VOLUME 7, ISSUE 4 OCTOBER 2011

Acoustics Today

Voices and Listeners
Imitation in Speech
Human Voice in Evolutionary Perspective
Phonetics of Endangered Languages
and more

A publication of
the Acoustical Society
of America

Speech Communication

Photo Courtesy of ETS-Lindgren

EDITOR
Dick Stern (AcousticsToday@aip.org)

CONTRIBUTORS
Molly Babel, Mary Bates,
Christian T. DiCanio,
Jody Kreiman, Elaine Moran,
Michael J. Owren, Patricia A. Shaw,
Diana Sidtis, Dick Stern,
Michael Vorländer, D. H. Whalen

ADDRESS
Acoustical Society of America
Suite 1NO1
2 Huntington Quadrangle
Melville, NY 11747-4502

ASA EXECUTIVE DIRECTOR
Charles E. Schmid

ASA EDITOR-IN-CHIEF
Allan D. Pierce

ADVERTISING
Robert Finnegan, Director, Journal Advertising & Exhibits, AIP

ACOUSTICAL SOCIETY OF AMERICA

The *Acoustical Society of America* was founded in 1929 to increase and diffuse the knowledge of acoustics and to promote its practical application. Any person or corporation interested in acoustics is eligible for membership in the Society. Further information may be obtained by addressing Elaine Moran, ASA Office Manager, Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Phone: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: AcousticalSociety.org

FUTURE ARTICLES

Standards

Psychological and Physiological Acoustics

Marine Mammals and Fish

Medical Acoustics

Physical Acoustics

ACOUSTICS TODAY (ISSN 1557-0215, coden ATCODK) October 2011, volume 7, issue 4 is published quarterly by the Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Periodicals Postage rates are paid at Huntington Station, NY, and additional mailing offices. POSTMASTER: Send address changes to Acoustics Today, Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. **Copyright ©2011, Acoustical Society of America. All rights reserved.** Single copies of individual articles may be made for private use or research. Authorization is given to copy articles beyond the use permitted by Sections 107 and 108 of the U.S. Copyright Law. To reproduce content from this publication, please obtain permission from Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA via their website: www.copyright.com/, or contact them at (978)-750-8400. Persons desiring to photocopy materials for classroom use should contact the CCC Academic Permissions Service. Authorization does not extend to systematic or multiple reproduction, to copying for promotional purposes, to electronic storage or distribution, or to republication in any form. In all such cases, specific written permission from the American Institute of Physics must be obtained. Permission is granted to quote from *Acoustics Today* with the customary acknowledgment of the source. To reprint a figure, table, or other excerpt requires the consent of one of the authors and notification to AIP. Address requests to AIP Office of Rights and Permissions, Suite 1NO1, 2 Huntington Quadrangle, Melville NY 11747-4502; Fax (516) 576-2450; Telephone (516) 576-2268; E-mail: rights@aip.org. *Acoustics Today* is also reproduced in the Acoustical Society of America's Digital Library shortly after a print copy becomes available. Members and non-member subscribers may download articles or issues for their personal or professional use. The articles may not be altered from their original printing and pages that include advertising may not be modified. Articles may not be reprinted or translated into another language and reprinted without prior approval from the Acoustical Society of America as indicated above. Non-member subscribers may purchase articles at a nominal cost under the same conditions that are specified for members and non-member subscribers.

SMALL ENOUGH TO DISAPPEAR IN YOUR HAND
—YET CAPTURES A WIDE WORLD OF SOUND!

the **RION NL-27** Sound Level Meter (Class 2)

Features:

- Wide 107 dB linearity range allows sound level measurements from 30 to 130dB without range switching
- Easy measurement of sound pressure level (Lp), equivalent continuous sound pressure level (Leq). maximum sound pressure level (Lmax), sound exposure (LE), and peak sound level (Lpeak)
- Manual store function and capability to transfer data via optional USB adapter cable
- 7.5 hours operation on two size AAA (IEC R03) alkaline batteries
- Especially useful for community noise enforcement, schools, clubs, exercise spas, police work, and industrial noise

Scantek, Inc.
Sound & Vibration Instrumentation and Engineering
www.scantekinc.com • info@scantekinc.com
800-224-3813

When "BUY" does not apply,
give RENTAL a try!

