sound. In the equations of motion, different combinations of properties can produce identical solutions, so if there are no constraints, that ping of the tuning fork could just as easily have come from a hollow flagpole, pogo stick, or ceramic plate. The problem is indeterminate; it has not one but many solutions.

Now consider that busy street in downtown Chicago. You hear traffic, people walking around you, and a siren wailing in the distance. What reaches your ears is the superposition (sum) of the sound pressure wave fronts emitted by all of these sources; you have access only to this sum, but somehow you extract from it and recognize individually the sounds emitted by each source. The problem is principally the same as having to solve for $x$, $y$, and $z$ in the expression $x + y + z = 20$. Again, there is no single solution.

In both examples, the only way perception can be correct is to bring additional information to bear on the problem. Understanding what that information is and how it is encoded in the auditory nervous system has been the fundamental challenge for research on sound source perception and the focus of Bill's work.

## Early Influences and Signal Detection Theory

Bill received his undergraduate degree from Colorado College, Colorado Springs, Colorado, in 1966 with a major in psychology and a minor in mathematics. He knew then that he wanted to be a professor and researcher studying objective, quantifiable ways of explaining how the brain works. That same year, Green and Swets (1966) published their seminal book on signal detection theory (SDT; see Yost et al., 2021). For Bill, the timing was perfect. SDT recognized that perception is covert, that the judgments of subjects in perceptual studies are merely their personal impressions, opinions, or beliefs regarding what they see or hear. SDT would provide a way to convert these subjective impressions into entirely objective measures of perception; in the words of Green (2020), "as objective as any of the quantities used in the so-called hard sciences." This development would bring a sea change in the conduct of perceptual research that would have a lasting impact on Bill's work and on the work of many other scientists of the time.

**Figure 2.** *Bill's academic family tree.* ***Left to right:*** *Lloyd Jeffress (academic grandfather), Don Robinson (academic father), and Bill.*

After graduating from Colorado College, Bill furthered his studies in the Psychology Department at Indiana University (IU), Bloomington, under the tutelage of James Egan, another giant of SDT. He then finished his PhD with Don Robinson (**Figure 2**) after the early departure of Egan from IU. After receiving his doctorate, Bill received a National Science Foundation postdoctoral award to work with Green at the University of California, San Diego. The influence of this early training is evident in Bill's consistent approach to research: model oriented, precise, and given to clear outcomes based on quantitative data. Although there would still be a place for qualitative data in Bill's research, he would be among the first in the field to apply the lessons of SDT to the study of human sound source perception.

## "The Basis for Hearing"

Bill published a call to action, encouraging researchers to focus more attention on sound source perception (Yost, 1991). The title would leave little doubt as to the importance he attached to the subject, "Auditory Image Perception and Analysis: The Basis for Hearing." The article underscored the role of sound source perception in communication and survival and offered compelling examples of how we rely on it every day to navigate our environment. Bill would make the case to a broader audience in three subsequent publications (Yost, 1992, 1993, 2008). In these publications, Bill identifies three major areas of research making up the bulk of the work on human sound source perception, still active today. What follows is an abbreviated review of the highlights of the work in each area, concentrating on the key contributions made by Bill.

### *Pitch*

Of the three primary qualities we perceive in sound, (pitch, loudness, and timbre), pitch is most closely tied to the properties of the sound source. Loudness varies with distance from the source, the driving force for vibration, and any obstacles that might block the sound's path on the way to our ears. Timbre is affected by room acoustics, how the source is supported, and how it is driven to vibrate. Pitch, however, is much less subject to these extraneous factors and depends more on the properties of the resonating source itself.

The long history of research on pitch shows that it corresponds largely to our perception of periodicity in sound. Many sounds in nature, particularly those having the most significance for us, are periodic, or at least roughly so. Speech and music are the most notable examples. These sounds have a harmonic structure whose periodicity is given by a fundamental frequency (F0) that with few exceptions dominates our perception of pitch. So strong is this tendency that we hear a pitch at F0 when there is little or no energy at F0; and even when the sound is inharmonic, we tend to hear a pitch corresponding to the closest match to F0 (see Yost, 2009, for a review and http://auditoryneuroscience.com/pitch for online demos).

Pitch contributes to sound source perception in a variety of ways. It tells of an animal's size through their vocalizations, generally lower pitch vocalizations corresponding to larger size. Larger sized animals are more attractive to potential mates and are a greater threat to competitors. In humans, pitch affects the meaning of a spoken sentence through prosody and conveys information about the talker's gender and even their emotional state. It also plays an important role in helping to segregate sound sources perceptually. The individual spectral components of multiple sources sounding simultaneously are conflated in a complex spectrum reaching the ears. But a separate pitch is heard for each source, effectively segregating the sounds by their harmonic structure. A popular demonstration of this is when a single component of an otherwise harmonic complex is slightly mistuned. The pitch of the mistuned component will "stand out" from that of the harmonic complex such that two pitches are heard simultaneously (Moore et al., 1986). The literature includes a variety of examples of segregation based on pitch (reviewed by Carlyon and Gockel, 2008).

Bill's work on pitch has focused on how it is encoded in the auditory system, the second part of the fundamental goal of research on sound source perception. The question has prompted an ongoing dispute, dating back to Helmholtz (1863, 1954), between two theories: one centering on the features of the time waveform and the other on its spectrum. Because the spectrum is a translation of the time waveform, early tests of the two theories based on acoustics alone proved difficult. Modern theory has since contributed what we have learned about the transformations of the signal taking place in the auditory periphery. We now know that individual fibers in the auditory nerve are selectively responsive to different frequencies, providing a *place code* for the sound spectrum. Temporal features of sound are also represented in the group *synchronous* response of nerve fibers to signals. The combination of these processes results is a neural activation pattern (NAP) in frequency and time that preserves much of the spectral and temporal information present in the airborne sound.

**Figure 3**, *left*, was derived from the NAP model of Patterson et al. (1995). It shows the simulated neural response to a 200-Hz harmonic complex, which produces a strong perception of pitch at 200 Hz. Looking horizontally, one can readily see the oscillations, shifted in phase vertically, and having a periodicity of 5 ms, the reciprocal of 200 Hz. Looking vertically, one can also make out the representation of the harmonic spectrum as neighboring activation maxima, with a spacing of 200 Hz. **Figure 3**, *right*, shows the simulated response to a special signal that Bill popularized and termed iterated rippled noise (IRN). IRN is created by passing random noise through a delay and add-back circuit and applying multiple iterations of the circuit (see Yost, 2009). There were three iterations of a delay of 5 ms for the signal in **Figure 3**.

IRN poses a challenge for models of pitch because it has no clear spectral or temporal structure but nonetheless

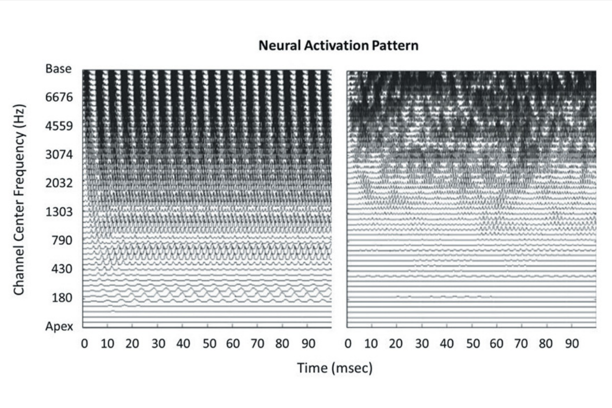



**Figure 3.** ***Left:*** *Neural activation pattern (NAP) of a 200-Hz harmonic complex from model of Patterson et al. (1995).* ***Right:*** *NAP of an iterated rippled noise (IRN), three iterations with 5-ms delay. Adapted from Yost (2009). See text for detailed discussion.*

produces a pitch corresponding to the inverse of the delay. Bill has investigated extensively how the pitch and pitch strength of IRN changes with its various parameters and has concluded that a temporal model that extracts periodicities in the fine structure of IRN best accounts for the data. Bill's work on IRN is his most frequently cited and has contributed to making IRN a standard stimulus in many other areas of research on hearing.

## *Temporal Modulations in Sound*

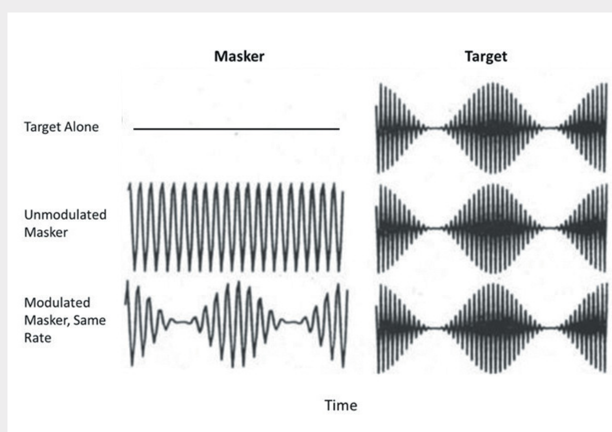
Most sounds of interest to us in nature change over time. The messages we convey through speech, the pleasure we take in music, and the actions we track in the sound events unfolding around us all derive from modulations in the sound spectra over time. Without these dynamics, the sounds about us would combine to produce a single unpleasant drone.

Two parallel and largely independent lines of research have evaluated the influence of temporal modulations on sound source perception. The first has focused on the phenomenon of *auditory streaming* (Bregman, 1990). This refers to the listener's subjective impression of when a sequence of sounds, typically tones, is heard to split into separate perceptual objects or entities (streams). Striking examples occur when the tones in the sequence are made to differ in frequency and rhythmic pattern (for demos, see http://auditoryneuroscience.com/index.php/scene-analysis).

The second line of research has focused on *auditory masking*, an objective measure of the influence one sound (the masker) has on the listener's ability to detect, discriminate, or recognize another (the target). The early view of auditory masking, dating back to Fletcher (1940), was that it is caused by the overlap of neural excitation produced by the target and masker in the auditory periphery. Bill would publish one of the early studies, indicating that the process is much more complex and possibly connected to auditory streaming (Yost et al., 1989).

**Figure 4** shows three conditions of that study. The listener's task was to detect an increase in the base modulation rate of a target tone (**Figure 4,** *right*). The target was either presented alone (**Figure 4,** *top*), presented with an unmodulated masking tone (**Figure 4,** *center*), or presented with the masking tone modulated at the same base frequency as the target (**Figure 4,** *bottom*). Little masking was expected in the two masking conditions because there was always a two-octave separation between target and masker; indeed, the unmodulated masker had little effect on threshold, consistent with that expectation. The modulated masker, on the other hand, produced a significant, unexpected increase in threshold, suggesting a perceptual interference created by the common modulation. The results are reminiscent of those from the streaming experiments where common temporal modulations in the frequencies of tones cause those tones to fuse into a single auditory image (Bregman, 1990). Bill's results on the effects of temporal modulations on masking and those of many other studies

**Figure 4.** *Three conditions adapted from the study by Yost et al. (1989). See text for discussion.*



conducted at this time would lead to a dramatic change in thinking about the factors that affect auditory masking.

## *Spatial Attributes of Sound*

As discussed, Bill views sound source perception as the primary function of our sense of hearing. He has argued that identifying the sources of sound in our environment is paramount to survival (Yost, 2008). From an evolutionary perspective, the job of the perceptual system is to make sense of the world so that the organism can interact effectively with it. More specifically, identification is required for organisms to discriminate predators, prey, and potential mates so that they can act accordingly to survive. But using pitch, temporal, and other cues to deduce that a sound source is a potential predator, for example, would not be especially helpful if we were not able to also identify its location and then avoid it.

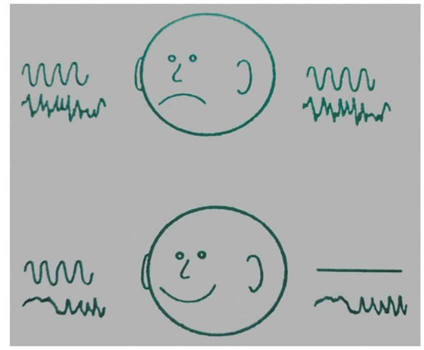


**Figure 5.** *Conditions in a typical masking level difference (MLD) experiment.* ***Top:*** *both the tone and noise are diotic or homophasic.* ***Bottom:*** *the noise is homophasic while the tone is presented monaurally to the left ear only, or antiphasic. Adapted from Green and Yost (1975).*

Sound source localization arises from our ability to process relatively small differences in the auditory signals between the two ears. A sound coming from the left of a listener will arrive at the left ear sooner in time than it will at the right ear. The sound will also generally be louder at the left ear than the right due to the head shadow. These interaural differences of time (ITDs) and level (ILDs) are the cues used to localize sound sources in the horizontal plane. Bill has contributed a wealth of information to our understanding of these spatial cues in numerous papers spanning over 40 years (e.g., Green and Yost, 1975; Yost and Pastore, 2019). For example, thanks to Bill's efforts, we better understand ILD and ITD sensitivity across frequency (e.g., Yost and Dye, 1988), by cochlear implant users (e.g., Doorman et al., 2014), and in the presence of time-varying amplitude fluctuations (e.g., Yost et al., 1989).

In addition to facilitating sound source localization, spatial cues can also provide additional benefits for detection, discrimination, and identification tasks that occur in the presence of one or more additional concurrent, spatially separated sound sources or maskers. When the task is speech perception, it is often described as solving the "cocktail party problem," a term coined by Cherry (1953). There are other terms for the general perceptual benefits that arise from spatial separation of sound sources, including spatial release from masking (SRM) and the masking level difference (MLD).

Bill has made significant contributions to the literature characterizing the MLD, which was first described nearly simultaneously by both Licklider (1948) and Hirsh (1948). If the same tone is presented to both ears using headphones (described as "homophasic" because the tone has the same phase to both ears) and then one adds a homophasic noise, the signal-to-noise ratio (SNR) can be manipulated so that the listener can just detect the tone. If the phase of the tone is changed in one ear so that it is different than that in the other ear (an "antiphasic" condition), the perceived location of the tone in the listener's head will change because of the interaural phase delay. Interestingly, the level of the noise will have to be increased to produce the same amount of masking. In this example, the difference in SNR between the homophasic and antiphasic conditions is the MLD.

**Figure 5** depicts an even simpler and more striking example. In **Figure 5**, *top*, the tone and noise are both homophasic and the sad face indicates that the listener is having difficulty detecting the tone. In **Figure 5**, *bottom*, the tone has been turned off in one ear and the noise remains homophasic, a condition in which the amount of masking (the MLD) is reduced, as indicated by the happy face. To summarize, simply eliminating the tone in one ear made the tone more easily detected!

The MLD is a particularly elegant example of taking a complex phenomenon (the perceptual benefits of spatially separated targets and maskers) and reducing the problem to its essence so that it can be studied systematically. Since

Hirsh's and Licklider's initial papers in 1948, Bill has explored the various conditions under which the MLD does and does not occur (Yost, 1988). The effect has been shown by various researchers for tones, speech, and other signals, using both interaural phase (or time) differences and ILDs, and even in temporal masking paradigms, in which the signal and noise are not presented concurrently. In addition to publishing many influential articles on various aspects of the MLD, Bill along with his friend and colleague Tino Trahiotis in 1998 organized *The MLD: A Collection of Seminal Papers* to commemorate the 50th anniversary of the Licklider and Hirsh papers and to highlight and celebrate the vibrant psychoacoustics community, many of whom contributed to our understanding of this interesting phenomenon. The image in **Figure 5** was taken from the cover of this collection.

Bill's more recent work has focused on the maximum number of spatially separated sound sources in an auditory scene that listeners are able to successfully process (Yost et al., 2018, 2019b). These studies have found that for talkers simulating a cocktail party or noisy restaurant auditory scene, the maximum size of the auditory scene appears to be four. More specifically, listeners were relatively accurate in both identifying and discriminating the total number of talkers and reporting talker locations when there were up to four talkers. Listeners could also judge loudness differences based on individual source levels when there were four or fewer sources. With five or more sources, discrimination of the total number of talkers and localization accuracy approached chance, and listeners tended to use overall level to perform the loudness difference task rather than individual source levels, indicating an inability to "hear out" individual sources or streams.

Most recently, Bill and his team have been interested in auditory motion and the effect of head turns on sound source localization, with a focus on cochlear implant (CI) users who are well-known for being poor localizers (Brown, 2018). It was established some time ago that head movements are integrally related to localization (Wallach, 1940). The work by Pastore et al. (2020) in this area established that head turns significantly improved localization abilities for single-sided deafened individuals implanted with a CI with their CI both off (monaural condition), and on (so-called bimodal listening condition).

In fact, auditory motion is a sorely understudied topic. One very good reason for this is the many technical challenges and other difficulties that interfere with the ability to exert sufficient scientific rigor so that the results are generalizable while also maintaining ecologically valid conditions. Ever fearless, Bill undertook the challenge, and the result is a listening room at Arizona State University, Tempe, that has been custom designed and purpose-built for auditory motion experiments (**Figure 6**). The room is sound deadened and contains a custom chair that allows precise measurement and control of rotational velocity and an array of loudspeakers with custom software that allows sound source motion to be accurately simulated.

Using this facility, Bill has collected a trove of interesting data, most of which have been used in published studies on the relationship between localization, source movement, and listener movement (e.g., Yost and Pastore, 2019). One goal of this work was to establish how individuals can use spatial cues during motion. Interaural difference cues are inherently head-centric and thus change with head turns as well as with any other movement of the source or the listener. How then does a listener disentangle a relatively complex scene wherein both the listener and the sound source are moving? Supported by compelling data, Bill has argued in several papers that sound source localization is not a purely a psychoacoustic phenomenon but rather is based on an integration of input from several systems, including auditory, visual, and very likely vestibular (e.g., Yost et al., 2019a, 2020).

**Figure 6.** *Bill's sound insulated room with rotating chair and surrounding speakers for studying auditory motion perception.*

## No End to an Era

In articles in *Acoustics Today* honoring prominent members of our Society (see https://bit.ly/3HC1udm), their retirement has sometimes been talked of as marking the end of an era. This does not apply to Bill. After 50 years of steady scholarly contributions and continuous service to the advancement of our science, he shows no sign of slowing down. In those 50 years, we have seen tremendous progress in our understanding of human sound source perception, in large part thanks to Bill. For those important questions that remain, all indications are that he will continue to be on the forefront of the research providing answers. In a recent special session of the Acoustical Society of America honoring Bill, the title of the first speaker's talk asked, "Does He Ever Sleep?" You might say he does but hasn't made a habit of it. Our science is better for the tireless efforts of Bill and for that we are most thankful.

### References


Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.

Brown, C. A. (2018). Corrective binaural processing for bilateral cochlear implant patients. *PLOS ONE* 13, e0187965.

Carlyon, R. P., and Gockel, H. E. (2008). Effects of harmonicity and regularity on the perception of sound sources. In Yost, W. A., Popper, A. N., and Fay, R. R. (Eds.), *Auditory Perception of Sound Sources*. Springer, Boston, MA, pp. 191-214.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America* 25, 975-979.

Dorman, M., Loiselle, L., Yost, W. A., Stohl, J, Spahr, A., Brown, C., and Cook, S. (2014). Interaural level differences and sound source localization for bilateral cochlear implant patients. *Ear and Hearing* 35(6), 633-640.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics* 12(1), 47-65

Green, D. M. (2020). A homily on signal detection theory. *The Journal of the Acoustical Society of America* 148, 222-225.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York, NY.

Green, D. M., and Yost, W. A. (1975). Binaural analysis. In Keidel, W., and Neff, W. D. (Eds.), *Auditory System. Handbook of Sensory Physiology,* vol. 5/2. Springer, Berlin, Heidelberg, Germany, pp. 461-480.

Helmholtz, H. (1863, 1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Dover, New York, NY.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *The Journal of the Acoustical Society of America* 20, 536-544.

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *The Journal of the Acoustical Society of America* 20, 150-159.

Lutfi, R. A. (2008). Sound source identification. In Yost, W. A., Popper, A. N., and Fay, R. R. (Eds.), *Auditory Perception of Sound Sources*. Springer, Boston, MA, pp. 13-42.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America* 80, 479-483.

Pastore, M. T. Natale, S. J., Clayton, C., Dorman, M., Yost, W. A., and Zhou, Y. (2020). Effects of head movements on sound-source localization in single-sided deaf patients with their cochlear implant on versus off. *Ear and Hearing* 41, 1660-1674.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (1995). Time domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America* 98, 1890-1894.

Russell, D. A. (2020). The tuning fork: An amazing acoustics apparatus. *Acoustics Today* 16(2), 48-55.

Szabo, B., Denham, S., and Winkler, I. (2016). Computational models of auditory scene analysis: A review. *Frontiers in Neuroscience* 15, Article 524. https://doi.org/10.3389/fnins.2016.00524.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology* 27, 339-368.

Yost, W. A. (1988). The masking-level difference and overall masker level: Restating the internal noise hypothesis. *The Journal of the Acoustical Society of America* 83, 1517-1521.

Yost, W. A. (1991). Auditory image perception and analysis: The basis of hearing. *Hearing Research* 56, 8-19.

Yost, W. A. (1992). Auditory perception and sound source determination. *Current Directions in Psychological Sciences* 1, 179-184.

Yost, W. A. (1993). Overview of psychoacoustics. In Yost, W. A., Fay, R. R., and Popper, A. N. (Eds.), *Human Psychophysics*. Springer, New York, NY, pp. 1-12.

Yost, W. A. (2008). Perceiving sound sources. In Yost, W. A., Popper, A. N., and Fay, R. R. (Eds.), *Auditory Perception of Sound Sources*. Springer, Boston, MA, pp. 1-12.

Yost, W. A. (2009). Pitch perception. *Attention, Perception and Psychophysics* 71, 1701-1715.

Yost, W., and Dye, R. H. (1988). Discrimination of interaural differences of level as a function of frequency. *The Journal of the Acoustical Society of America* 83, 1846-1851.

Yost, W., and Pastore, M. T. (2019). Sound-source localization when listeners and sound sources rotate: The auditory Filehne illusion. *The Journal of the Acoustical Society of America* 146, 3047.

Yost, W. A., Pastore, M. T., and Dorman, M. F. (2020). Sound source localization is a multisystem process. *Acoustical Science and Technology* 41, 113-120.

Yost, W. A., Pastore M. T., and Pulling, K. R. (2018). Loudness of an auditory scene composed of multiple talkers. *The Journal of the Acoustical Society of America* 144(3), EL236-EL241.

Yost, W. A., Pastore, M. T., and Pulling, K. R. (2019a). Sound-source localization as a multisystem process: The Wallach azimuth illusion. *The Journal of the Acoustical Society of America* 146, 382-398.

Yost, W. A., Pastore, M. T., and Pulling, K. R. (2019b). The relative size of auditory scenes of multiple talkers. *The Journal of the Acoustical Society of America* 146, EL219-EL224.

Yost, W. A., Patterson, R. D., and Feth, L. L. (2021). David M. Green and psychoacoustics. *Acoustics Today* 17(3), 51-59.

Yost, W. A., Popper, A. N., and Fay, R. R. (Eds.) (2008). *Auditory Perception of Sound Sources*. Springer, Boston, MA.

Yost, W. A., Sheft, S., and Opie, J. (1989). Modulation detection interference: Effect of modulation frequency. *The Journal of the Acoustical Society of America* 86, 2138-2148.

Yost, W. A., and Trahiotis, C. (1999). *The MLD: A Collection of Seminal Papers,* Parmly Hearing Institute, Chicago, IL.

## About the Authors

**Robert A. Lutfi** rlutfi@usf.edu

*Auditory Behavioral Research Lab (ABRL)*
*Department of Communication Sciences and Disorders*
*University of South Florida*
*4202 E. Fowler Avenue, PCD1017*
*Tampa, Florida 33612, USA*

**Robert A. (Bob) Lutfi** is a professor of communication sciences and disorders, University of South Florida, Tampa, and formally an emeritus professor at the University of Wisconsin, Madison. His research focuses on how human detection and recognition of sound sources in noise is affected by lawful and random variation in sound as occurs in nature. The research is funded by the National Institute of Deafness and Other Communication Disorders (NIDCD), National Institutes of Health, Bethesda, Maryland. Bob is a Fellow of the Acoustical Society of America, a member of the US Intelligence Science and Technology Experts Group (ISTEG), and a friend and student of Bill Yost.

**Christopher A. Brown**
cbrown1@pitt.edu

*Department of Communication Sciences and Disorders*
*University of Pittsburgh*
*6074 Forbes Tower*
*Pittsburgh, Pennsylvania 15260, USA*

**Christopher A. (Chris) Brown** is an associate professor of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania. Chris' research interests include speech understanding in noise by normal-hearing and impaired listeners, spatial hearing, and cochlear implant signal processing. His work is funded by the National Institute of Deafness and Other Communication Disorders (NIDCD), National Institutes of Health, Bethesda, Maryland and has resulted in two federal patents. Chris considers Bill Yost a friend, mentor, and colleague.

# Additive Manufacturing Enables New Ideas in Acoustics

Christina J. Naify, Kathryn H. Matlack, and Michael R. Haberman

## Introduction

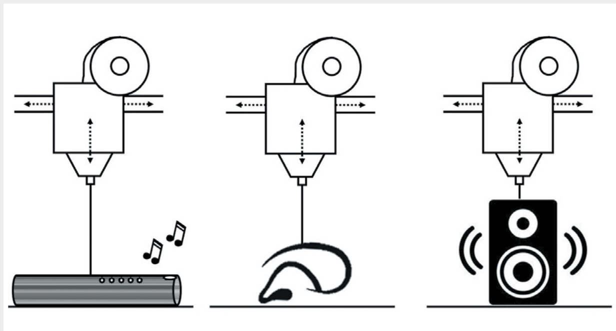
Additive manufacturing (AM), also known as three-dimensional (3D) printing, refers to the process of building a solid object in a layer-by-layer manner to create a desired object. This fabrication approach opens a multitude of new possibilities across virtually all areas of acoustics, ranging from advances in musical instruments, new acoustic materials, and acoustic metamaterials to new opportunities for ultrasonic nondestructive evaluation to new transducer geometries and concepts and customized hearing aids.

AM fabrication is a generic term and can be achieved using a range of techniques. Most AM fabrication techniques create, or "print," objects in a process that begins by first placing a small amount of material in a thin layer on a platform. More material is then placed on the first layer at locations that are defined by the geometry or shape of the object being built and then joined, or "added," to the existing layer. The process is repeated in a layer-by-layer fashion until the entire object is created. In general, the key feature of AM is that material is added where it is wanted by adhering material to previously deposited layers. Overviews of some of the most common AM approaches are included in this article, but the approaches outlined here are only a small slice of a very rich technological field.

The "additive" in AM is in contrast to more traditional subtractive manufacturing in which one begins with a solid block and material is removed where desired to create the final object. Although the approach of AM has been around since the 1980s in academic and industrial laboratories (Ngo, 2018), the past decade has seen an explosion in the availability of 3D printers (the device used to fabricate objects using AM technology) as well as in the price reduction for purchasing printers and supplies. These two factors mean that this exciting technology is available to virtually all individuals or institutions that would like to explore the possibilities afforded by AM. Indeed, it's not uncommon to start a conversation with someone about AM, only to find out that they have a 3D printer in their home office or garage.

The appeal of AM to a wide audience is easy to see once you know about the possibilities it provides. Users are often attracted by the versatility of AM. Because parts are built layer by layer, the final objects can take almost any shape that the builder can imagine. This means that a user can often use a single 3D printer to iterate on designs or print a wide range of objects without having to create custom molds or tooling often required in conventional fabrication methods. **Figure 1** shows a cartoon example of a single 3D printer being used to print a flute (Ritz, 2015), an ear pinna (Prepeliță, 2020), and an acoustic speaker housing

**Figure 1.** *Additive manufacturing (AM) offers manufacturing versatility. Traditional methods would require different approaches to build items such as musical instruments (**left**); hearing prosthetics (**center**); or speakers (**right**), but AM technology has the potential to build each of these on the same desktop-style 3-dimensional (3D) printer without special molds or custom tooling. In fused deposition modeling (FDM), this is achieved using a material print head, or extruder, that is moved in horizontal and vertical directions (**arrows**) while molten material is deposited in the desired location.*




https://doi.org/10.1121/AT.2022.18.1.49

(Ishiguro and Poupyrev, 2014). These are all examples that can be found in industrial, academic, or hobby applications.

Google searches for AM turn up a variety of applications as wide as one's imagination, with top "hits" including things like figurines of cartoon characters, cell phone cases, and more. It's worth pointing out that although most people vaguely familiar with AM imagine desktop-sized printers that extrude plastics, people have developed printers that can make food (Sun, 2015; see https://bit.ly/3r0Tjli); electronics (Goh, 2021); biological tissue (Mannoor, 2013; see https://bit.ly/3nIKeMd); and even buildings (Paolini, 2019).

AM has seen significant activity in both academic instruction and research. The recent past has seen the emergence of AM graduate certificate or degree programs, and academic journals are published that focus solely on additive processes, applications, and development. A wide range of conferences, from general to content specific, also exist. Within the Acoustical Society of America, two special sessions focused on AM have taken place since 2017, and recently a special issue of *The Journal of the Acoustical Society of America* (see https://bit.ly/3gzUNws) was published highlighting research at the intersection of AM and acoustics.

Given the wide range of potential applications for AM in acoustics, it's no surprise that acousticians of all types would find uses for a new fabrication approach. This article introduces AM and its applicability to a wide range of applications in acoustics. We begin by introducing the basic technology, then review some case studies of current research and educational and industrial uses of AM related to acoustic applications, and conclude with some perspectives on the future of AM in acoustics.

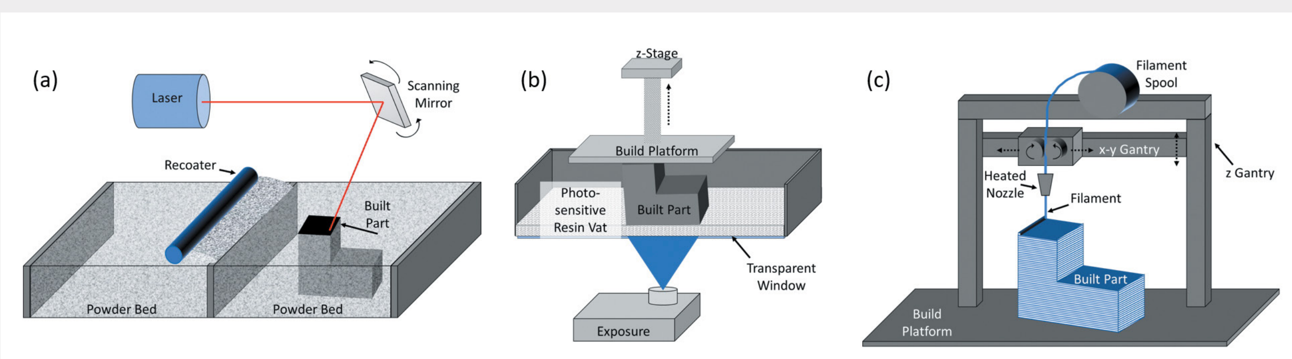
## Overview of Different Additive Manufacturing Techniques

If you wanted to build something like a scale model of a building, you could start with a block of metal, plastic, or wood and cut out the shape of the building using a saw or other standard machining equipment. This approach is known as subtractive manufacturing because you remove, or subtract, the material you don't want from a large piece of that material.

An alternative approach to building the same model would be to use something like LEGO bricks to create the same shape by adding material point-by-point only where it is needed. Although a rudimentary example, building with LEGO bricks is an example of AM where small pieces are added to make the final object. That object can take almost any shape you can imagine, as is the case with building with LEGO bricks.

The concept of construction by adding material where you want it can be accomplished in a variety of ways. Three of the most common approaches are described here, but we note that this list is far from exhaustive. The key difference between the three methods described here is the base material, in that the material that is added sequentially to additively build up the object. In the example above, the base material is a plastic LEGO brick. The method of joining in that case is snapping the bricks together. Commercial additive approaches typically use one of three base material forms, powder, liquid, or thin strips, all of which are joined by adhering a new layer of material to

**Figure 2.** *Three forms of AM.* ***a:*** *Powder-based printing.* ***b:*** *Stereolithography.* ***c:*** *Fused deposition modeling. See text for more details.*



the previously added material, like laying down a new layer of LEGO bricks on top of the existing layer.

As with any building project, the first step to build the object is to create a design and generate building plans. In the case of AM, the design is done on a computer by drawing a 3D representation of the object. A wide variety of commercial tools are available to do this, and the goal is to create a computer file that the printer can understand. Because AM is mostly done in a layer-by-layer approach, the 3D object one wishes to create must be broken down into 2D cross sections, known as slices. For example, to build a solid cube, the plans would consist of identical slices of squares. To build a sphere, each slice would be a different sized circle. The job of the printer is to fabricate the object in successive 2D slices until a 3D object is constructed. Three common AM approaches are shown in **Figure 2**.

In the first example, the base material is in powder form, with grains having dimensions on the order of tens of microns. The process begins by spreading the powder to create a thin layer on the print bed, which is the "ground" for the object. One such printing technique, known as laser powder bed fusion (LPBF), then fuses grains together using a laser that heats the powder beyond its melting point. Only the powder that should be part of the final object is melted, and the rest of the powder remains in powder form. If the first layer is a circle, the laser will fuse powder together within the circle, leaving the rest of the powder unaltered. Another layer is then created by spreading powder from a stock basin using a rolling cylinder called the recoater, and the powder is again fused together at locations required to create the object. This process repeats, sometimes thousands of times, until the object is complete. This method is extremely versatile in that the base material can be a metal like steel or a polymer like nylon. **Figure 2a** shows an example of LPBF including two powder beds, one for the built part and one for the stock powder.

In the second example the base material is in liquid or resin form, and it is cured/solidified in a layer-by-layer manner using a light source such as a laser or projector, a process called photopolymerization. The most common AM techniques based on photopolymerization is stereolithography (SLA), which is used to print plastic objects ranging in size from a few nanometers to centimeters. **Figure 2b** illustrates a common approach where the "new" build material is added to the bottom of the part as it is moved in the positive $z$-direction by a gantry. In this configuration, light is projected through a transparent window onto a thin layer of the uncured resin directly below the part.

The third method, material extrusion, is likely the most familiar to the reader. In this approach, the base material is a thin strip of plastic called a filament that is deposited along a predefined path that makes up the layer of the object being created. The nozzle deposits the filament by melting the base material, like how a glue gun converts a glue stick to a semiliquid substance that can be deposited at will. The filament is used to trace out each 2D shape on top of the previous one to create the 3D structure.

Material extrusion approaches can print a variety of materials and have the added advantage of being able to print parts made from two or more materials or even functional materials such as piezoelectrics (Chen, 2020). A common filament-based AM approach is fused deposition modeling (FDM), which is shown in **Figure 2c**. FDM employs the heated nozzle mounted on a gantry to extrude semisolid filaments like plastics, metals, or composite materials. This type of printer is by far the most accessible in terms of cost, with base models starting at a few hundred dollars. For this reason, and the fact that they are easy to use by beginners and have small footprints of less than half a meter in each direction, these are the most common type of printers available.

It is worth noting that each of these methods has advantages and disadvantages, the details of which are beyond the scope of this article. As a short summary, powder-based methods are very expensive but produce objects with good mechanical strength and desirable attributes such as smooth surfaces. Liquid-based photopolymerization methods can be messy and cannot be used to build objects larger than a few centimeters. However, small objects, such as those used in many high-precision engineering applications, are readily fabricated using this approach. Filament-based methods are low cost and easy to use, but the finished products have rough surfaces and have poor mechanical strength.

## Musical Acoustics

Access to musical instruments takes on a new twist when AM is incorporated. In some cases, by using existing plans online or drawing one's own plans, it's possible to use AM to build a kazoo, guitar, or flute. Recent review articles such as that by Michon et al. (2018) have highlighted studies using AM and acoustics. An interesting use case is the fabrication of instruments where the user has access to the drawings or 3D model of an instrument. An example of this were researchers who had drawings of clarinet mouthpieces from the 1890s and fabricated them using AM (Cottrell and Howell, 2019). The authors printed multiple mouthpieces using different techniques such as FDM or stereolithography and received qualitative feedback from professional musicians regarding quality. AM effects such as surface finish or material strength were found to vary from process to process by the evaluating musician's response and these features affected the perceived quality of each mouthpiece.

String instruments produced using AM are rare due to material availability, limiting materials selection to plastics, metals, or ceramics, all of which produce a dynamic response that differs from traditional wood designs (Qian, 2019). Studies of AM ukuleles using FDM showed large differences in both measured A-weighted sound level and timbre compared with a traditional wood instrument. This is because the AM-constructed material has a different stiffness from the wood instrument. The two instruments compared in Qian's study (2019) are shown in **Figure 3a**.

**Figure 3.** *3D-printed musical instruments.* ***a:*** *Printed ukulele (**left**) and wood instrument (**right**). Reproduced from Qian (2019), with the permission of the American Institute of Physics (AIP).* ***b:*** *A recorder that has been optimized using computer software (**left**) and a printed prototype (**right**). Reproduced from Thacker and Giordano (2021), with permission from the American Institute of Physics (AIP).*


AM offers new capabilities enabled by the ability to create complex designs that are not possible using standard manufacturing techniques. For example, Ritz et al. (2015) investigated the field of microtonal music, which employs more equally spaced intervals in an octave than are employed in a standard 12 semitone equal temperament used in most Western music. By using AM technology to create a double-helix flute, they were able to exploit geometry to create these tones. Taking a more physics-based design approach, Thacker and Giordano (2021) used fluid computational approaches to design improved recorder instruments which were then fabricated with AM and their performance compared with the fluid model (**Figure 3b**).