At Scantek, Inc. we specialize in **Sound and Vibration Instrument Rental** with ***expert assistance***, and fully calibrated instruments for:

Applications
- Building acoustics
- Sound power measurement
- Community noise
- Building vibration
- Industrial noise
- Human body vibration
- Machine diagnostics
- Vibration measurement

Instruments
- analyzers
  - *FFT and real-time*
  - *1/3 and 1/1 octave bands*
- noise and vibration dosimeters
- vibration meters
- human body dose/vibration
- A-weighted sound level meters
- rangefinders
- GPS
- windscreens
- wide range of microphones and accelerometers

Scantek, Inc.
Sound & Vibration Instrumentation
and Engineering
www.scantekinc.com
info@scantekinc.com
800-224-3813

TAPPING
just got easier!

The rugged brand new Norsonic N-277 Tapping Machine is ideal for making structureborne impact noise tests for floor/ceiling combination in the field and in the laboratory. This third-generation unit meets all international and US standards.

- Impact sound transmission testing according to ISO140 part VI, VII and VIII, ASTM E-492 and ASTM E-1007.
- Remote operation from hand switch or PC; Mains or battery operation.
- Low weight 10 kg (22 lb) incl. battery and wireless remote option.
- Built in self check of hammer fall speed, and tapping sequence for automatic calibration of major components.
- Retractable feet and compact size provide easy transportation and storage.

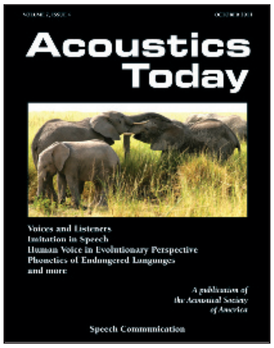
Scantek, Inc.
Sound & Vibration Instrumentation
and Engineering

www.scantekinc.com
info@scantekinc.com
800-224-3813

Acoustics Today

A Publication of the Acoustical Society of America

Volume 7, Issue 4 **October 2011**



Acoustics Today

Voices and Listeners
Imitation in Speech
Human Voice in Evolutionary Perspective
Phonetics of Endangered Languages
and more

A publication of
the Acoustical Society
of America

Speech Communication

Cover: Playback experiments (in which recorded vocalizations are broadcast in the field to freely-behaving animals and responses are recorded) have shown that the extent of elephants' defensive responses (bunching together, retreating) to the voices of elephants from the other family groups encountered within their range can be predicted by the frequency with which those animals are encountered. Photo by Jody Kreiman

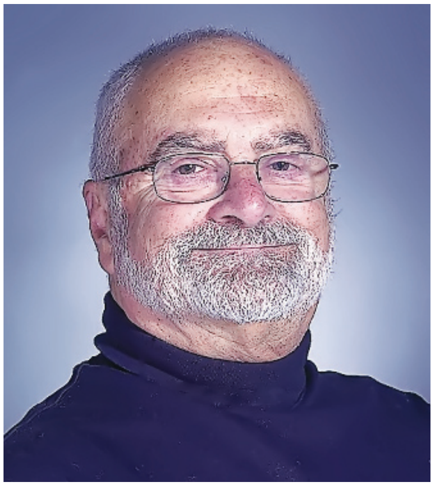
Articles:

Departments:

Business Directory

FROM THE EDITOR

Dick Stern
1150 Linden Hall Road
Boalsburg, Pennsylvania 16827



I am extremely grateful to our guest editor, Jody Kreiman, for her ability to assemble an outstanding group of authors for this comprehensive issue on speech communication. Each discussed their area of expertise so as to provide the reader a glimpse into this fascinating subject.