## Applications to Acoustic Metamaterials

Acoustic metamaterials are engineered structures that can control acoustic waves in ways that are not possible with typical "bulk" materials such as steel or plastics (Haberman and Norris, 2016). Acoustic metamaterials are usually produced by constructing a material from periodically repeating "unit cells" that, when properly designed, exhibit novel acoustic behavior like band gaps (a range of frequencies where waves cannot propagate through the material). One example of this behavior can be observed in an artistic structure in Madrid that is composed of periodically arranged long metal cylinders (Martinez-Sala et al., 1995). This structure has a band gap in the audible range (around 1,600 Hz), so if you were to play a 1,600-Hz tone on one side of the structure, it would be filtered by the structure rendering it inaudible on the other side. Like this structure, acoustic metamaterials can act as a "filter" for acoustic waves.

Although research on acoustic metamaterials has been ongoing for several decades, many of these ideas have only recently taken form with the advancement of AM. The reason is that AM can create objects with intricate and complex geometries, such as the structures shown in **Figure 4, a** and **c**, that are impossible to create using traditional manufacturing techniques. Metamaterial

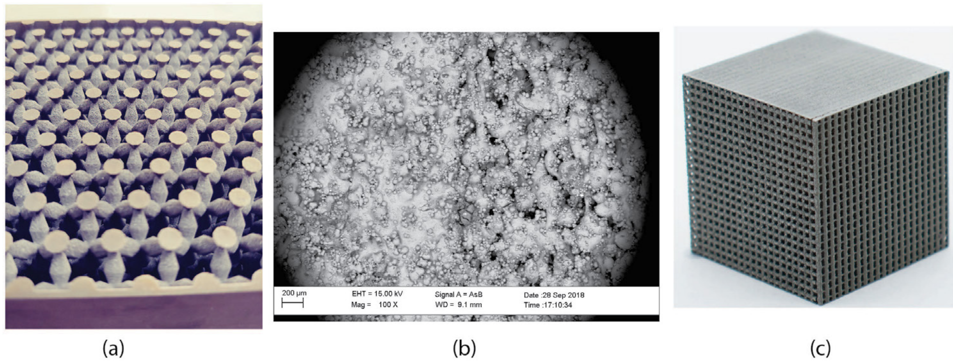



**Figure 4.** *Examples of AM for phononic crystals and acoustic metamaterials.* ***a:*** *Acoustic metamaterial that may enable acoustic cloaking, fabricated from titanium using a laser powder bed fusion (LPBF) technique known as direct metal laser sintering. Reproduced from Cushing (2022), with permission from the AIP.* ***b:*** *Acoustic metal foam with prescribed porosity enabled by metal LPBF. Reproduced from Konarski (2021), with permission from the AIP.* ***c:*** *AM acoustic "filter" that can be used to improve nondestructive evaluation measurements. Reproduced from Smith and Matlack (2021), with permission from the AIP.*

structures have enabled researchers to explore materials that exhibit acoustic phenomena like acoustic cloaking (Cushing, 2022). In addition, small changes to complex geometric structures can allow researchers to finely tune the frequency range over which the acoustic metamaterial operates (Arretche and Matlack, 2018). Recent work has even shown that AM can be used to create metamaterials with porosity as an additional design variable, as shown in **Figure 4b** (Konarski, 2021). Because AM makes it possible to fabricate complex geometries in many different length scales, acoustic metamaterials can now operate over a wide range of frequencies, from hertz up to megahertz frequencies.

AM has also opened the door to various applications of acoustic metamaterials, such as eliminating damaging vibrations from structures (Arretche and Matlack, 2018; Gerard et al., 2021). Recent work has shown how acoustic metamaterials that are highly anisotropic, meaning that their mechanical properties are different along different directions, can be used to guide acoustic waves that have large wavelengths compared with the size of the metamaterial (Yves and Alù, 2021). Such materials could be fabricated using AM techniques. Continued advancements in AM, including honing our ability to print multiple materials in the same structure or even print more advanced materials like piezoelectrics (Lewis, 2006), will certainly push what is possible in terms of acoustic wave control with acoustic metamaterials.

## Applications in Ultrasonic Nondestructive Evaluation

The discrete layer-by-layer approach of AM means that the resulting materials, and thus their mechanical response, can be very different from their traditionally manufactured counterparts. Furthermore, seemingly minor differences from one print to another, even using the same machine, can result in materials with different mechanical properties. One application of acoustics has been to use ultrasound as a nondestructive evaluation (NDE) tool to determine the mechanical and microstructural properties of AM parts. For AM to be successfully adopted by industries that require highly precise part creation (e.g., nuclear, automotive, aircraft), reliable and fast NDE methods to qualify, characterize, and quantify damage in these new materials is crucial. The ability to accurately measure the properties of materials created using AM remains a critical challenge that, when properly addressed, will enable more widespread adoption of AM. Ultrasonic NDE methods are one promising approach to evaluate AM materials.

Various ultrasonic inspection techniques have been applied to AM structures, particularly metals. Ultrasonic parameters such as the wave velocity, attenuation, and nonlinearity coefficients have been shown to be capable of sensing AM-specific features such as texture or pores (Kim et al., 2021) and nanometer-sized defects (Bellotti, 2021). One way to assess these properties is to use a Rayleigh wave measurement setup as shown by Bellotti et al. (2021). In that work, waves were generated in an AM-fabricated specimen using a transducer configured to excite waves on the surface of the material. An air-coupled transducer then measured the response of the material using the airborne acoustic wave generated from the surface motion of AM material. Using this approach, the properties of the AM-fabricated material can be extracted. This ultrasonic technique measured the acoustic response from steels manufactured using various AM techniques and compared observations with steels made using traditional methods. The results showed different acoustic responses for the different manufacturing methods, indicating that materials created using AM fabrication techniques should be carefully evaluated prior to use in critical components. Other recent NDE research relevant to AM fabrication has shown the possibility of using ultrasonic methods to determine the mechanical parameters and print quality in real time as the part is printed (Gillespie, 2021). Additional recent work has also shown how acoustic metamaterials can be used to enhance ultrasonic NDE measurements by creating filtering materials (Smith and Matlack, 2021) to better isolate the ultrasonic response or a portion of the structure of interest. This and similar approaches to ultrasonic NDE are uniquely enabled by AM.

## Acoustic Transduction Materials and Devices

The ability to generate and sense sound has always been central to the study, application, and enjoyment of acoustics. As a result, there has been considerable exploration of the structure and materials used to construct acoustical devices. Present-day examples range from consumer electronics with exotic geometries and transduction components to mass-manufactured microelectromechanical systems (MEMS) microphones and accelerometers in smart devices to high-precision ultrasonic measurement systems. Noting that key developments in the science and technology of acoustic transducers are often enabled by technological advances, AM technology provides an exciting opportunity to explore how to improve well-established approaches to acoustic transduction.

An obvious application example of AM technology is the ability to create highly unique geometries that cannot be created using traditional manufacturing techniques. A notable recent work by Nielsen et al. (2021) considered this case in a numerical study that considered the values and distribution of "stiffness, mass, and damping of both the speaker diaphragm and surround" to optimize loudspeaker response.

One of the more interesting prospects of AM technology is the potential to directly print the transducer components or materials (Chen et al., 2020) sand do so in a streamlined process that could be extended to include fabrication and assembly of electrical, mechanical, and transducing components (Ambriz et al., 2017). Kierzewski et al. (2020) created a macroscopic embodiment of piezoelectric material that was first described by Bauer et al. (2004). The work of Kierzewski et al. (2020) leveraged the geometric freedom offered by AM paired with a multistep assembly process to essentially replicate the response of a condenser microphone and extend it to a collection of cavities. Although not currently ideal for application, this work shows the relative ease of creating true transducing "materials" using AM techniques that have not yet been fully leveraged for this type of technology.

Additive technology is also of interest for the direct manufacture of transduction materials. The most common materials in accelerometers and underwater transducers are piezoelectric ceramics. There is considerable progress on various manufacturing techniques to print piezoelectric ceramics using approaches like selective laser sintering and paste extrusion followed by postprocessing (reviewed by Chen et al., 2020). Cui et al. (2019) have taken a very different approach by investigating novel AM techniques to create lattice structures that display piezoelectric coupling with tailored anisotropy and directional sensitivity that could ultimately be used for a wide range of applications. The opportunities that arise in being able to create active materials with conformal geometries, tailored piezoelectric coupling constants, and multimaterial components have a vast potential to significantly alter how vibroacoustic transducers are created in the future.

## Hearing Prostheses and Hearing Aids

In the early stages of development, AM technology was collectively referred to as "rapid prototyping" due to the fact that part quality was insufficient for use in functional parts or products. One of the earliest examples of the transition of AM technology from rapid prototyping to "rapid production" was in the field of hearing aid technology (Widmer and Dutta, 2005). Most hearing aids are constructed using a hard, external shell that fits within the ear canal and contains the electroacoustic components: microphone(s), signal processing and amplification electronics, sound amplification transducer, and the battery. The electroacoustic components are the same within any given product line, and the signal-processing and amplification characteristics can be tailored to the individual user according to the specific hearing impairment using a simple programming step after audiological evaluation.

In contrast, the shape of the shell itself is highly unique due to the need to correctly position the device in the ear canal and maintain a physical seal between the device and ear canal. This shape cannot be easily reconfigured and thus each individual shell must be custom constructed. Furthermore, traditional fabrication methods, such as UV-cured polymers molded to fit each user, require a significant effort by individual technicians with years of experience. This approach to mass customization of components resulted in low device-to-device repeatability, even when starting with the same biometric customer information.

For all the reasons given, these types of hearing devices are particularly well suited for the strengths of LPBF and SLA technology. Namely, AM fabrication of custom hearing devices exploits the fact that (1) each part has a highly irregular and custom shape in order to fit within the ear canal of the individual user; (2) the components are physically small, and thus parts for different customers may be fabricated in a single build and each build can be different without changing the fabrication settings; (3) the final product is subjected to mild environmental mechanical loading over the entire life cycle; and (4) 3D optical scanning can be used to gather the ear canal geometry, and thus final parts can be created with virtually no contact with the customer. Given all of these benefits, there has been a massive transition to rapid production of hearing devices since the early 2000s, including nonacoustic or mechanical improvements such as incorporating antibiofilm properties into the printed material to reduce ear infection (Vivero-Lopez et al., 2021).

As an example, **Figure 5** shows photos of 3D-printed hearing aids printed with a digital light processing polymer, a liquid resin-type process, which has been implanted with an antibiofilm drug in varying quantities. In this case, the drugs are mixed into the liquid resin

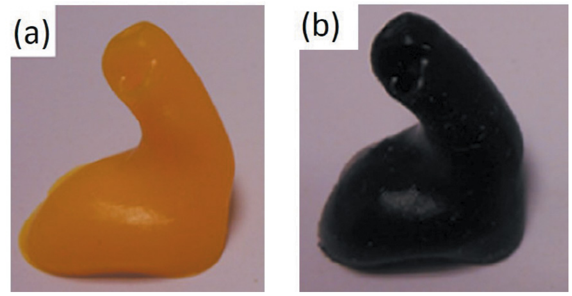



**Figure 5.** *AM hearing aids made from different resins.* ***a:*** *Sample is a resin.* ***b:*** *Sample is the result of printing the same hearing aid shell from a flexible resin. The samples were manufactured with differing levels of an antibiotic in the material to investigate the ability of the material to reduce ear infection from long-duration hearing aid use. Reproduced from Vivero-Lopez et al. 2021, Figure 1), with permission from Elsevier.*

before printing. The study also used different materials, both a hard and a flexible resin, to assess the resulting variations in the printed final product.

## Conclusion

This article provided a multidisciplinary review of current uses for AM in acoustics. AM can be applied to a wide swath of acoustic applications, including musical instruments, scale models, acoustic metamaterials, ultrasonic NDE, transducers, and hearing prostheses. These examples highlight how AM is pushing the frontiers of acoustics, and conversely, how acoustics can be used to advance different AM technologies. Advances in both acoustic design and AM technology mean that the range and depth of these applications will certainly expand in the future. For example, advances in AM that push the limitations of a smallest printable feature size would expand the frequency range of acoustics applications, and advances that increase the maximum size of AM parts could enable larger scale acoustics applications such as noise control of large structures. From another perspective, advances in acoustics and ultrasonics could enable better quality control of AM parts that would allow safety-critical industries to adopt AM technologies. Additionally, reduced costs for basic printing approaches mean that many of these applications will become more and more accessible to acoustic researchers and hobbyists alike.

## Acknowledgments

Funding was provided in part by the Internal Research and Development Program, Applied Research Laboratories, The University of Texas at Austin.

### References

Ambriz, S., Coronel, J., Zinniel, B., Schloesser, R., Kim, C., Perez, M., Espalin, D., and Wicker, R. B. (2017). Material handling and registration for an additive manufacturing-based hybrid system. *Journal of Manufacturing Systems* 45, 17-27. https://doi.org/10.1016/J.JMSY.2017.07.003.

Arretche, I., and Matlack, K. H. (2018). On the interrelationship between static and vibration mitigation properties of architected metastructures. *Frontiers in Materials* 5, 1-16. https://doi.org/10.3389/fmats.2018.00068.

Bauer, S., Gerhard-Multhaupt, R., and Sessler, G. M. (2004). Ferroelectrets: Soft electroactive foams for transducers. *Physics Today* 57(2), 37-43. https://doi.org/10.1063/1.1688068.

Bellotti, A., Kim, J.-Y., Bishop, J. E., Jared, B. H., Johnson, K., Susan, D., Noell, P. J., and Jacobs, L. J. (2021). Nonlinear ultrasonic technique for the characterization of microstructure in additive materials. *The Journal of the Acoustical Society of America* 149, 158-166. https://doi.org/10.1121/10.0002960.

Chen, C., Wang, X., Wang, Y., Yang, D., Yao, F., Zhang, W., Wang, B., Sewvandi, G. A., Yang, D., and Hu, D. (2020). Additive manufacturing of piezoelectric materials. *Advanced Functional Materials*, 30, 2005141. https://doi.org/10.1002/adfm.202005141.

Cottrell, S., and Howell, J. (2019) Reproducing musical instrument components from manufacturers' technical drawings using 3D printing: Boosey & Hawkes as a case study. *Journal of New Music Research* 48(5), 449-457. https://doi.org/10.1080/09298215.2019.1642362.

Cui, H., Hensleigh, R., Yao, D., Maurya, D., Kumar, P., Kang, M. G., Priya, S., and Zheng, X. R. (2019). Three-dimensional printing of piezoelectric materials with designed anisotropy and directional response. *Nature Materials* 18, 234-241. https://doi.org/10.1038/s41563-018-0268-1.

Cushing, C.W, Kelsten, M.J. Su, X., Wilson, P.S., Haberman, M.R., Norris, A.N., (2022). Design and characterization of a three-dimensional anisotropic additively manufactured pentamode material. *The Journal of the Acoustical Society of America*, 151(1), 168-179. https://asa.scitation.org/doi/10.1121/10.0009161

Gerard, N. H. Oudich, M., Xu, Z., Yao, D., Cui, H., Naify, C. J., Ikei, A., Rohde, C. A., Zheng, X. R., and Jing, Y. (2021). Three-dimensional trampoline-like behavior in an ultralight elastic metamaterial. *Physical Review Applied* 16, 024015. https://doiorg/10.1103/PhysRevApplied.16.024015.

Gillespie, J., Yeoh, W. Y., Zhao, C., Parab, N. D., Sun, T., Rollett, A. D., Lan, B., and Kube, C. M. (2021). In situ characterization of laser-generated melt pools using synchronized ultrasound and high-speed X-ray imaging. *The Journal of the Acoustical Society of America* 150(4), 2409-2420. https://doi.org/10.1121/10.0006386.

Goh, G. L., Zhang, H., Chong, T. H., and Yeong, W. Y. (2021). 3D printing of multilayered and multimaterial electronics: a review. *Advanced Electronic Materials* 7, 2100445. https://doi.org/10.1002/aelm.202100445.

Haberman, M. R., and Norris, A. N. (2016). Acoustic metamaterials. *Acoustics Today* 12(3), 31-39.

Ishiguro, Y., and Poupyrev, I., (2014). 3D printed interactive speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, Association for Computing Machinery, Toronto, ON, Canada, April 26 to May 1, 2014, pp. 1733-1742. https://doi.org/10.1145/2556288.2557046.

Kierzewski, I., Bedair, S. S., Hanrahan, B., Tsang, H., Hu, L., and Lazarus, N. (2020). Adding an electroactive response to 3D printed materials: Printing a piezoelectret. *Additive Manufacturing* 31, 100963. https://doi.org/10.1016/j.addma.2019.100963.

Kim, C., Yin, H., Shmatok, A., Prorok, B. C., Lou, X., and Matlack, K. H. (2021). Ultrasonic nondestructive evaluation of laser powder bed fusion 316L stainless steel. *Additive Manufacturing* 38, 101800. https://doiorg/10.1016/j.addma.2020.101800.

Konarski, S. G., Rohde, C. A., Gotoh, R., Roberts, S. N., and Naify C. J. (2021). Acoustic measurement and statistical characterization of direct-printed, variable-porosity aluminum foams. *The Journal of the Acoustical Society of America* 149, 4327-4336 https://doi.org/10.1121/10.0005273.

Lewis, J. A. (2006). Direct ink writing of 3D functional materials. *Advanced Functional Materials*, 16, 2193-2204. https://doi.org/10.1002/adfm.200600434.

Mannoor, S., Jiang, Z., James, T., Kong, Y. L., Malatesta, K. A., Soboyejo, W. O., Verma, N., Gracias, D. H., and McAlpine, M. C. (2013). 3D printed bionic ears. *Nano Letters* 13(6), 2634-2639. https://doi.org/10.1021/nl4007744.

Martinez-Sala, R., Sancho, J., Sanchez, J. V., Gomez, V., and Llinares, J. (1995). Sound attenuation by sculpture. *Nature* 378, 241. https://doi.org/10.1038/378241a0.

Michon, R., Chaf, C., and Gransow, J. (2018). 3D printing and physical modeling of musical instruments: Casting the net. In *Proceedings of the 15th Sound and Music Computing Conference* (SMC2018), Limassol, Cyprus, July 4-7, 2018.

Ngo, T., Kashani, A., Imbalzano, G., Nguyen, K., and Hui, D. (2018). Additive manufacturing (3D printing): A review of materials, methods, applications and challenges. *Composites Part B: Engineering* 143, 172-196. https://doi.org/10.1016/j.compositesb.2018.02.012.

Nielsen, D. G. Agerkvist, F. T., and Jensen, J. S. (2021). Achieving a flat, wideband frequency response of a loudspeaker unit by numerical optimization with requirements on its directivity. *The Journal of the Acoustical Society of America* 150, 663-672. https://doi.org/10.1121/10.0005731.

Paolini, A., Kollmannsberger, S., and Rank, E. (2019). Additive manufacturing in construction: A review on processes, applications, and digital planning methods. *Additive Manufacturing* 30, 100894. https://doi.org/10.1016/j.addma.2019.100894.

Prepelită, S. T., Bolaños, J. G., Geronazzo, M., Mehra, R., and Savioja, L. (2020). Pinna-related transfer functions and lossless wave equation using finite-difference methods: Validation with measurements. *The Journal of the Acoustical Society of America* 147, 3631-3645. https://doi.org/10.1121/10.0001230.

Qian, P., Niu, X., Lin S., Ma, L., Ma, J., Yu, S., Chen, Z., Li, G., Fu, S., and Lin, J. (2019). A comparison on sound quality of PLA 3-D printing ukulele and single board wooden ukulele. *Proceedings of Meetings on Acoustics* 39, 035007. https://doi.org/10.1121/2.0001284.

Ritz, C., Dabin, M., Narushima, T., Grady, K., and Beirne, S. (2015). 3D printing for custom design and manufacture of microtonal flutes. *SPIE Newsroom* 0-3. https://doi.org/10.1117/2.1201508.006082.

Smith, E. J., and Matlack, K. H. (2021). Metal additively manufactured phononic materials as ultrasonic filters in nonlinear ultrasound measurements. *The Journal of the Acoustical Society of America* 149, 3739-3750. https://doi.org/10.1121/10.0004995.

Sun, J., Zhou, W., Huang, D., Fuh, J. Y., and Hong, G. S. (2015). An overview of 3D printing technologies for food fabrication. *Food Bioprocess Technology* 8, 1605-1615. https://doi.org/10.1007/s11947-015-1528-6.

Thacker, J. W., and Giordano, N. (2021), Regime change in the recorder: Using Navier-Stokes modeling to design a better instrument. *The Journal of the Acoustical Society of America* 150, 43-50. https://doi.org/10.1121/10.0005317.
Vivero-Lopez, M., Xu, X., Muras, A., Otero, A., Concheiro, A., Gaisford, S., Basit, A. W., Alvarez-Lorenzo, C., and Goyanes, A. (2021). Anti-biofilm multi drug-loaded 3D printed hearing aids. *Materials Science and Engineering: C* 119, 111606. https://doi.org/10.1016/j.msec.2020.111606.
Widmer, C., and Dutta, J. (2005). *Method for Manufacturing an Ear Device and Ear Device*. US Patent No.6863151, March 8, 2005.
Yves, S., and Alù, A. (2021). Extreme anisotropy and dispersion engineering in locally resonant acoustic metamaterials. *The Journal of the Acoustical Society of America* 150, 2040-2045. https://doi.org/10.1121/10.0006237.

## About the Authors

**Christina J. Naify**
christina.naify@arlut.utexas.edu

*Applied Research Laboratories*
*University of Texas at Austin*
*1000 Burnet Road*
*Austin, Texas 78758, USA*

**Christina J. Naify** is a research associate in the Applied Research Laboratories, University of Texas at Austin. Dr. Naify is the current chair of the Tutorials/Short Course/Hot Topics Committee and the Structural Acoustics and Vibration Technical Committee. Her research interests include additive manufacturing, materials characterization, and acoustic metamaterials for air and water environments. Dr. Naify served as guest editor for the special issue on “Additive Manufacturing and Acoustics” of *The Journal of the Acoustical Society of America*.

**Kathryn H. Matlack**
kmatlack@illinois.edu

*Department of Mechanical Science and Engineering*
*University of Illinois at Urbana-Champaign*
*1206 W. Green Street*
*Urbana, Illinois 61801, USA*

**Kathryn H. Matlack** is an assistant professor in the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, where she leads the Wave Propagation and Metamaterials Laboratory. She is a member of the Structural Acoustics and Vibrations Technical Committee of the Acoustical Society of America. Her research interests include phononic materials, mechanical metamaterials, ultrasonic nondestructive evaluation, and additive manufacturing. Dr. Matlack served as guest editor for the special issue on “Additive Manufacturing and Acoustics” of *The Journal of the Acoustical Society of America*.

**Michael R. Haberman**
haberman@utexas.edu

*Walker Department of Mechanical Engineering*
*The University of Texas at Austin*
*204 E. Dean Keeton Street*
*Austin, Texas 78712, USA*

**Michael R. Haberman** is an assistant professor in the Department of Mechanical Engineering, The University of Texas at Austin (UT), with an appointment in the Applied Research Laboratories at UT. His research interests include elastic and acoustic wave propagation in complex media, acoustic metamaterials, and transduction materials. Dr. Haberman served as guest editor for the special issue on “Additive Manufacturing and Acoustics” of *The Journal of the Acoustical Society of America* (*JASA*). He serves as an associate editor of *JASA* and chair of the Engineering Acoustics Technical Committee of the Acoustical Society of America.

# The Perception and Measurement of Headphone Sound Quality: What Do Listeners Prefer?

Sean E. Olive

Headphones are the primary means through which we listen to music, movies, and other forms of infotainment. They have become an indispensable accessory for our mobile phones, providing a 24/7 connection to our entertainment, colleagues, and loved ones. This trend is reflected in the exponential growth in sales. The global market for wireless headphones alone was estimated at $15.9B in 2020 and is projected to rise to $45.7B by 2026, a compound annual growth rate of 19.1% (PRNewsWire, 2021). With this growth has come a renewed interest in improving the sound quality of headphones.

Unfortunately, headphone sound quality has not kept pace with consumers' demands and expectations. Two recent studies have measured the variance in frequency response of more than 400 headphones and found no correlation between their retail price and frequency response (Breebaart, 2017; Olive et. al., 2018a). They included the three most common types: headphones that fit around the ear (AE), on the ear (OE), and in the ear (IE). It seems that headphone designers are aiming at a target frequency response that is as random and variable as the weather.

Another telling sign that headphone sound quality has not kept pace is that headphone industry standards have not changed fundamentally since the 1990s. The International Electrotechnical Commission (IEC) 60268-7 (2010) standard specifies multiple ways to measure the frequency response of a headphone for both free-field (FF) and diffuse-field (DF) targets, with the warning: "subjective assessments are still useful because the objective methods whose results bear good relation to those from subjective assessments are under research stage" (IEC, 2010, Section 8.6.1). This does not inspire confidence.

The International Telecommunication Union Radiocommunication Assembly (ITU-R) BS.708 (1990) standard recommends that professional headphones be designed to the DF target curve to achieve best sound, but most headphone designers have rejected this suggestion and probably for good reasons. Recent psychoacoustic investigations provide evidence that listeners prefer alternative headphone targets to DF and FF target standards (Olive et al., 2013a).

The chaos that exists within the headphone industry today is reminiscent of the loudspeaker industry 30 years ago when there was insufficient knowledge on listeners' loudspeaker preferences and which loudspeaker measurements best predict them. The situation improved after Floyd Toole, an acoustician at the National Research Council of Canada, published seminal scientific papers that provided guidelines in how to measure and design loudspeakers that most listeners prefer (Toole, 1985, 1986). Later, a mathematical model was developed that could predict listeners' preference ratings of the loudspeakers based on objective measurements alone (Olive, 2004). The science provided important answers on what loudspeaker listeners prefer, design guidelines, and new measurement standards (American National Standards Institute/Consumer Technology Association [ANSI/CTA] Standard, 2015) that became widely accepted and adopted throughout the industry.

## Headphone Sound Quality

In 2012, the seminal papers for headphone sound quality did not exist, and this was reflected in the headphone standards and the large variance in headphone sound quality. Skeptics argued that the variance in headphone sound was explained by a need to satisfy individual tastes in sound that vary like individual tastes in music, food, and preferred companions. If listeners could not agree on what sounds good, then a single optimal frequency response or headphone target curve could not be defined.



https://doi.org/10.1121/AT.2022.18.1.58

These same arguments were undoubtably made about loudspeakers 40 years ago and until research proved listeners largely agreed on what is a good loudspeaker.

With the lessons learned from the loudspeaker industry, the author and his colleagues embarked on a seven-year research project to improve the consistency and sound quality of headphones. There were three fundamental questions we hoped to answer.

(1) What is the preferred headphone target curve? Should the reference be a loudspeaker in a FF, a DF, or a semireflective field (SRF) found in a typical listening room?
(2) Do listeners agree on what makes a headphone sound good? To what extent does listening experience, age, gender, and geographical location influence sound quality preferences?
(3) Can listeners' subjective ratings of headphones be predicted based on an objective measurement?

These research questions were addressed for the three main headphone types, but the scope of this article is largely restricted to AE and OE headphones. The preferred target curve for IE headphones is almost identical to those for the AE and OE targets, except it has an additional 4 dB of bass (Olive et al., 2016). Each question is addressed separately, followed by conclusions.

## The Search for the Preferred Headphone Target Curve

Over the past 50 years, headphone researchers have focused their attention on determining what the ideal reference sound field should be for headphone reproduction and how to measure it. Three types of reference sound fields have been proposed: a FF, a DF and a SRF that lies somewhere between the two extremes. What these sound fields are, how they are measured or derived, and psychoacoustic investigations of headphone target curves based on them are described.

### *Free-Field Headphone Target Curve (1970s)*

The reference FF was generated by placing a loudspeaker in front of the listener in a reflection-free room. A tedious subjective loudness-matching procedure was used where a test subject would listen to narrow bands of noise at different frequencies alternately with the FF (with the headphone removed) and then with the headphone. While listening to the headphones, the levels for each band would be adjusted to match the loudness of the loudspeaker. This would be repeated for several test subjects to calculate the loudness transfer function that defined the headphone FF target curve.

Theile (1986) conducted formal listening tests and found the DF target to be preferred to the FF target, which produced an unnatural timbre and in-head localization effects. Although the FF target fell out of favor beginning in the 1980s, it remains part of the current headphone IEC (2010) standard today.

### *Diffuse-Field Headphone Equalizations (1980s to Present)*

A DF occurs when a sound source is placed in a reverberation room with little or no absorption, so the listener receives a random and equal distribution of sounds from all directions. The headphones are calibrated to the DF using a subjective loudness procedure or alternative methods. In one method, a probe microphone is placed in the ear canals of the listener to measure and then match the transfer function of the headphone to that of the sound field (Theile, 1986).

A second approach is to substitute the listener with a head and torso simulator (HATS); this produces faster, more reproducible, and safer measurements than putting probe microphones in the listeners' ears. A third option is to use a headphone known to be DF calibrated as the reference and compare its performance with the headphone under test.

Møller et al. (1995) derived a headphone target curve based on different sound fields using a large set of head-related transfer functions (HRTFs) measured at the blocked ear canal. HRTFs define the transfer functions, both the frequency and phase responses at the entrance to the ear, for each direction and distance of a sound source. They capture both interaural time (ITD) and intensity (IID) differences and spectral cues that humans use to localize sound sources in space (Blauert, 1983). By selecting HRTFs from the appropriate directions and distances and integrating them, Møller et al. (1995) were able to derive transfer functions of reference sound fields ranging from the FF to the DF and anything in between. This method eliminated the need for a physical reference sound field, making headphone calibration more practical and reproducible. A headphone could be measured and equalized to the DF target curve using a calibrated dummy head or ear simulator.

The DF target was not seriously challenged until Lorho (2009) reported 80 listeners (25% audio engineers, 25% music students, and 50% naive listeners) on average preferred a significantly modified version of the DF target where its main feature, a wide 12 dB peak at 3 kHz, was reduced to just 3 dB. This paper sparked new interest to find better alternative headphone target curves to the ones recommended in the current headphone standards.

## *Semireflective Field Headphone Equalizations (2012 to Present)*

Because stereo recordings are optimized for reproduction through loudspeakers in semireflective rooms, they should sound best through headphones that emulate this sound field. Sank (1980) made similar proposals three decades earlier but never conducted formal listening tests that compared these targets with the DF target.

Loudspeakers with flat on-axis and smooth off-axis frequency responses tend to produce the highest subjective ratings in formal listening tests (Toole, 2018). When placed in a typical room, they produce a uniform quality of direct, early, and late reflected sounds that in summation produce the steady-state in-room response of the loudspeaker. Due to the frequency-dependent directivity of the loudspeaker and absorption characteristics of the room, the in-room response will not be flat like the FF response nor the same as the DF response where the room absorption has been removed. Instead, the in-room response gently falls about 1 dB per octave from 20 Hz to 20 kHz.

Fleishmann et al. (2012) reported the first formal listening test results where three SRF headphone targets were evaluated. The targets were based on measurements of the steady-state in-room response of a 5.1-channel loudspeaker setup in a standard listening room and then equalized by three expert listeners to match the timbre of the speakers. Two of the SRF targets were found to be slightly preferred to the DF target, depending on the music programs. Other targets included the Lorho target, a flat target, and three unequalized headphones that generally received lower ratings than the two SRF targets. Unfortunately, no measurements or details of the loudspeakers and the three SRF targets were given. The conclusions were that the SRF targets were equal to or better than the DF target, but the Lorho target was not.

A similar study (Olive et al., 2013a) reported evidence that listeners strongly preferred headphones equalized to SRF targets to, in descending order of preference, two DF targets (Möller et al., 1995); two high-quality headphones; the Lorho target; and the FF target. The trained listeners described both DF targets as having too much emphasis in the upper midrange (2-4 kHz) and lacking bass. The Lorho target had too little energy at 2-4 kHz, which made instruments sound "muffled and dull." The FF target was strongly criticized for its strong emphasis between 2 and 4 kHz, lack of bass, and harsh and nasal colorations. Listeners described the highest rated the SRF target as having "good bass with an even spectral balance." The measured frequency responses of the headphone targets correlate to and confirm listeners' descriptions of their sound quality (see Olive et al., 2013b, Figure 2). The highest rated target curve in this study soon became known in the audio industry as the Harman target curve and is widely influencing the design, testing, and review of headphones.

## Do Listeners Agree on What Makes a Headphone Sound Good?

Although the initial test results of the Harman target curve were encouraging, they were based on a small sample of 10 trained listeners. To better understand if certain demographic factors influence the acceptance of the curve, it was tested using a larger number of listeners from a broad range of ages, listening experiences, and geographic regions.

The target curve was benchmarked against three headphones considered industry references at the time in terms of sound quality or commercial sales (Olive et al., 2014). They ranged in price from $269 to $1,500 and included dynamic and magnetic planar transducer designs. A total of 283 listeners participated from four different countries (Canada, United States, Germany, and China) and included a broad range of ages, listening experiences, and genders. Most of the participants were Harman employees.

A novel virtual headphone test methodology allowed controlled, rapid, double-blind comparisons among the different headphones. Virtual versions of the different headphones were reproduced over a single high-quality replicator headphone by equalizing it to match the measured frequency response of each headphone. This removed

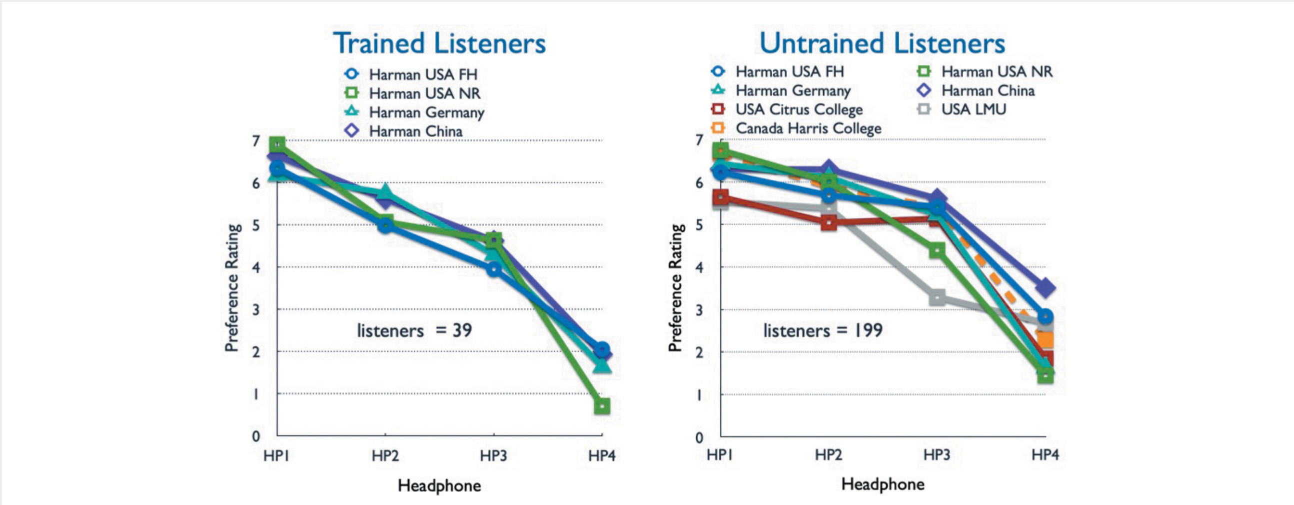



**Figure 1.** *The mean preference ratings are shown for 11 different groups of listeners categorized as trained (**left**) and untrained (**right**). The tests were administered in four different countries: Canada, United States, Germany, and China. HP1 is the Harman target curve and HP2 and HP3 are high-quality, high-priced headphones. HP4 was the most popular headphone in terms of sales (Olive et al., 2014).*

any potential biases related to visual (brand, model, price, design) and tactile (weight, clamping force, feel of materials) cues that might cloud their judgments of sound quality. A prior validation study confirmed that subjective ratings of virtual versus actual headphones (with the listener unaware of the headphone brand, model or appearance) had a correlation of 0.86 to 0.99 depending on the headphone type (Olive et al., 2013b). A limitation of the method is that it does not reproduce nonlinear distortions in the headphones. However, the high correlations between virtual and actual headphone comparisons and evidence from other studies indicate that these distortions are generally below masked thresholds (Temme et al., 2014).

The results show that headphone preferences were remarkably consistent across the 11 test locations for both trained and untrained listeners (**Figure 1**). As expected, the trained listeners were more discriminating and consistent than the untrained listeners.

Headphone preferences were also relatively consistent across different age groups and the four countries. The exception was listeners in the 55+-year age category who tended to prefer HP2, a brighter headphone with less bass than the Harman target curve. A possible explanation could be age related hearing loss; additional treble and less bass can help improve intelligibility. More research is needed to provide definitive answers.

## Preferred Level of Bass and Treble in Headphones

The same group of listeners participated in a second experiment where they adjusted the bass and treble levels of the headphone (Olive and Welti, 2015) several times according to taste using different samples of music. The listeners' preferred levels were influenced by several factors, including the music program, as well as by the subject's age, gender, and prior listening experience (see **Figure 2**). The program interactions between preferred bass and treble levels are expected due to variability in the quality of music recordings; often they require adjustments in bass and treble on playback to restore a proper balance. Toole (2018) refers to these errors as audio's "circle of confusion." The confusion arises from not knowing the source of these errors: the recording, the loudspeaker, or its interaction with the room acoustics. The solution is a meaningful loudspeaker standard common to both the professional and consumer audio industries.