Dick Stern

FROM THE GUEST EDITOR

Jody Kreiman

Some jobs are more fun than others, and guest editing this issue of *Acoustics Today* was one of the good ones. The speech communication technical committee encompasses such a wide range of topics that I found it impossible to pick one on which to focus the issue, so instead I asked colleagues whose work I admire to write about what they were doing. The four papers that appear here are not especially "speechy" in the traditional sense of speech production, perception, and acoustics, but they highlight the growing scope of what now falls under the "speech communication" umbrella, and the interdisciplinary approaches that characterize much current research on speech. They also share a focus on the role that speech in particular and vocalization in general play in the biology and social lives of humans and non-human animals. As all the authors point out, considering speech in the broader context of social and biological functioning may enhance or even fundamentally alter our understanding of *why* speech behavior varies as it does. And the "why" questions are the most fun of all.

VOICES AND LISTENERS: TOWARD A MODEL OF VOICE PERCEPTION

Jody Kreiman
Department of Head/Neck Surgery
University of California, Los Angeles, School of Medicine
Los Angeles, California 90095

and

Diana Sidtis
Department of Communicative Sciences and Disorders
New York University
New York, New York 10012

and

Nathan Kline Institute for Psychiatric Research
Orangeburg, New York 10962

Introduction

As humans, we are exquisitely tuned to voices and all that they are capable of conveying (Table 1). On hearing someone speak, we quickly infer details about gender, age, education, and geographical background (Sebastian and Ryan, 1985). We listen for signs of interest, well-being, competence, and cooperation, or coldness, ineptness, and resistance. Along with these, mood, emotional conditions, personality, and psychological status are simultaneously assessed by the listener, with varying accuracies. These speaker characteristics constitute a very large, complex array and pose huge challenges to analytic approaches.

Not least important among the characteristics listeners extract from voices is the identity of the person who is speaking. The person may be someone familiar; or, much less commonly, we may try to identify a stranger, for example in a forensic situation. In this paper we will describe some of the important differences between these two classes of stimuli—familiar and unfamiliar voices—and the cognitive and neuropsychological processes used in their perception. We then present a preliminary model of the manner in which listeners tackle each kind of information. taking into account underlying brain structures involved in these disparate processes. Finally, we explore the implications of our model for measurement of quality in the voice clinic and elsewhere.

"The wide distribution of voice recognition abilities across species, combined with the clear survival value of such abilities and their strikingly full-blown ontogenetic appearance, suggests that familiar voice recognition is evolutionarily very old."

Which came first: Familiar or unfamiliar voices?

Unfamiliar voices surround us in life, from the sound of the cashier greeting us at the market, to students talking in the hall outside a classroom, to the voices of other patrons conversing in a background of chattering and cheering at a sports event. When we pay attention to such voices, they can provide substantial amounts of information about the speaker, as noted above, and as a result it is easy to assume (as we ourselves have done in the past) that the unfamiliar voice is somehow the basis of the perceptual processes used to extract information from all voices. After all, we reasoned, every voice was unfamiliar before it was familiar, so logically familiarity develops out of unfamiliarity, which implies that the unfamiliar is foundational.

In the beginning was the familiar voice

A substantial body of evidence suggests that the assumption that unfamiliar voices are fundamental is fundamentally wrong. First, we note that the ability to recognize a familiar voice (and especially the voice of a parent, offspring, or mate) is very widespread among animals. Many, many species, including deer (Torriani *et al.*, 2006), sheep (e.g., Sebe *et al.*, 2010), wolves (Goldman *et al.*, 1995), mares (Wolskia *et al.*, 1980), many marine mammals (e.g., Insley, 2001; Pitcher *et al.*, 2010), rodents (Fuchs *et al.*, 2010), bats (Voigt-Heucke *et al.*, 2010), amphibians (Bee and Gerhardt, 2002; Simmons, 2004), and birds ranging from penguins (e.g., Jouventin and Aubin, 2002) to parrots (Berg *et al.*, 2011) also recognize the familiar voices of their kin. Recognition often begins very early in life, or even immediately; for example, the developing human fetus has been shown to recognize the voice of its mother (Kisilevsky *et al.*, 2003). Scientists have only begun to appreciate the social complexity and sophistication of these behaviors. Recent studies reveal that seal mothers time their departure for food gathering to coincide with successful voice recognition by their pups, so that reuniting on their return will be successful (Charrier *et al.*, 2001). In comparison, mother evening bats recognize the