Female listeners preferred less bass and treble than their male counterparts. Younger and less experienced listeners

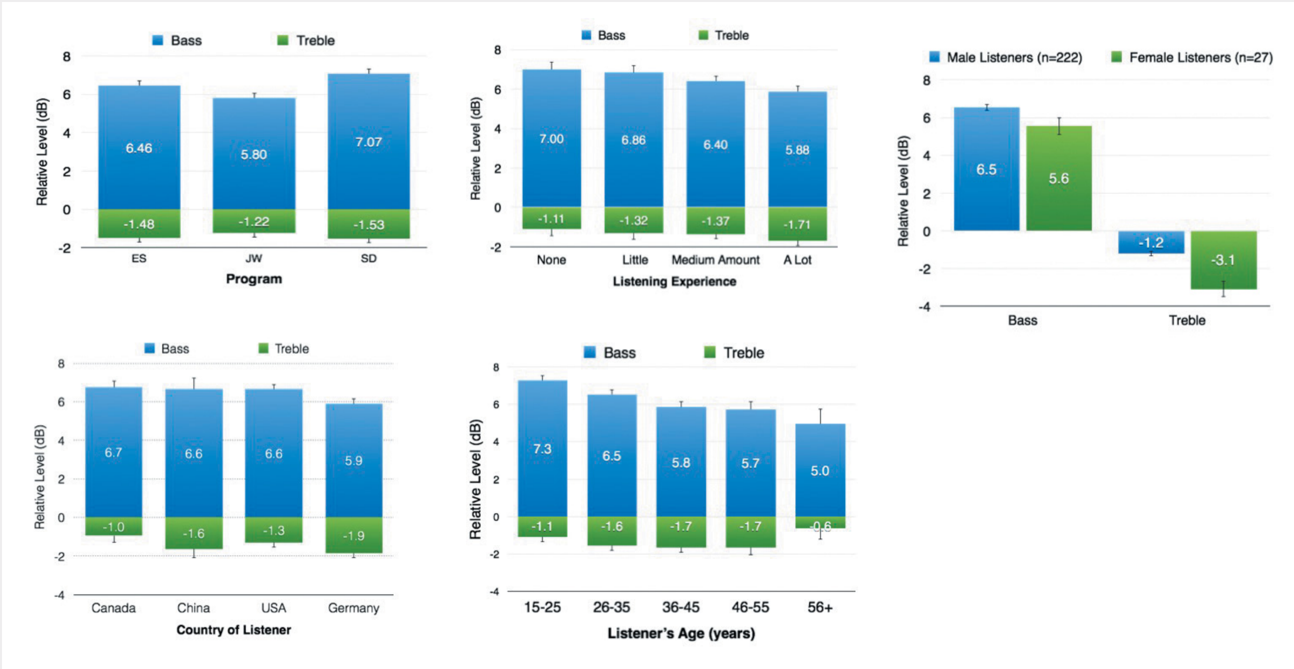



**Figure 2.** *The mean bass and treble levels and 95% confidence intervals for a headphone calibrated to match a flat in-room loudspeaker response. Each graph shows the interaction effect between the preferred levels and program, gender, listening experience, age, and the country of the test location (Olive and Welti, 2015).*

preferred more bass and treble than their older, more experienced counterparts. The older listeners (55+ years) were the exception here, preferring significantly more treble and less bass, consistent with their preference for headphone HP2. Altogether, these findings suggest that a single headphone target may not be sufficient to satisfy variations in the recordings, individual tastes, listening experience, and hearing loss. A simple solution for headphone personalization is to provide a simple bass and treble control that allows listeners to compensate for these variations.

## Testing the Harman Target with Larger Sample of Headphones

The next goal was to test the Harman target using a larger population of headphones. A total of 31 different headphone models from 18 manufacturers were evaluated by 130 listeners, with an approximately equal number trained and untrained (Olive et al., 2018a). The headphones ranged in price from $60 to $4,000, including open and closed back designs with dynamic or magnetic planar drivers. The same virtual headphone double-blind method was used to eliminate biases from visual and tactile cues. The results establish that, on average, both trained and untrained listeners preferred the headphone equalized to the Harman target in 28 of the models tested. Four models with frequency responses close to the Harman target were equally preferred.

## Segmentation of Listeners Based on Preferred Headphone Sound Profiles

Although the study established that listeners, on average, preferred the Harman target to other headphones tested, it had not explored whether segments or classes of listeners exist based on similarities in their headphone preferences and what those sound quality features or profiles are. Also, it did not identify possible underlying demographic factors that might predict membership in each class. There was already prior evidence that younger males and less experienced listeners preferred higher levels of bass and treble in their headphones compared with females, experienced, and older listeners (Olive et al., 2013a; Olive and Welti, 2015). A reasonable hypothesis was that segmentation of headphone preferences may relate to bass and treble levels, possibly predicted by these demographic factors.

A statistical method known as agglomerative hierarchical clustering exposed three different segments or classes of listeners based on similarities in their headphone preferences. By calculating the average response of the top five preferred headphones in each class, it was clear that the preferred bass level is the main feature that defines membership in a class. Class 1 includes most listeners (64%) who prefer headphones that closely comply with the Harman target. Class 2 listeners (15%) prefer the Harman target with 4-6 dB more bass. Class 3 listeners (21%) prefer the Harman target curve with 2 dB less bass.

**Table 1** shows the different demographic categories and the distribution or percentage represented in each class. For example, 69% of the males in the study are members of Class 1 (Harman Target Lovers) compared with 56% of females. Class 1 has roughly equal representation from trained (70%) and untrained (65%) listeners. Class 2 (More Bass Is Better) has the fewest members overall and is represented by all categories except female; only 4% of females tested want more bass than the Harman target provides. Class 3 (Less Bass Is Better) members are disproportionately represented by females (40% of females are in this class versus 13% of males) and listeners over the age of 50 (50%). Hearing loss may be a confounding factor here. More research is needed to better understand the role it plays in headphone sound quality preferences.

The main takeaway is that the Harman target is a good design target for headphones because it satisfies the tastes of a majority of listeners (64%) over a broad range of age groups, genders, and levels of listening experience. The two smaller classes of listeners who prefer headphones with more bass or less bass can be accommodated through a simple bass tone control on the headphone or via an app on the audio device. The bass adjustment would also help compensate for inconsistences in the quality of recordings that contain either too much or not enough bass and treble. A word of caution: the research suggests adding too much bass beyond the Harman target may alienate many listeners given that the "more bass is better" segment is a small segment with little female and older listener membership. Conversely, reducing the bass too much may alienate trained and experienced listeners who are underrepresented in the "less bass is better" segment.

## Predicting Listener's Headphone Sound Quality Preferences

Conducting controlled headphone listening tests is a challenging, time-consuming, and expensive proposition. An alternative solution is to model and predict listeners' headphone preference ratings using objective measurements that are relatively faster, more reproducible, and cost effective.

The 31 headphones from the same study discussed by Olive et al. (2018a) were sorted into 4 categories of sound quality based on listeners' preference scores: excellent (90-100% preference rating), good (65-76%), fair (42-54%), and poor

**Table 1.** *Distribution of listeners within each category.*

| Distribution of Listener Categories Within Each Class (in %) Based on Preferred Headphone Sound Profile | | | |
|---|---|---|---|
| **Category** | **Class 1: Harman Target Lovers** | **Class 2: More Bass Is Better** | **Class 3: Less Bass Is Better** |
| Males | 0.69 | 0.18 | 0.13 |
| Females | 0.56 | 0.04 | 0.40 |
| Trained | 0.70 | 0.30 | 0.00 |
| Untrained | 0.65 | 0.10 | 0.25 |
| **Age (years)** | | | |
| 20s | 0.69 | 0.17 | 0.15 |
| 30s | 0.74 | 0.13 | 0.13 |
| 40s | 0.67 | 0.10 | 0.24 |
| 50+ | 0.30 | 0.20 | 0.50 |

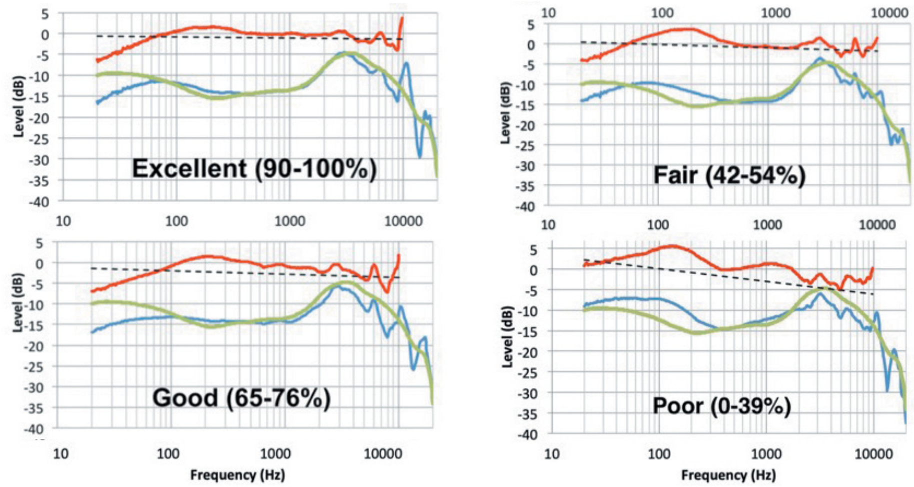
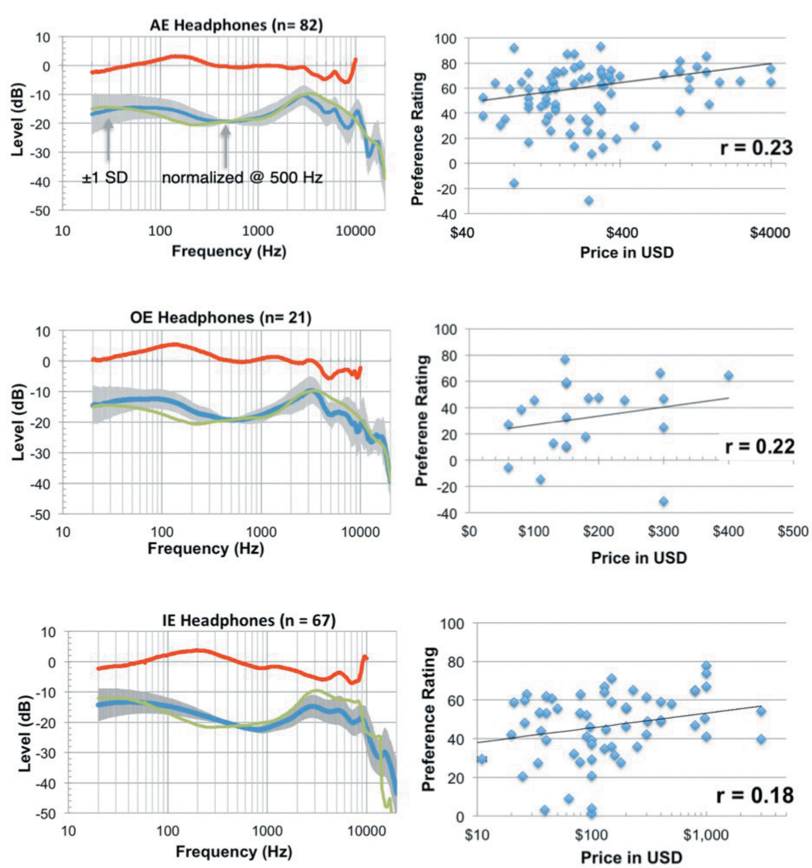



**Figure 3.** *The average frequency response (**blue**) and error curves (**red**) for 31 around the ear (AE) headphones assigned to four categories of sound quality based on their preference rating given in controlled listening tests. The Harman target curve used to calculate the error curve (**green**). **Dotted curve**, regression line that best fits the error curve (Olive et al., 2018a).*

**Figure 4.** ***Left:*** *average frequency response (**blue**), standard deviation (**gray area**), and error response curve (**red**) for AE, on the ear (OE), and in the ear (IE) headphone with the Harman target curve (**green**).* ***Right:*** *predicted preference ratings versus their retail price are plotted for the headphones with the best-fit regression line and correlation coefficient shown (Olive et al., 2018b).*



(0-39%). Frequency response measurements of the headphones were made using an ear simulator according to IEC 60318-1 (2009) equipped with a custom pinna that better simulates headphone leakage on humans.

In each category, the average frequency response for the headphones is plotted with the Harman target curve and the error curve that is the difference between the two (see **Figure 3**). **Figure 3**, *black dotted line*, is a regression line that best fits the error curve. The relationship between objective and subjective headphone measurements seems clear: the more the frequency response of a headphone deviates from the Harman target curve, the lower the listeners rated its sound quality.

A linear model was developed that predicts headphone preference ratings using two variables based on the standard deviation and the absolute slope of the error curve. The correlation between the predicted and measured ratings is 0.86, with an error of 6.7 ratings on a 100-point scale. A similar model was developed for IE headphones that produces slightly better ($r = 0.91$) predictions (Olive et al., 2016).

The two models were used to predict preference ratings for 158 headphones, including AE, OE, and IE types (Olive et al., 2018b). **Figure 4**, *left*, shows the average magnitude response, standard deviation, Harman target, and error response curve for each headphone type. **Figure 4**, *right*, plots the retail price versus the predicted preference rating for each headphone model tested. On average, the AE headphones come closest to the Harman target and produce the highest preference ratings, the OE headphones are the worst, and the IE headphones fall in between. The retail price of a headphone is not a good indicator of its sound quality based on the relatively low correlation values shown here.

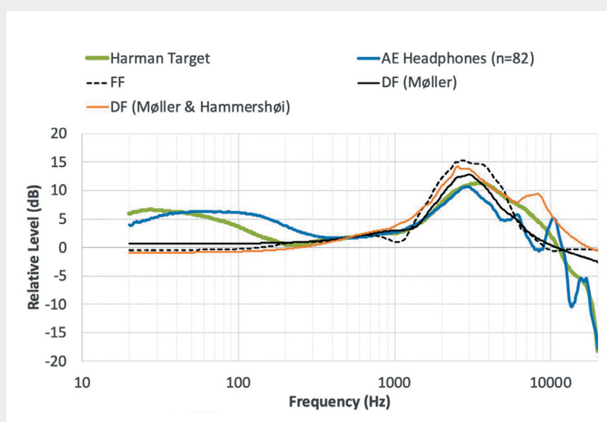
These findings are generally in agreement with those reported by Breebaart (2017). The two studies together provide evidence that headphone designers are aiming at a target curve that is closer to the Harman target than the DF or FF target curves recommended by the current headphone standards. **Figure 5** shows the average response of the 82 AE headphones in **Figure 4** compared with the Harman, DF, and FF target curves (Møller et al., 1995). Although the DF and FF targets specify a flat response below 200 Hz, the average AE headphone and Harman targets have 5-6 dB more bass, which better approximates the preferred in-room response of a full-range loudspeaker calibrated in a typical listening room.

## Conclusions

Our understanding of the perception and measurement of headphone sound quality has not kept pace with consumer demand and expectations. Two independent studies measured over 400 headphones and came to similar conclusions: there is little correlation between the price of a headphone and its frequency response, the single best indicator of its sound quality. Most professional and consumer headphone designs today do not comply with the FF and DF targets recommended by current headphone standards, which warns "the objective methods whose results bear good relation to those from subjective assessments are under research stage" (see IEC 60268-7, 2010, Section 8.6.1). The research stage is largely completed, the results are in, and the headphone standards need to be updated.

Listeners largely agree on what makes a headphone sound good. For stereo reproduction, the preferred headphone target approximates the in-room response of an accurate loudspeaker calibrated in a semireflective room. This makes perfect sense because stereo recordings are intended to sound best through accurate loudspeakers in semireflective rooms. What makes a headphone sound good is the same as what makes a loudspeaker sound good.

**Figure 5.** *Proposed headphone target curves normalized at 500 Hz: the Harman AE headphone target (**green**), two diffuse-field (DF; **orange** and **black**) and free-field (FF; **dashed**) targets (Møller et al., 1995), and the average frequency response (**blue**) of 82 different models of AE headphones (Olive et al., 2018b).*



The Harman target curve is one example that is preferred by a majority (64%) of listeners from a broad range of ages, listening experiences, and genders. Slight adjustments in the bass and treble levels may be necessary to compensate for variance in the quality of recordings and to satisfy individual tastes. The Less Bass Is Better class (21% of listeners) includes a disproportionate percentage of females and older listeners and none of the trained listeners. The More Bass Is Better class is skewed toward males versus females by a factor of 4 to 1. There is no evidence that sound quality preferences are geographically influenced. Recognition of good sound reproduction seems to be universal.

Objective measurements of the headphones using standard ear simulators can predict how good they sound. The further the frequency response a headphone deviates from the Harman target response, the lower its perceived sound quality will be. A simple linear model based on these deviations can predict how listeners would rate it in controlled listening tests.

The reaction from the headphone industry to this new research has been largely positive. There is evidence that the Harman target curve is widely influencing the design, testing, and review of many headphones from multiple manufacturers, providing a much needed new reference or benchmark for testing and evaluating headphones. Several headphone review sites provide frequency response measurements of headphones showing the extent to which they comply with the Harman target (Vafaei, 2018; Audio Science Review, 2020); in cases where they fall short, corrective equalizations are often provided.

As expected, there are also critics whose headphone tastes in sound may not agree with the research. The Harman target is intended as a guideline and is not the last word on what makes a headphone sound good. One legitimate criticism is the limited number of headphones, programs, female listeners tested, and questions raised about the confluence of variables like hearing loss and its effect on headphone preference. Future studies will hopefully address this. Finally, I hope that this article encourages others to continue the research and improve our knowledge of the perception and measurement of headphone sound quality. Although listeners largely agree on what makes a headphone sound good, there are still many unanswered questions and more to learn.

## References


American National Standards Institute/Consumer Technology Association (ANSI/CTA) Standard (2015). *Standard Method of Measurement for In-Home Loudspeakers*. Report ANSI/CTA-2034-A, ANSI/CTA, Arlington, VA.

Audio Science Review (2021). *Headphone Review and Discussions*. Available at https://tinyurl.com/cvbwmc2v.

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA.

Breebaart, J. (2017). No correlation between headphone frequency response and retail price. *The Journal of the Acoustical Society of America* 141(6), EL526-EL530. https://doi.org/10.1121/1.4984044.

Fleischmann, A., Silzle, F., and Plogsties, J. (2012). Identification and evaluation of target curves for headphones. *Proceedings of the 133rd Audio Engineering Society Convention*, San Francisco, CA, October 26-29, 2012. Available at https://www.aes.org/e-lib/browse.cfm?elib=16482.