Table I. Some of the kinds of judgments that listeners can make from voices

Meaning of the spoken message
Physical characteristics of the speaker
Age
Appearance (size, attractiveness)
Drunk?
Healthy?
Personal identity
Race, ethnicity
Sex
Sexual orientation
Smoker?
+/- Teeth?
Tired?
Psychological characteristics of the speaker
Relaxed?
Competence
Lying?
Mood or emotional status
Intelligence
Personality
Psychiatric status
Under stress?
Social characteristics of the speaker
Education
Occupation
Regional origin
Role in the conversation
Social status

voices of their offspring immediately after birth, suggesting calls have a genetic component (Scherrer and Wilkinson, 1993). These biological scenarios cast an eerie doubt on the traditional assumption that all voices, at the first instant, are unfamiliar.

Voice recognition facilitates reunions between foraging parents and offspring that are mobile or located in a crowded crèche, helps animals ensure that care is provided to the correct infant, and promotes bonding between mothers and infants. The wide distribution of voice recognition abilities across species, combined with the clear survival value of such abilities and their strikingly full-blown ontogenetic appearance, suggests that familiar voice recognition is evolutionarily very old. In fact, it may have appeared by the time that frogs emerged (Burke and Murphy 2007; Bee and Gerhardt, 2002; see Kreiman and Sidtis, 2011, for more review). Studies showing that primate brains may have voice-sensitive areas analogous to those seen in human infants as young as 7 months (Petkov *et al.*, 2008; Petkov *et al.*, 2009; Grossmann *et al.*, 2010) further point to a long evolutionary history of voice recognition abilities (Belin and Grosbras, 2010). Producing and recognizing familiar voice patterns thus antedates, by millions of years, the more lauded evolutionary development of speech and language in human communication and cognition. For discerning cohort, friend from foe, and recognizing intimate family members—and being able to achieve this at a distance and in the dark—the preeminence of the familiar voice pattern in evolutionary biology can hardly be exaggerated (Sidtis and Kreiman, 2011).

Recognition of the familiar voices of animals that are not first-degree relatives is less common, but helps maintain proximity and promotes group cohesion in social animals by providing a means of separating insiders from outsiders, even at a distance (Fig. 1). For example, female vervet monkeys can recognize the voices of their own offspring but also of unrelated juveniles, and can associate those voices with the correct mother (Cheney and Seyfarth, 1980); and female baboons recognize both the screams and threat grunts of unrelated individuals (Cheney and Seyfarth, 1999). Playback experiments (in which recorded vocalizations are broadcast in the field to freely-behaving animals and responses are recorded) have shown that the extent of elephants' defensive responses (bunching together, retreating) to the voices of elephants from the other family groups encountered within their range can be predicted by the frequency with which those animals are encountered. Response patterns imply an ability to recognize about 100 individuals (McComb *et al.*, 2002). In the vast landscape of biological vocal recognition, not to be neglected is the ability of nonvocal reptiles to recognize alarm calls of other species (Vitousek *et al.*, 2007). Although these abilities are impressive, they pale in comparison to prodigious human abilities to recognize the voices of people we are not related to. Besides our friends, family, neighbors, and other associates (the "familiar-intimate" set), thanks to the media we are easily able to recognize and identify scores of people we have never spoken to or even met (the "familiar-famous" voices: actors, politicians, announcers, broadcasters), as well as fictional beings of endless variety (Bugs Bunny, Hal the computer, and Robby the Robot, for example). In fact, studies of familiar face recognition (Bahrick *et al.*, 1975) and informal voice recognition challenges suggest that there may not be an upper limit to the number of voices humans can recognize (Ladefoged and Ladefoged, 1980).