International Electrotechnical Commission (IEC) (2009). *Electroacoustics - Simulators of Human Head and Ear - Part 1: Ear Simulator for the Measurement of Supra-Aural and Circumaural Earphones*. Report IEC 60318-1, IEC, Geneva, Switzerland.

International Electrotechnical Commission (IEC) (2010). *Sound System Equipment - Part 7: Headphones and Earphones*. Report IEC 60268-7, IEC, Geneva, Switzerland.

International Telecommunication Union Radiocommunication Assembly, (ITU-R) (1990). *Determination of the Electroacoustical Properties of Studio Monitor Headphones*. Report ITU BS.708, International Telecommunication Union, Geneva, Switzerland.

Lorho, G. (2009). Subjective evaluation of headphone target frequency responses. *Proceedings of the 126th Audio Engineering Society Convention*, Munich, Germany, May 7-10, 2009. https://www.aes.org/e-lib/browse.cfm?elib=14966.

Møller, H., Jensen, C. B., Hammershoi, D., and Sørensen, M. F. (1995). Design criteria for headphones. *Journal of the Audio Engineering Society* 43, 218-232. https://www.aes.org/e-lib/browse.cfm?elib=10274.

Olive, S. E. (2004). A multiple regression model for predicting loudspeaker preference using objective measurements: Part 2 - Development of the Model. *Proceedings of 117th Audio Engineering Society Convention*, San Francisco, CA, October 28-31, 2004. https://www.aes.org/e-lib/browse.cfm?elib=12847.

Olive, S. E., and Welti, T. (2015). Factors that influence listeners' preferred bass and treble balance in headphones. *Proceedings of the 139th Audio Engineering Society Convention*, New York, NY. https://www.aes.org/e-lib/online/browse.cfm?elib=17940.

Olive, S. E., Welti, T., and Khonsaripour, O. (2016). A statistical model that predicts listeners' preference ratings of in-ear headphones: Part 2 - Development and validation of the model. *Proceedings of 143rd Audio Engineering Society Convention*, New York, NY, October 29 to November 1, 2015. https://www.aes.org/e-lib/browse.cfm?elib=19436.

Olive, S. E., Welti, T., and Khonsaripour, O. (2018a). A statistical model that predicts listeners' preference ratings of around-ear and on-ear headphones. *Proceedings of the 144th Audio Engineering Society Convention*, Italy, May 23-26, 2018. https://www.aes.org/e-lib/browse.cfm?elib=19436.

Olive, S. E., Welti, T., and Khonsaripour, O. (2018b). A survey and analysis of consumer and professional headphones based on their objective and subjective performances. *Proceedings of the 145th Audio Engineering Society Convention*, New York, NY. https://www.aes.org/e-lib/browse.cfm?elib=19774.

Olive, S. E., Welti, T., and McMullin, E. (2013a). Listener preference for different headphone target response curves. *Proceedings of the 134th Audio Engineering Society Convention*, Rome, Italy, May 4-7, 2013. https://www.aes.org/e-lib/browse.cfm?elib=16768.

Olive, S. E., Welti, T., and McMullin, E. (2013b). A virtual headphone listening test methodology. *Proceedings of the Audio Engineering Society 51st International Conference: Loudspeakers and Headphones*, Helsinki, Finland, August 22-24, 2013. https://www.aes.org/e-lib/browse.cfm?elib=16874.

Olive, S. E., Welti, T., and McMullin, E. (2014). The influence of listeners' experience, age, and culture on headphone sound quality preferences. *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, October 9-12, 2014. https://www.aes.org/e-lib/online/browse.cfm?elib=17500.

PRNewsWire (2021), *Global Wireless Headphones Market to Reach $45.7 Billion by 2026*. Available at https://tinyurl.com/234dhxs4. Accessed November 2, 2021.

Sank, J. R. (1980). Improved real-ear tests for stereophones. *Journal of the Audio Engineering Society* 28, 206-218. https://www.aes.org/e-lib/browse.cfm?elib=3994.

Temme, S., Olive, S. E., Tatarunis, S., Welti, T., and McMullin, E. (2014). The correlation between distortion audibility and listener preference in headphones. *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA. October 9-12, 2014. https://www.aes.org/e-lib/online/browse.cfm?elib=17441.

Theile, G. (1986). On the standardization of the frequency response of high-quality studio headphones. *Journal of the Audio Engineering Society*, 34, pp. 956-969. https://www.aes.org/e-lib/browse.cfm?elib=5233.

Toole, F. E. (1985). Subjective measurements of loudspeaker sound quality and listener preferences. *Journal of the Audio Engineering Society* 33, 2-31. https://www.aes.org/e-lib/online/browse.cfm?elib=4465.

Toole, F. E. (1986). Loudspeaker measurements and their relationship to listener preferences. *Journal of the Audio Engineering Society* 34, Part1, 227-235; Part. 2, . 323-348. https://www.aes.org/e-lib/browse.cfm?elib=5276.

Toole, F. E. (2018). *Sound Reproduction, the Acoustics and Psychoacoustics of Loudspeakers and Rooms*, 3rd ed. Routledge, New York, NY.

Vafaei, S. (2018). Raw frequency response. *Rtings.com*. Available at https://bit.ly/3oBUnKG. Accessed November 2, 2021.

## About the Author

**Sean E. Olive**
sean.olive@harman.com

*Harman International*
*8500 Balboa Boulevard*
*Northridge, California 91329, USA*

**Sean E. Olive** studied music at the University of Toronto, Toronto, ON, Canada, and sound recording at McGill University, Montreal, QC, Canada, where he received master's and PhD degrees. He is a senior fellow in acoustic research at Harman International, Northridge, California, focused on research in sound quality of reproduced sound. He has published over 50 research papers for which he received Fellow of the Audio Engineering Society (AES), two AES publications awards, and the ALMA Titanium Award. Sean is the past president of AES.

# Recent Acoustical Society of America Awards and Prizes

*Acoustics Today* is pleased to present the names of the recipients of the various awards and prizes given out by the Acoustical Society of America. After the recipients are approved by the Executive Council of the Society at each semiannual meeting, their names are published in the next issue of *Acoustics Today*.

Congratulations to the following recipients of Acoustical Society of America medals, awards, prizes, and fellowships, who will be formally be recognized at the Spring 2022 Plenary Session. For more information on the accolades, please see:

- https://acousticalsociety.org/acoustical-society-of-america-awards
- http://acousticalsociety.org/prizes
- https://acousticalsociety.org/fellowships-and-scholarships

***Gold Medal***

**Michael J. Buckingham**
(University of California, San Diego, La Jolla)
for theoretical and experimental contributions to ocean acoustics and for service to the Society

***Helmholtz-Rayleigh Interdisciplinary Silver Medal in Physical Acoustics and Engineering Acoustics***

**George Augspurger**
(Perception, Inc., Los Angeles, California)
for contributions to the design of recording studios, performance venues, and loudspeakers and for decades of patent reviews

***R. Bruce Lindsay Award***

**Meaghan O'Reilly**
(Sunnybrook Research Institute, Toronto, ON, Canada)
for contributions to biomedical ultrasound applications in the central nervous system

Congratulations also to the following members who were elected Fellows in the Acoustical Society of America in Spring 2022.

- **Julien Bonnel**
  (Woods Hole Oceanographic Institution, Woods Hole, Massachusetts) for advances in time-frequency analysis of underwater sound
- **Lori Holt**
  (Carnegie Mellon University, Pittsburgh, Pennsylvania) for understanding neural processing and perception of complex auditory phenomena over a life span
- **Rochelle Newman**
  (University of Maryland, College Park) for the understanding of speech perception and language development in challenging listening conditions
- **Andi Petculescu**
  (University of Louisiana at Lafayette) for exploring the acoustics of extraterrestrial environments
- **Erica Ryherd**
  (The Durham School of Architectural Engineering and Construction, University of Nebraska-Lincoln) for advancements to acoustics in the health care industry



https://doi.org/10.1121/AT.2022.18.1.68

# Ask an Acoustician: Arthur N. Popper

*Arthur N. Popper*
*and Micheal L. Dent*

**Figure 1.** *Art Popper with his grandkids (**left to right**) Emma, Sophie, and Ethan.*

## Meet Arthur N. Popper

In this, the last "Ask an Acoustician" essay, we feature Arthur N. (Art) Popper. I thought it was fitting that Art, as editor of *Acoustics Today* (*AT*), write the final article in this series of interviews. Art received his BA from New York University, The Bronx, New York, and his PhD from the City University of New York. He had faculty positions at the University of Hawai'i, Honolulu, and Georgetown University, Washington, DC, before moving to the University of Maryland, College Park, in 1987 (where I got to know him when I was a graduate student). Art is a Fellow of the Acoustical Society of America (ASA) and has served on many committees of the ASA over the years. Currently, in addition to serving as the *AT* editor, he is the coordinating editor for animal bioacoustics for *The Journal of the Acoustical Society of America*. I will let Art tell us the rest.

## A Conversation with Arthur N. Popper, in His Own Words

### *Tell us about your work.*

I am "semiretired" in that I no longer have a regular appointment at the University of Maryland where I worked for over 25 years. But I continue to be very active professionally, doing research, writing, editing, and a bit of consulting. Perhaps the thing I enjoy most is editing *AT*. Indeed, *AT* takes up a good deal of my time since I not only work with authors to develop topics but also review and edit all articles and essays.

Of the time I spend on *AT*, the most time consuming and interesting is working with the authors to hone their contributions. By this, I mean that our goal for *AT* is to have scholarly content while communicating science and technology in ways that every member of the ASA can read and understand. The "problem" is that most of us are trained to write for peers and at very technical levels, and so communicating complex material to a broad audience is a challenge. Fortunately, most authors are responsive to my "pushing" them to communicate with our very broad ASA audience, but this may take anywhere from 3 to 12 iterations of a manuscript. I am pleased, however, that authors not only thank me for working with them but also often tell me that they learned a good deal about how to communicate their work to a broader audience, which might include a dean, a CEO, the public, or their grandparents.

I also continue editing a series of books, the Springer Handbook of Auditory Research (volume 74 is in press) (Fay and Popper, 2014), I have also been writing several scholarly papers each year, mostly related to my interests in the effects of anthropogenic sound on fishes and other aquatic life (e.g., Popper and Hawkins, 2019), and I am part of several research projects on the potential effects of anthropogenic sound on fishes.

### *Describe your career path.*

My path is one of serendipity, which I discussed in Popper (2014). Opportunities arose and I followed their trail. Indeed, I keep being amazed that if I'd made a different decision at various points, my career might be very different.

Just as a few examples, I got started doing research on fish because, on my way to school one day (New York University), I had a few minutes to spare before my bus would come and so I stepped into a new pet shop. I found a tank holding fish without eyes, Mexican blind cave fish. I got very curious about these fish and asked one of my professors, Douglas Webster (who later became a good friend), about them. He invited me to do research on hearing in his laboratory. This led to my working in the world-renowned Ichthyology


https://doi.org/10.1121/AT.2022.18.1.69

Department at the American Museum of Natural History (AMNH; New York) where, one day, one of the investigators happened to show me an otolith (a fish "ear bone"); little did I know then that otoliths would become an integral part of my future research. Ultimately, my time as an undergraduate working at the AMNH led me to the museum's Department of Animal Behavior where I met William N. Tavolga (see https://www.ahukini.net/tavolga). Later, Bill became my doctoral advisor, life-long mentor, and very close friend to my wife Helen and myself. And Mexican blind cave fish became the subjects of my doctoral research and my first two research publications!

More recently, I was called by a researcher for the US Congress and asked about the response of salmon to sound. She then asked me to review the literature on how sounds can be used to control fish movement. That path let me to other opportunities that ultimately resulted in a substantial shift in my research from doing basic science to applying the work I'd done for the first half of my career to real-life problems.

I could go on, but my point is that I have been truly fortunate to not only have a great career but to also be able to take advantage of opportunities that arose unexpectedly.

### *What is a typical day for you?*

Unlike when I was "working," my day starts with reading *The New York Times*. A positive thing about semiretirement is that I no longer must attend department meetings, sit on tenure committees, or seek grant funding. (Although I do miss teaching and working with students.) Most of my day is devoted to writing and editing, meeting with colleagues on joint research and writing projects (via Zoom), or working with groups around the world on issues related to the effects of anthropogenic sound on the aquatic environment.

I do interrupt my day with various nonwork things. I am reasonably active in the community in which we live, and so I work with various community groups dealing with such things such as strategic planning and development of electric car charging stations. I also try to get to our local fitness center to work out or swim at least four days a week.

### *How do you feel when experiments/projects do not work out the way you expected them to?*

Part of doing science is that things don't always work out. Indeed, I try to teach students that one of the most important things they can learn while they are a student is that experiments often do not work the first, second, or even fifth time and that they must accept this and come up with ways to solve problems. So, I try to figure out why things may not have worked and ask whether I was asking the wrong question or if I'd tried to answer the question in the wrong way. And then I explore other approaches. I try not to give up but to be creative.

### *Do you feel like you have solved the work-life balance problem? Was it always this way?*

As a semiretiree, my focus is on family and work comes second. So, my work-life balance now is primarily life-work. When I was working, this was harder, but I think I handled things pretty well and that my priority was always Helen and our girls, Michelle and Melissa. Of course, I am truly fortunate that Helen has always been extraordinarily supportive of my work, and, as a biologist herself, she has some appreciation of the work I do[1].

### *What makes you a good acoustician?*

I want to rephrase this question: What makes me a good scholar? I say this because I don't classify myself as an acoustician per se but rather as a biologist (or neuroscientist or neuroethologist or ichthyologist, depending on who I'm talking with) interested in how biological systems deal with sound. Then, the answer to the question is curiosity, looking at issues with an open mind, and enjoying being a problem solver.

But the other part of the answer is that my work has benefitted immensely because I have been fortunate to have a great network of colleagues (many who have become close friends) with whom I've collaborated for much of my career. A critical part of these collaborations is that each participant brings a different skill set and way of thinking to our work, and this strongly enhances what we are doing. I've actually written about my collaborations in a recent paper (Popper, 2020).

My point is one of the things that has made me good at what I do is being able to collaborate and share ideas. This is not only productive but is perhaps one of the most enjoyable parts of my career.

### *How do you handle rejection?*

I cannot recall how I dealt with rejection early in my career, but at this stage in my life, I expect it and try not to get too upset (although I don't like it and never have).

*1 Full disclosure: Helen is the copy editor for* AT.

In some cases, I say fine and just go on. In other cases, I will "stew" on the rejection for a long time, especially if there is nothing I can do about it. In other cases, especially with papers, I try to evaluate why the paper was rejected and make corrections.

So, the answer to the question is that I handle rejection in different ways depending on the circumstances. The only rejection I cannot handle is when one of our grandkids says that she or he would rather pal around with a friend than spend time with grandpa (**Figure 1**)!

### *What are you proudest of in your career?*

I think it is the way my work has evolved. I started out asking questions about what fish hear, and over the years, the questions I asked and the research approaches I have used have evolved: first to asking questions about the evolution of hearing (a topic that still holds great interest to me) and most recently to being heavily involved, on an international scale, in setting guidelines and criteria for the potential effects of anthropogenic sound on aquatic animals. The point is that I am proud that the questions I have asked and the approaches I have taken to answer the questions were never static. Indeed, I think that an evolution in research questions and approaches is critical for any good scholar.

At the same time, there are a couple of specific things I've done that I think have been of considerable value. My discovery of the organization of sensory cells in the ear of fishes (which was simultaneously discovered by several others in Europe) has had a significant impact on understanding fish hearing (Popper, 1976). And my being able to cochair an international group developing criteria and guidelines for the effects of sound on fishes has become an informal standard around the world. Knowing that our group has had a real impact is quite a nice feeling.

### *What is the biggest mistake you've ever made?*

I'm sure I've made mistakes in both science and life. I do think these were mainly where I made the wrong choice when I had options. However, I try never to go back and ask "what if" because I know that I cannot change where I'm going. For example, what if I'd chosen to take a postdoc with Arthur Myrberg at the University of Miami, Coral Gables, Florida, rather than take the job offered me at the University of Hawai'i (my first job)? I'll never know but I am really pleased where this decision took me.

### *What advice do you have for budding acousticians?*

Find great mentors at every stage of your career and be a great mentor to others. Develop great networks. Value and enjoy collaboration. Read what other authors of "Ask an Acoustician" essays have said about this and figure out what will work best for you.

### *Have you ever experienced imposter syndrome? How did you deal with that if so?*

In hindsight, I suspect so. Mostly in terms of collaboration, wondering whether I'm contributing equally to the collaborations. However, over the years, I realized that collaborations are really a sharing of ideas and skills and my collaborators keep "coming back for more," so I feel comfortable in saying that, despite how I might feel, I am giving as much as I get to our shared endeavors.

### *What do you want to accomplish within the next 10 years or before retirement?*

Since I am already semiretired, I am now contemplating a second retirement where I actually learn to not work and find fun things to do. I have no idea what those things will be, but I will be entering full retirement on December 31, 2024. Of course, everyone who knows me, from colleagues to Helen to grandkids, laughs at the idea that I will "really" retire!


#### Bibliography

Fay, R. R., and Popper, A. N. (2014). A brief history of SHAR. In Popper, A. N., and Fay, R. R. (Eds.), *Perspectives on Auditory Research*. Springer, New York, NY, pp. 1-8.

Popper, A. N. (1976). Ultrastructure of the auditory regions in the inner ear of the lake whitefish. *Science* 192(4243), 1020-1023.

Popper, A. N. (2014). From cave fish to pile driving: A tail of fish bioacoustics. In Popper, A. N., and Fay, R. R. (Eds.), *Perspectives on Auditory Research*, Springer, New York, NY, pp. 467-492.

Popper, A. N. (2020). Colleagues as friends. *ICES Journal of Marine Science* 77(6), 2033-2042.

Popper, A. N., and Hawkins, A. D. (2019). An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes. *Journal of Fish Biology* 94, 692-713.



#### Contact Information

**Arthur N. Popper** apopper@umd.edu

*Department of Biology*
*University of Maryland, College Park*
*College Park, Maryland 20742, USA*

**Micheal L. Dent** mdent@buffalo.edu

*Department of Psychology*
*University at Buffalo*
*State University of New York (SUNY)*
*B76 Park Hall, Buffalo, New York 14260, USA*

# Identity Struggles of a Black STEM Academic

*Tyrone Porter*

Sylvester James "Jim" Gates, Jr. was the keynote speaker at the Acoustical Society of America (ASA) Acoustics in Focus meeting in May 2021, and he gave a riveting account of his academic and professional life as a Black STEM scholar (see https://bit.ly/3CWGOti). Listening to the keynote speech, I thought about my own experiences as a Black STEM scholar and pondered on the identity struggles that I have faced through the years.