In contrast, it is not clear how much attention listeners of any species actually pay to unfamiliar voices under normal circumstances. Most animals, including humans, treat unfamiliar voices as part of the background of noise that surrounds them every day. As an example, imagine yourself on a busy street, surrounded by strangers talking to each other or on their cell phones. The voices we hear under these circumstances, although ubiquitous, barely penetrate consciousness. In fact, in a study in which the original caller was surreptitiously replaced with a different talker during a telephone survey call, only 6% of subjects noticed the change (Fenn *et al.*, 2011). In contrast, the voice of an approaching friend jumps out from a background of unknown voices, much as the sound of our own name emerges from the unattended

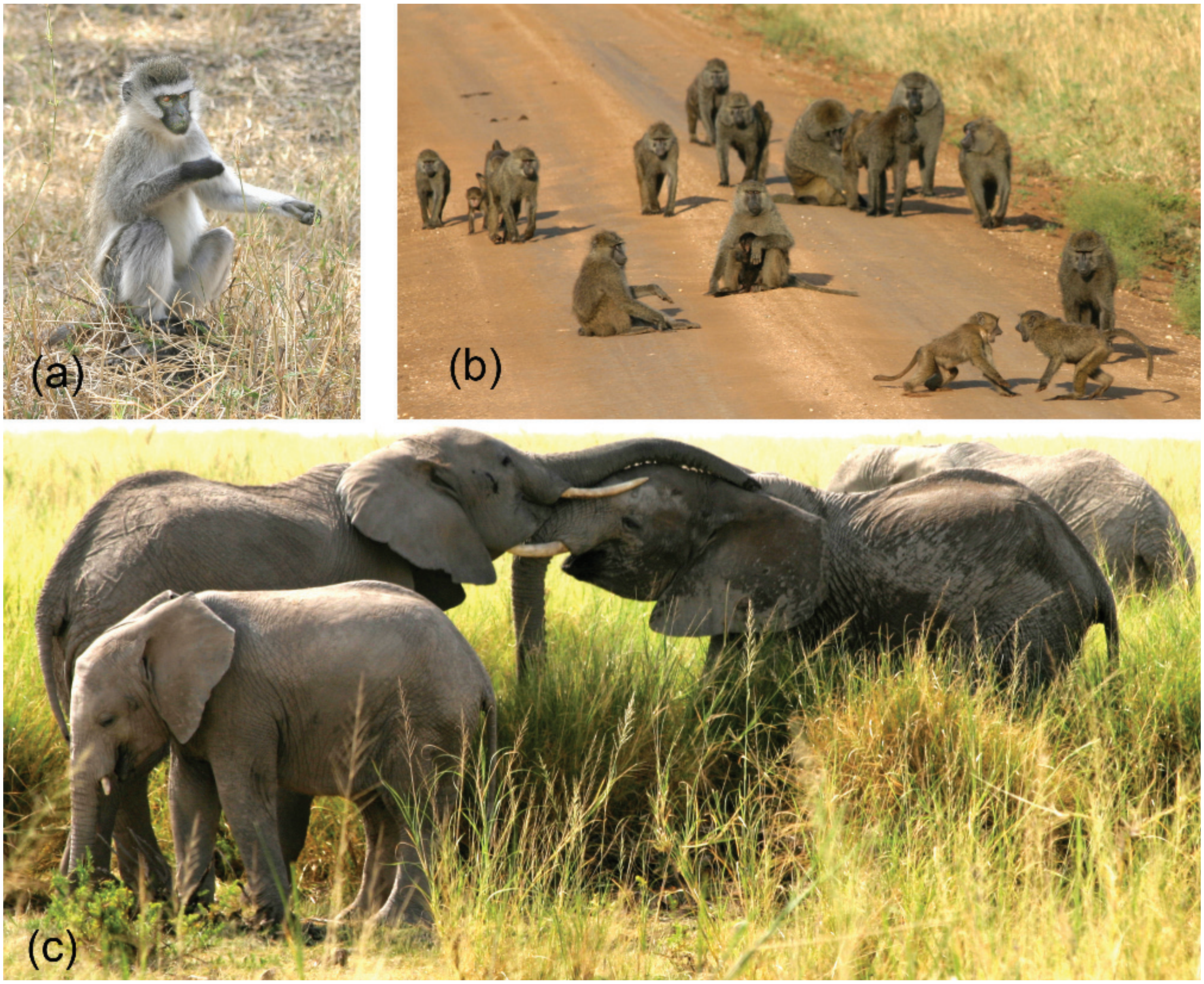


Fig.1. Some non-human animals that recognize the voices of familiar non-family members. A: vervet monkeys. B: baboons. C: elephants.

chatter in a crowded room. From these several perspectives, we must conclude that it is the familiar voice pattern that plays the dominant role in animal biology and human culture (Sidtis and Kreiman, 2011).

The brain behind the voice

These findings suggest that there should be differences in the neuropsychological and cognitive processes involved in perceiving familiar versus unfamiliar voices. That is, if recognizing a familiar voice is "basic" in some way, we might expect that there exist specific, efficient neuropsychological mechanisms to support this ability. Similarly, if unfamiliar voices are not important or salient stimuli, we might expect a messier set of processes to be engaged if and when we are forced to deal with them.

In fact, a substantial number of studies point to such differences. Recognizing a familiar voice and discriminating among unfamiliar voices are dissociated neuropsychological abilities, meaning that either one can be independently disrupted by neurological damage, leaving the other entirely intact (Van Lancker and Kreiman, 1987). Familiar voices engage a large expanse of cerebral systems. Upon recognizing a familiar voice, parietal lobes establishing associations in declarative memory, subcortical structures modulating memory, motivation and emotion, frontal lobes organizing and integrating behaviors, and temporal lobes processing auditory patterns and selected auditory features all participate (see Kreiman and Sidtis, 2011, for extended review). Although multiple cerebral structures play significant roles in processing familiar voices, studies of performance following brain lesions and in functional imaging give a role to the right cerebral hemisphere as a final common pathway for voice recognition, especially of familiar stimuli (e.g., Van Lancker *et al.*, 1989; Neuner and Schweinberger, 2000; Belin *et al.*, 2000; Latinus and Belin, 2011b; Gainotti, 2011).

Consistent with right hemisphere participation in familiar voice perception and recognition, evidence suggests that familiar voices comprise distinctive, integral, heterogeneous patterns, which can be accessed as unique, holistically stored units. These integral patterns resist systematic decomposition into bundles of separable features. Parameters like F0, timbre, and intensity—cornerstones of voice quality analysis—interfere with each other perceptually, such that irrelevant, unattended variation on one parameter facilitates or interferes with listeners' judgments of the other, depending on whether that irrelevant variation is or is not correlated with the attended dimension (Melara and Marks, 1990; Li and Pastore, 1995). Similarly, studies using unfamiliar voices show that the harmonic and inharmonic (noise) parts of the voice interact perceptually, so that listeners' sensitivity to either depends on energy levels in both (Kreiman and Gerratt, in press); and sensitivity to tremor rates in voice depends on the magnitude of the tremor, and vice versa (Kreiman *et al.*, 2003). Further, listeners' relative inability to reliably and consistently isolate single dimensions in a voice pattern is the largest source of error in voice quality ratings (Kreiman *et al.*, 2007). These findings argue against reliance on feature-based models of voice quality of the sort that underlie most clinical voice evaluation protocols (about which more in a moment). As most studies of voice and voice quality perception use unfamiliar voices as stimuli, understanding the functional and perceptual roles of auditory-acoustic cues or features in the perception of familiar voices has only been crudely begun (Van Lancker *et al.*, 1985). These early attempts have shown that individual familiar voice patterns vary greatly in how (and how much) cues such as F0 or breathiness contribute to the recognition process.