Like most people, I have multiple identities. I am an American, a Detroit native, a husband, a father, an engineer, and a sports enthusiast. I am also Black, and there are many people who have been conditioned to associate that identity with criminality or a lack of intelligence. These prejudices have established a racial hierarchy that exists throughout America and has contributed to the underrepresentation of Blacks and other ethnic minorities in STEM (Science, Technology, Engineering, and Mathematics). Based on these biases, bigots have questioned whether I deserved various opportunities academically or professionally. I became an activist to combat prejudice and systemic racism in society generally and in STEM specifically. While I have persevered through the years, there were pivotal moments in my past that caused me to question my academic and scientific identities and whether America was capable of change. In this essay, I share on these pivotal moments and how they have shaped my identities and my commitment to the fight for racial equity and inclusivity in STEM and in acoustics.

After completing my undergraduate degree in electrical engineering at Prairie View A&M University (see https://www.pvamu.edu/), a Historically Black College/University (HBCU) in Prairie View, Texas, I started my doctoral studies in bioengineering at the University of Washington (UW; Seattle). My doctoral adviser was Larry Crum, a former ASA president and a recipient of an ASA Gold Medal. This was my first extended experience at a Predominantly White Institution (PWI) and I was uncomfortable immediately.

One of the greatest attributes of an HBCU is the efforts made by virtually all staff, students, and faculty to connect with each other and create a supportive community. I didn't see the same level of effort toward community building at the UW and found it difficult to connect with others.

Indeed, my time in graduate school coincided with the anti-affirmative action movement (for background, see https://en.wikipedia.org/wiki/Affirmative_action), which had gained momentum after the passing of Proposition 209 (Prop 209) in California as an amendment to the state constitution. Prop 209 prohibited the use of race, gender, ethnicity, or national origin in hiring, contracting, or admission decisions in state institutions, effectively ending affirmative action initiatives in California universities and colleges.

Inspired by Prop 209, citizens of Washington State were able to get an equally damaging bill on the ballot, known as Initiative 200 (I-200). There were numerous town hall meetings to discuss I-200 and its potential impact, and proponents argued that Black students like myself were admitted into the UW solely based on race. Although proponents never presented any evidence to support their narrative, the claims perpetuated the idea that students of color like myself were inferior academically. Persons of color have been fighting to change this perception for years, whereas those who have benefited from exclusionary practices have fought to maintain it.

At the UW, I usually was the only Black student in my classes, and I worried constantly that the other students or the professors believed I was admitted to the UW only because I was Black. This created undue stress that initially negatively affected my academic performance. I began to question whether I deserved to be at the UW and whether I had the capacity to succeed in the graduate program and assume the identity of a bioengineer. In fact, there were only two other Black students in the bioengineering graduate program at the time so I had trouble



https://doi.org/10.1121/AT.2022.18.1.72

finding other students with whom to discuss my feelings. Moreover, adjusting to the new environment was not a topic of discussion within the acoustics research group, so I had to find a solution on my own. I have heard similar stories from peers, some of whom decided to transfer to another university or terminate their degree early.

The fact is that studies have shown that one of the greatest barriers to persons of color thriving in STEM academic programs is establishing a scientific identity and connecting with others in the discipline. I was able to connect with other STEM students of color in the Minority Science and Engineering Program (MSEP) and the Graduate Opportunities and Minority Achievement Program (see https://grad.uw.edu/gomap-is-now-gsee/), and my academic performance improved over time. Additionally, I began to visit the Ethnic Cultural Center (ECC; see http://depts.washington.edu/ecc/) frequently because this was a safe space for students from underrepresented groups to discuss the challenges of simply existing at a PWI. We created a supportive community, which enabled most to excel instead of survive at the UW, and we worked collectively to push the university to expand access and support for underrepresented students.

Additionally, I cofounded the Minority Think Tank (MTT) with other students from the ECC. We organized seminars and launched programs that were designed to combat exclusivity and counter the false narrative presented by anti-affirmative action advocates. I assumed the identity of an activist during this period, and my involvement in these activities served as a stress release. However, it was difficult to balance my identities as a doctoral student and as an activist.

Fortunately, Larry Crum and the other bioengineering faculty and staff were understanding and extremely supportive. I worked on a collaborative project studying the impact of high-intensity focused ultrasound combined with pH-sensitive polymers on the permeability of cell membranes. There were many days where I split my time equally between the laboratory and meetings at the ECC. Throughout this period, Larry and other faculty in the department fueled my passion for science and occasionally attended events that my peers in MTT and I organized. I successfully managed my dual identities with the support of Larry and the other faculty, completing my doctoral studies in 2003 while also creating new programs to recruit and support marginalized students that were adopted by the university.

When I started my faculty position at Boston University (BU; Boston, Massachusetts) I found myself once again in the all-too-familiar position of being "the only or one of the few." I continued my commitment to increasing diversity, equity, and inclusion (DEI) in STEM programs but found it more difficult to balance my efforts with my obligations as a faculty member. I had more time available for extracurricular activities as a graduate student but starting a faculty career was *significantly* more demanding, and in a way, the expectations on me, as the only Black faculty member, were greater than for White colleagues. In addition to building a research program and teaching core undergraduate engineering courses, I agreed to serve as faculty advisor for the BU chapter of the National Society of Black Engineers and the Black Student Union.

By working with this group, I discovered that the students craved contact time with faculty of color so I attended meetings of the organizations as often as possible to increase interaction with the students. Additionally, I regularly met with the presidents of the chapters to discuss monthly events, membership recruitment, and leadership skills. By working cooperatively with the presidents, the chapters were operated more effectively, which led to an increase in membership enrollment and participation.

Although working with the students directly was gratifying, it also was draining and took time away from building my research program. This was a major risk given that scholarly output and teaching are more valued in promotion and tenure decisions than efforts to increase DEI in higher education.

The fact is that DEI efforts are commonly referred to as "invisible work" because there is no clear way to capture the impact of these efforts in a measurable way in faculty annual reports or promotion and tenure dossiers. Moreover, faculty who write evaluation letters for promotion and tenure cases are rarely if ever asked to comment on the efforts of the candidate to address underrepresentation of marginalized groups in their respective discipline. Knowing that my DEI efforts could go unnoticed created tremendous stress and anxiety because they were a reflection of my identity and core values. I felt like the system was forcing me to decide between my racial and scientific identities. Fortunately, the engineering senior leadership at BU valued my DEI efforts and honored me with a Faculty Service Award. I did successfully navigate the system to promotion with tenure while being true to both identities. But the academic system needs to evolve and embrace scholars who desire to expand access to higher education in addition to generating new knowledge. A system that values DEI efforts in addition to a scholar's h-index will attract more scholars of color to pursue careers as STEM faculty who can then serve as role models for students of color and help them establish STEM identities.

Moving forward, I plan to embrace both identities and infuse DEI in all aspects of my profession. Recently, I changed the name of my research group to the Diverse Engineering Applications Laboratory (D.E.A.L.), signaling that we value diversity in personal background as well as in scholarly pursuits. The guiding principle for the laboratory is "Diversity drives innovation, creativity, and personal growth." I have also begun to pen more essays such as this one sharing my experiences balancing my identities as a scholar and as an activist. I hope that my efforts contribute to transforming higher education so that DEI work is no longer "invisible."

### Contact Information

**Tyrone Porter** tmp6@utexas.edu
*Department of Biomedical Engineering*
*University of Texas at Austin*
*107 W. Dean Keeton Street*
*Austin, Texas 78715, USA*

# Voces de los acústicos hispanohablantes en América Latina[1]

*Zachery O. L'Italien, Fernando del Solar Dorrego, Ana M. Jaramillo, y Mariana Botero*

La Sociedad Americana de Acústica (Acoustical Society of America; ASA) implementó su primer plan estratégico en 2015 (ver el plan estratégico actual en bit.ly/ASAStrategicPlan2022), definiendo sus objetivos a futuro, uno de los cuales es el "incremento de la participación de los miembros y la diversidad." Ana Jaramillo, miembro de ASA, y originaria de Medellín, Colombia, ha participado regularmente de las conferencias de la ASA desde 2007, y en ellas notó la baja participación de hispanos en la Sociedad. En un esfuerzo para abordar esta baja representación de acústicos hispanos, un pequeño grupo de miembros hispanohablantes de la ASA (Spanish-Speaking Acousticians, SSA) se reunió en el congreso de la ASA en 2018 en la ciudad de Minneapolis, en donde intercambiaron ideas sobre cómo servir e interconectar mejor a la comunidad acústica hispana de la Sociedad. Este encuentro resultó en la creación del capítulo regional "Spanish-Speaking Acousticians in the Americas," que fue aprobado en diciembre de 2019 (ver ssaasa.org). Sus miembros fundadores fueron: Fernando del Solar Dorrego (actual Presidente), Ana Jaramillo (actual Representante ante la ASA), Sandy Guzman y Alex Padilla. El comité actual se puede ver en la **Figura 1**. Hoy en día, el capítulo cuenta con más de 180 miembros, distribuidos principalmente en las Américas, pero con miembros también en Europa y Asia.

**Figura 1.** ***Arriba, izquierda a derecha:*** *Ana Jaramillo, representante ante ASA nacional; Fernando del Solar, presidente; Mariana Botero, secretario; Zachery L'Italien, tesorero.* ***Abajo, izquierda a derecha:*** *Juan Francisco Mayorga, miembro vocal; Andrés Millán, miembro vocal; Felipe Raimann, vice-presidente.*

Si bien el capítulo está logrando más participación y exposición, nuestra base de datos muestra que menos del 20% de nuestros miembros son miembros de la ASA (miembro, asociado, estudiante u otros), e incluso una menor proporción participa activamente de las actividades de la ASA. El Directorio de Membresía de la ASA muestra que solo 105 miembros de la ASA residen en países de habla hispana (un tercio en España y dos tercios en América Latina). De estos miembros, sólo cinco son miembros honoríficos (Fellows) y 17 son miembros estudiantes. Existen, obviamente, miembros de origen hispano que residen en otros países, como Estados Unidos. Un estudio demográfico de la ASA de 2019 muestra que sólo el 4% de sus miembros son hispanos. Sin embargo, a lo largo de los años, se han publicado muchos artículos en el *Journal of the Acoustical Society of America* (*JASA*) por parte de acústicos de Latinoamérica, muchos de ellos sin ser miembros de nuestro capítulo, como también se han hecho presentaciones en conferencias, y otras formas de participación en las actividades de la Sociedad. Es importante mencionar que la ASA ofrece un descuento en su membresía para la categoría asociado electrónico (*electronic associate member*) para miembros que residen en países en vía de desarrollo, tales como los de la región de América Latina (ver https://acousticalsociety.org/asa-membership/).

Actualmente, el 65% de los miembros de Spanish-Speaking Acousticians (SSA) estudian o trabajan en el área de acústica arquitectónica (AA), el 10% en acústica musical, y menos del 1% en áreas como acústica psicológica y fisiológica, o acústica computacional. Con el objetivo de minimizar esta desproporción, nuestro capítulo tiene como reto aumentar la diversidad técnica de la membresía, invitando investigadores,

*1 The English-language version of this article appears in the winter 2021 issue of* Acoustics Today *(see https://doi.org/10.1121/AT.2021.17.4.75). We include it in Spanish as part of the interest of the Acoustical Society of America to reach out to a broader international community. Translation is by the authors.*


https://doi.org/10.1121/AT.2022.18.1.75

profesionales y estudiantes de todas las áreas técnicas para que participen en las actividades del grupo.

Dentro de la ASA, han existido esfuerzos enfocados en aumentar la participación de acústicos latinoamericanos, no relacionados con SSA. En 2017, Ernesto Accolti publicó una serie de entrevistas a acústicos latinoamericanos destacados en la revista *Acoustics Today* (ver bit.ly/AcousticianInterviews). En 2020, varios artículos de *Acoustics Today* fueron traducidos al español y publicados en la revista *Ingenierías* como parte del Año Internacional del Sonido (ver bit.ly/SpanishArticles). Ambas iniciativas ayudaron a incrementar el acceso de los hispanohablantes a la ASA y les dieron visibilidad a acústicos latinoamericanos destacados en diversas áreas. Tenemos la esperanza de que las iniciativas promovidas por nuestro capítulo continúen esta tendencia de aumentar la participación de los acústicos hispanohablantes en la ASA, al mismo tiempo que se le da visibilidad a los retos que afrontan los profesionales en América Latina y promueven el desarrollo de la acústica en la región.

En los próximos meses se espera ver un aumento en la membresía de la ASA por parte de nuestros miembros, así como su participación activa. También se espera continuar la diversificación de las áreas técnicas en el Capítulo. En 2021 comenzamos una colaboración con el Comité para la mejora de la Diversidad Racial e Inclusividad (CIRDI; ver bit.ly/CIRDIASA), otros grupos de la ASA, y algunas entidades externas a la ASA. Como meta, esperamos ver más oportunidades para que los acústicos hispanos sean parte activa de la Sociedad, tanto virtualmente como en persona, ya que sabemos que la ASA puede convertirse en un recurso invaluable para los acústicos de la región, y ellos mismos como una fuente de crecimiento para la ASA.

El Comité de SSA busca entender los problemas y limitaciones que enfrentan los acústicos en América Latina. En Junio de 2021, se envió una encuesta a todos los miembros del Capítulo, en la cual se pedía describir su perspectiva acerca de las oportunidades de desarrollo profesional y educación en sus países de origen, en términos de acceso a la educación formal, oportunidades laborales, y estabilidad económica. Las respuestas, en su mayoría provenientes de acústicos en el área de acústica arquitectónica, demuestran que la acústica no es una profesión o área de estudio suficientemente desarrollada en América Latina y que muchos acústicos en este campo enfrentan innumerables retos en su desarrollo profesional. Una queja común expresada por quienes respondieron a la encuesta es la falta de regulación por parte de los gobiernos e instituciones que generen unas bases adecuadas para promover la acústica como parte del diseño arquitectónico, la construcción, y el ruido ambiental. Sin embargo, aunque dichas regulaciones podrían generar más oportunidades para acústicos trabajando en el área de acústica arquitectónica y control de ruido, es igualmente importante el desarrollo industrial, científico y tecnológico, el cual podría incentivar a los acústicos a investigar, trabajar y estudiar en la región.

Uno de los encuestados, tras haber emigrado de su país de origen, comentó que "No hay oportunidades de trabajo en acústica, las pocas empresas tienen cupos llenos y los niveles de conocimiento actuales en campos de la acústica (acústica arquitectónica, relación acústica-estructura, ultrasonidos, acústica física y otros) son medio-bajos." Para las economías latinoamericanas es particularmente difícil establecer empresas con capacidad de dar empleo y compensación adecuada. Aquellos con el nivel educativo y la experiencia necesaria son comúnmente considerados "sobrecalificados", así que las posiciones de trabajo se llenan con empleados sin la educación y experiencia requerida.

A pesar de que la mayoría de los encuestados son acústicos arquitectónicos, también hay acústicos latinoamericanos en otras áreas que contribuyen regularmente a la ASA a través de las revistas *JASA*, *POMA* (*Proceedings of Meetings on Acoustics*) y otros medios. Sin embargo, se puede asumir que dichos acústicos enfrentan retos similares con relación al desarrollo profesional y laboral. Por ejemplo, en la entrevista para *Acoustics Today* de Gabriela Virginia Santiago, una acústica venezolana enfocada en la neurociencia cognitiva, ella comenta que "al menos en Venezuela la acústica es vista como un campo de estudio completamente nuevo y mucha gente no conoce mucho de acústica. Sin embargo, creo que el interés por la acústica está creciendo en las generaciones más jóvenes de Sudamérica" (ver bit.ly/SantiagoAccolti-es). Jorge Arenas ha notado el mismo interés y comenta que la academia fue esencial en establecer criterios acústicos de control en la legislación chilena, y esto se dio como resultado de una colaboración global (ver bit.ly/ArenasAccolti-es). Otros como ellos, trabajan duro para darle visibilidad a la acústica entre la población general y son pioneros en fijar las bases para el futuro del medio.

Con la industria de la acústica creciendo en Latinoamérica, el comité SSA cree que la colaboración entre colegas tiene un gran potencial para ayudar a avanzar el campo, enfatizando la importancia de conectarse entre acústicos y facilitar un espacio para el crecimiento regional. "Hay países más avanzados que otros, dentro de la región veo un futuro prometedor ya que no se ha parado de crecer y con ello se requiere de mayor conocimiento y más aplicaciones de la acústica para resolver diferentes problemáticas de países en vías de desarrollo," comenta uno de los miembros encuestados de Ecuador.

Asimismo, una miembro y profesora de ingeniería civil de Uruguay ve un inmenso potencial de crecimiento, pero le preocupa que la prioridad de satisfacer necesidades básicas compromete el avance en áreas especializadas, como la acústica. Ana Jaramillo recuerda cuando, después de obtener su título de doctorado en los Estados Unidos, regresó a Colombia y encontró un desarrollo significativo en el ámbito de la acústica arquitectónica, demostrado por varias empresas de consultoría bien establecidas y un mejor intercambio de conocimiento entre universidades y profesionales, en varias ciudades del país. Esta construcción de relaciones profesionales y académicas, junto con el compartir conocimiento podría empezar a interconectar países en todo el continente, para combinar conocimiento y experiencia, especialmente ahora que el mundo está mejor preparado para comunicarse a través de plataformas virtuales.

También han existido esfuerzos para implementar cursos en programas de pregrado y postgrado de acústica en universidades latinoamericanas. Estos programas en acústica son ofrecidos más comúnmente en universidades de Argentina y Chile, como algunos miembros comentaron, pero también hay algunos más recientemente establecidos, como una maestría de la Universidad de San Buenaventura de Bogotá, Colombia. Otros países latinoamericanos ofrecen cursos de acústica como parte de un programa educativo diferente, donde el enfoque principal no es la acústica. Sin embargo, cursos y materias interesantes tienen gran potencial de fomentar el interés en el campo de la acústica. Por ejemplo, en el Instituto Tecnológico de Buenos Aires (ITBA), en Buenos Aires, Argentina, Fernando del Solar Dorrego acompaña a sus estudiantes a realizar una escucha crítica en el Teatro Colón, donde tienen una oportunidad muy llamativa de conectar teoría con experiencia.

Reconocemos que, a pesar del crecimiento reciente, todavía hay fallas, retos y oportunidades de mejora, como se ha mencionado en este artículo. Prevemos el crecimiento continuo del campo de la acústica en Latinoamérica y vemos nuestro capítulo como un gran paso para hacer que la ASA sea más accesible para acústicos hispanohablantes. Confiamos en que los posibles frutos de las crecientes contribuciones de acústicos de habla hispana sean sustanciales y resulten en un crecimiento aún mayor de la Sociedad y del campo de la acústica, llevando a su vez grandes oportunidades para nuestros colegas en Latinoamérica y la comunidad acústica en general.

El comité SSA está emocionado con la cantidad de apoyo y entusiasmo que hemos recibido de la comunidad de la ASA, y estamos trabajando arduamente para continuar organizando eventos, crecer nuestra membresía, colaborar con otros grupos y beneficiar a la comunidad Hispana. El comité SSA trabaja diligentemente para traer más voces hispanas a la ASA, generando eventos patrocinados (tanto en inglés como en español) con acústicos de alto renombre de todo el mundo. Pensando en el futuro del capítulo como parte de la ASA, esperamos expandir nuestras actividades y hacer de nuestro capítulo un recurso cada vez más útil, fomentando mayor acceso y oportunidades para que los acústicos latinoamericanos puedan participar en las conferencias nacionales, publicaciones, comités técnicos y crecimiento regional de la ASA. ¡El comité SSA espera con ansias el futuro de la acústica en Latinoamérica y en el futuro de la ASA!

En conclusión, queremos invitar a todos los miembros de habla hispana en la ASA, independientemente de su ubicación geográfica, a unirse a nuestro grupo. Queremos conectarnos con la mayor cantidad posible de colegas de todas las áreas técnicas de la ASA. Para unirse al capítulo SSA, por favor visita nuestra página en https://ssaasa.org/unete/.

### Contact Information


**Zachery O. L'Italien** zlitalien@mchinc.com

*McKay Conant Hoover, Inc.*
*5655 Lindero Canyon Road, Suite 325*
*Westlake Village, California 91362, USA*

**Fernando del Solar Dorrego**
fernando@patagonacoustics.com

*Patagon Acoustics*
*Juan Francisco Segui 3511 4th*
*1425 Buenos Aires, Argentina*

**Ana M. Jaramillo** ana@olsonsound.com

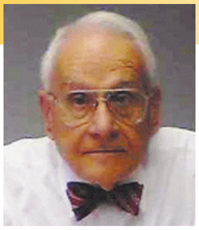
*Olson Sound Design, LLC*
*8717 Humboldt Avenue North*
*Brooklyn Park, Minnesota 55444, USA*

**Mariana Botero** acustica@abotero.com.co

*A Botero y CIA SCA*
*Calle 19 #9-50 Edificio Diario del Otún, Oficina 2105A*
*Pereira, Colombia 66000*

# Obituary

## Mahlon Daniel Burkhard, 1923–2021



**Mahlon Daniel Burkhard**, chairman of the board of HEAD acoustics, Inc., Brighton, Michigan, died on August 26, 2021, in Adamstown, Maryland, at the age of 98. He was born in Seward, Nebraska, on January 14, 1923.

Mahlon's dedication to acoustics for more than 70 years may have begun during service in the US Navy Pacific Fleet in 1943 as a landing craft pilot delivering US Marines to the beaches of Iwo Jima and Okinawa. He was in a craft less than 1,000 feet from the battleship USS Missouri when it fired its 16-inch guns over him, causing partial hearing loss and distortion in one ear for the rest of his life.

After his military duty, Mahlon finished his undergraduate degree at Nebraska Wesleyan University, Lincoln, and then received an advanced degree in acoustics from The Pennsylvania State University, University Park. He began his career at the National Bureau of Standards in Washington, DC, helping to establish the first standards for hearing aids. He moved into private industry as director of research at Industrial Research Products in Elk Grove Village, Illinois, heading teams developing and commercializing the Knowles Electronic Manikin for Acoustic Research (KEMAR), a line of digital audio signal-delay products, and multiple studies of electret materials and processes for electret condenser microphones. Under his direction, the communication microphones for the NASA Apollo lunar program were built.

Mahlon made many contributions to both the American National Standards Institute (ANSI) and International Electrotechnical Commission (IEC) standards committees, was a chair of the Engineering Acoustics Technical Committee of the Acoustical Society of America (ASA), and was active on the ASA Committee on Medals and Awards.

The 1972 development of Mahlon's digital signal delay was at the outdoor music pavilion at Ravinia, Highland Park, Illinois. Tests were performed in a long tunnel beneath the facility, using a speaker and microphones. Prototype delays were timed according to these tests and refined by listening.

That work began a long collaboration and friendship with Christopher Jaffe, the Ravinia acoustician. Jaffe challenged Mahlon to develop multitap nonrecursive delays as reverberators, extending the natural reflection character in performance spaces. Their first installation was for the 1980 NBC television series *Live from Studio 8H* featuring the New York Philharmonic under the direction of Zubin Mehta. Jaffe used Knowles electret condenser microphones flush mounted into stage ceiling reflectors there and in other concert spaces to feed "electronic reflected energy" systems (ERES) and produce "electronic forestage canopies" where physical ones could not exist, for example, the acoustic renovation of the Oakland Paramount Theater, Oakland, California.

Throughout his career, Mahlon was awarded seven patents and made many significant impacts. He attempted retirement in the late 1980s but immediately moved to Connecticut to work with Jaffe on performing arts electroacoustic systems and become president of Sonic Perceptions, Inc., Norwalk, Connecticut, the new Jaffe Acoustics-owned firm that introduced HEAD acoustics GmbH binaural technology to North America. Mahlon retired again in the mid-1990s to become chairman of the board of HEAD acoustics, Inc.

He was preceded in death by his wife Charlotte and son Douglas. He is survived by his sons John, David, and Ronald; five grandchildren; and three great-grandchildren.

### *Selected Publications of Mahlon Daniel Burkhard*

Burkhard, M. D. (Ed.) (1978). Manikin measurements. *Proceedings of Conferences Organized by M. D. Burkhard*, Industrial Research Products, Inc., Elk Grove Village, IL, Zurich, Switzerland, March 4, 1976; Washington, DC, April 5, 1976. Available at https://www.grasacoustics.com/files/m/a/KEMAR_Manikin_Measurements.pdf.

Burkhard, M. D., and Corliss, E. L. R. (1954). The response of earphones in ears and couplers. *The Journal of the Acoustical Society of America*, 26, 679.

Burkhard, M. D., and Sachs, R. M. (1975). Anthropometric manikin for acoustic research. *The Journal of the Acoustical Society of America*, 58, 214.

*Written by:*

**Wade R. Bray** wbray@headacoustics.com
*HEAD acoustics, Inc., Brighton, MI*

**Mead C. Killion** abonso@aol.com
*MCK Audio, Inc., Elk Grove Village, IL*

**Klaus Genuit** Klaus.Genuit@head-acoustics.com
*HEAD acoustics GmbH, Herzogenrath, Germany*

## Obituary

# Tony Frederick Wallace Embleton, 1929–2020

**Tony Frederick Wallace Embleton** died on November 13, 2020, in Woodbridge, ON, Canada. He was a leader in acoustics in the United States and Canada and also became known internationally. Tony attracted a steady stream of postdoctoral fellows and sabbatical visitors from India, Japan, Europe, and the United States.

Tony was born in Hornchurch, Essex, United Kingdom, on October 1, 1929. He earned a PhD in physics in 1952 from the Imperial College London, United Kingdom, under R. W. B. Stephens. A one-year postdoctoral fellowship at the National Research Council (NRC) in Ottawa, ON, Canada, turned into a four-decade career where he attained the rank of principal research officer. He was section head from 1985 until his retirement in 1990.

Tony's research addressed many significant and practical concerns. In the early 1950s, the intense noise generated by large suction rolls in the production of paper was a serious industrial concern. Together with George Thiessen, then head of the Acoustics Section at the NRC, Tony demonstrated a considerable noise reduction by substituting the simple patterns of holes in the cylinders with more complex patterns.

In the early 1960s, Tony's research included noise reduction in centrifugal blowers, axial-flow compressors, and stator blading for noise reduction in turbomachinery. The latter research increased efficiency in jet engines while quieting them and was eagerly adopted around the world. His work on mufflers for percussive pneumatic machines not only quietened pneumatic drills but led to higher drilling speeds, less icing in the muffler, and less vibration. While this work was in progress, he undertook to provide the laboratory with the absolute measurement of sound pressure by developing a reciprocity system for the calibration of microphones.

In the 1970s, Tony turned his interest to sound propagation outdoors. This led to a series of theoretical and experimental studies of sound propagation outdoors with NRC colleagues that addressed (1) the effect of the ground on near-horizontal sound propagation; (2) the measurement of ground impedance and acoustic characteristics of actual ground surfaces (e.g., asphalt, gravel, grass, snow); (3) the phase and amplitude fluctuations due to turbulence; and (4) the refraction due to wind and temperature.

Tony's main professional home was the Acoustical Society of America (ASA). He served as vice president; president; standards director; technical chair for the meetings in Ottawa and Honolulu, Hawai'i; and general chair of the 1981 meeting in Ottawa. He was a recipient of the R. Bruce Lindsay Award, a Silver Medal in Noise, and the ASA Gold Medal. Tony was also active in the Canadian Acoustical Association (CAA), the Institute of Noise Control Engineering (INCE), and the International INCE.

Tony's research was significant and of the highest quality; his service to societies was diligent, efficient, and cooperative. He was always willing to help, whether you were a society president or a student new to acoustics. At meetings, he was often surrounded by people, young and old, wanting to access his vast store of knowledge.

Tony was preceded in death by his wife, Eileen, in 2016. They are survived by their daughter Sheila and granddaughter Anne.

### *Selected Publications by Tony Frederick Wallace Embleton*

Embleton, T. F. W. (1971). Mufflers. In Beranek, L. L. (Ed.), *Noise and Vibration Control*. McGraw-Hill, New York, NY, pp. 362-405.

Embleton, T. F. W., and Daigle, G. A. (1991), Atmospheric propagation, In Hubbard, H. H. (Ed.), *Aeroacoustics of Flight Vehicles, Theory and Practice. Vol. 2: Noise Control*. Reference Publication 1258, NASA Langley Research Center, Hampton, VA, pp. 53-99. Reprinted by the Acoustical Society of America, 1994.

Embleton, T. F. W., and Thiessen, G. J. (1958). Efficiency of a linear array of point sources with periodic phase variation. *The Journal of the Acoustical Society of America* 30, 1124-1127.

Embleton, T. F. W., Piercy, J. E., and Olson, N. (1976). Outdoor propagation over ground of finite impedance. *The Journal of the Acoustical Society of America* 59, 267-277.

*Written by:*

**Gilles A. Daigle** gilles_daigle@sympatico.ca

**Michael R. Stinson** mikestinson42@gmail.com

*National Research Council, Ottawa, ON, Canada*

# Obituary
## John Richard Preston, 1945–2021

**John Richard Preston** passed away on April 20, 2021, in State College, Pennsylvania.

John received a BSc degree in physics from the University of Massachusetts, Amherst; MSc in physics from the University of Maryland, College Park; and MSEE degree in physics from George Washington University, Washington, DC. He received his PhD in acoustics from The Pennsylvania State University (Penn State), University Park.

Initially, John worked at Tetra Tech, Inc., Rosslyn, VA, from 1973 to 1983 and served as vice president at Amron Corporation, Washington, DC, from 1983 to 1989. After 16 years in the private sector, John joined the NATO Centre for Maritime Research and Experimentation (then SACLANTCEN), La Spezia, Italy, as a research scientist from 1989 to 1995. He then joined the Applied Research Laboratory at Penn State from 1995 to 2015 when he retired but remained as emeritus research associate.

Within the community, John was recognized as an exceptionally gifted scientist, both for his attention to experimental detail and his collaborative nature. Collaborative measurements are a crucial part of underwater acoustics research. This is driven by the large resource requirements to conduct them and the technical breadth of data that requires both experimentalists and theorists to interpret. He was chief scientist on several ocean acoustic experiments involving multiple ships and international partners, and he participated in numerous others.

John's major scientific contributions have been in the collection and analysis of data using towed arrays. For the interpretation of long-range reverberation, he developed polar plots in which the towed array beam time series are georeferenced and overlaid on the underlying bathymetry. This allows scattering features to be mapped and potential targets identified and is now standard in many naval operational systems. Later, he pioneered the extraction of quantitative environmental information from reverberation data, in a number of NATO Rapid Environmental Assessment exercises. He was elected a Fellow of the Acoustical Society of America.

At Penn State, John was asked by the Office of Naval Research to specify a towed research array with a high dynamic range and to oversee its construction, maintenance, and deployment at sea; this became the Five Octave Research Array (FORA). In contrast to many research arrays, FORA worked. This was not simply good luck. The design, choice of manufacturer, and maintenance of it were critical items. As well as the hardware aspects, John spent a great deal of effort making sure the data were accurately calibrated and of the highest quality. Many researchers used the data he collected. He participated in the geoclutter program and follow-on experiments. These determined that geoclutter (spurious seabed scattering that interferes with target detection) was, in some circumstances, really bioclutter (fish).

In summary, during a career spanning 45 years, he pioneered numerous efforts in underwater acoustic measurements and analysis and research array developments. He is clearly recognized as a key leader in the field and a strong and valuable collaborator and will be greatly missed by his colleagues and family.

### *Selected Publications by John Richard Preston*


Preston, J. R. (2000). Reverberation at the Mid-Atlantic Ridge during the 1993 ARSRP experiment seen by R/V Alliance from 200–1400 Hz and some modeling inferences. *The Journal of the Acoustical Society of America* 107, 237-259.

Preston, J. R. (2007). Using triplet arrays for reverberation analysis and inversions. *IEEE Journal of Oceanic Engineering* 32, 879-896.

Preston, J. R., and Ellis, D. D. (2009). Extracting bottom information from towed-array reverberation data. Part I: Measurement methodology. *Journal of Marine Systems* 78, S359-S371.

Preston, J., and Nisley, R. (1978). Single frequency modulation model for surface reflection of a cw tone. *The Journal of the Acoustical Society of America* 64, 601-604.



***Written by:***

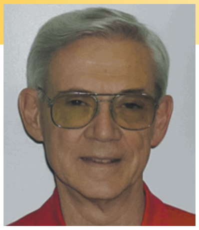
**David Bradley** David.Bradley@unh.edu
*University of New Hampshire, Durham*

**Dale Ellis** daledellis@gmail.com
*Dale Ellis Scientific Inc., Dartmouth, NS, Canada*

**Paul Hines** phines50@gmail.com
*Dalhousie University, Halifax, NS, Canada*

# Obituary

## Charles Schoff Watson, 1932–2021



**Charles Schoff (Chuck) Watson**, a Fellow of the Acoustical Society of America, died at age 89 on September 10, 2021, in Bloomington, Indiana. Chuck was a prolific contributor to hearing and communication science over a long, fruitful career, and he remained active in research to the end, receiving his last National Institutes of Health (NIH) research grant shortly before he passed away.

Born and raised in Chicago, Illinois, Chuck went to Indiana University (IU), Bloomington, for his undergraduate and graduate degrees. In 1963, he earned his doctorate in experimental psychology, with James Egan as his advisor. As a graduate student, he collaborated on a seminal study of the effects of intense noise on the mammalian ear with his lifelong friend and colleague James D. Miller.

Chuck took his first academic position at the University of Texas at Austin, where he examined basic issues in psychoacoustics and signal detection theory. At the Central Institute for the Deaf (CID) in St. Louis, Missouri, he created one of the first hearing research laboratories to be fully computerized. There, he began a series of studies of listeners' abilities to discriminate complex patterns using brief 10-tone sequences as surrogates for spoken words. This research helped move the field of psychoacoustics away from the study of simple sounds to the investigation of more naturalistic, complex stimuli. This work also revealed very large influences of trial-to-trial uncertainty on the discrimination of complex sounds, providing important early examples of informational masking. Chuck left the CID to serve as the first director of research at Boys Town National Research Hospital, Omaha, Nebraska, from 1977 to 1983, where he established a premier auditory research program (see https://bit.ly/AT-boystown). In 1983, Chuck returned to Indiana University to serve as chair of the Department of Speech and Hearing Sciences. There, he continued his work with complex sounds but also began to lead larger projects such as the six-year Benton-IU Project, which examined the influences on success in learning to read in elementary-school children, and included a large-scale study of individual differences in auditory abilities.

In 1989, Chuck, Diane Kewley-Port, and Dan Maki founded Communications Disorders Technology (CDT), a small business to develop new technologies for hearing and communication science. Under Chuck's leadership, CDT developed training systems for improving speech perception and production and also the National Hearing Test, a digits-in-noise hearing test that can be taken over the telephone. Just a few days before Chuck died, he was gratified to receive an NIH grant that would enable development of the National Hearing Test to continue.

Outside his work life, Chuck had many interests, most notably tennis. He met his wife Betty on the tennis court in St. Louis, and they played countless games together over the next five decades. Chuck is survived by Betty, his wife of 51 years; daughters Ann, Mary, Katharine, and Elizabeth; five grandchildren; and brother Donald Watson.

### *Selected Publications by Charles Schoff Watson*


Kidd, G. R., Watson, C. S., and Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America* 122, 418-435.

Watson, C. S. (1987). Uncertainty, informational masking and the capacity of immediate auditory memory. In Yost, W. A., and Watson, C. S. (Eds.), *Auditory Processing of Complex Sounds*. Erlbaum Associates, Hillsdale, NJ, pp. 267-277.

Watson, C. S., Kidd, G. R., Horner, D. G., Connell, P. J., Lowther, A., Eddins, D. A., Krueger, G., Goss, D. A., Rainey, B. B., Gospel, M. D., and Watson, B. U. (2003). Sensory, cognitive, and linguistic factors in the early academic performance of elementary school children: The Benton-IU Project. *Journal of Learning Disabilities* 36, 165-197.

Watson, C. S., Kidd, G. R., Miller, J. D., Smits, C., and Humes, L. E. (2012). Telephone screening tests for functionally impaired hearing: Current use in seven countries and development of a U.S. version. *Journal of the American Academy of Audiology* 23, 757-767.



*Written by:*

**Gary R. Kidd** kidd@indiana.edu

*Indiana University, Bloomington*

## Advertisers Index



## Advertising Sales & Production


**Debbie Bott,** Advertising Sales Manager
*Acoustics Today*, c/o AIPP, Advertising Dept
1305 Walt Whitman Rd, Suite 110, Melville, NY 11747-4300
Phone: (800) 247-2242 or (516) 576-2430 | Fax: (516) 576-2481 |
Email: dbott@aip.org


*For information on rates and specifications, including display, business card and classified advertising, go to Acoustics Today Media Kit online at:* https://publishing.aip.org/aipp_at-ratecard-2022 *or contact the Advertising staff.*

## Business Directory

HBK
HOTTINGER BRÜEL & KJÆR

TECHNOLOGY BY
Brüel & Kjær

DESIGNED FOR YOUR JOB

# Need an easier way to ensure noise compliance? JOB DONE.


For noise measurement surveys, you need a sound level meter solution that gets your job done faster, easier and problem-free. The new B&K 2245 gives you absolute confidence and control through user-friendly mobile apps and functionality tailored for your task, including customizable checklists, sound isolation markers, on-site analysis, photo embedding, and more.

To simplify your job-to-do, visit **hbkworld.com/2245**

HBM
Brüel & Kjær