While familiar voice recognition engages pattern recognition processes of the right hemisphere, discriminating among unfamiliar voices or "identifying" a voice heard only once or twice before (for example, in a voice lineup) engages auditory temporal receiving areas on both sides of the brain (Van Lancker *et al.*, 1989), and seemingly involves both pattern recognition and featural analysis/matching skills. Error patterns in long-term memory tasks suggest that unfamiliar voices are encoded in terms of a generalized template or "prototype," along with a set of deviations from that prototype which are forgotten over time so that memory tends to converge on average-sounding voices no matter what voice was heard originally (Papcun *et al.*, 1989). Similarly, memory tests in change deafness studies (testing listeners' awareness of abrupt voice quality changes during normal interaction) suggest that listeners remember only coarse differences between unfamiliar voices under normal circumstances (a "gist-based" representation, Fenn *et al.*, 2011, p. 1454), and that memory for specific acoustic details of a voice may be weak or entirely absent. In contrast, for familiar voices, a complex, unique perceptual pattern is stored along with an array of personally-relevant associations (appearance, biographical and episodic history, affective nuances, and so on); recognition occurs within a second or two; and the "cues" triggering recognition vary widely with vocal pattern (Schweinberger *et al.*, 1997a). These findings have led us to conclude that all voices are fundamentally patterns, and that pattern recognition and featural analysis reciprocally operate, in different degrees, for all voice perception processes, depending on the status of the voice with respect to its familiarity to the listener.

A large body of behavioral evidence also supports the notion that voices are best viewed as patterns. In a "repetition priming" protocol, listeners' accuracy in judging whether or not a voice sample was famous improved when they had previously heard a *different* sample of the target voice, so that the advantage transferred between tokens of speech and did not depend on the specific acoustic details of an individual sample (Schweinberger *et al.*, 1997b). Adaptation studies provide similar evidence. In these studies, the experimenter creates a stimulus continuum by "morphing" between two voices—for example, those of a male and a female. When listeners hear tokens taken from one end of the continuum, their judgments of ambiguous stimuli from the middle of the continuum shift, so that hearing a relatively male sample 3 or 4 times makes the ambiguous sample sound more female, and hearing tokens from the female end of the continuum makes it sound more male. These effects have been shown for judgments of speaker identity (familiar voices: Zäske *et al.*, 2010; trained to recognize: Latinus and Belin, 2011a), but also for

Table 2: Some of the factors affecting listeners' ability to identify an unfamiliar voice

The a priori "distinctiveness" of the target voice
The speaker and listener's accents and/or the language spoken
The presence of disguise or mimicry
The duration and phonetic content of the speech sample
Whether or not the same sample is used at learning and test
Filtering (for example, by a telephone or recording)
The listener's inherent ability to remember voices
The listener's attention at learning and at test
The listener's sex
The listener's professional training
The delay between hearing a voice and identifying it
The number and kind of distracter voices in the lineup
The instructions the listener receives

Fig. 2. A fox and a hedgehog.

perception of emotion (Bestelmeyer *et al.*, 2010), speaker sex (Schweinberger *et al.*, 2008), speaker age (Zäske and Schweinberger, 2011), and ratings of roughness (Gerratt *et al.*, 1993) from voice, and are interpreted as reflecting adaptation of a central representation (a pattern), rather than the effects of specific acoustic characteristics of the stimuli. Finally, studies of familiar voice recognition (e.g., Van Lancker *et al.*, 1985) have demonstrated that the acoustic cues to personal identity vary from voice to voice, and the importance of a given cue depends on the context of the complete voice pattern in which that cue operates, and not on the value of the cue itself. Thus, unusual pitch contours or a marked foreign accent (for example) may be essential cues to a speaker's identity, or not, depending on the other cues that are available to listeners. It is thus impossible to devise a set of features that are important for recognition of all voices: The importance of a given cue depends on the pattern in which the cue appears and on the status of the voice as familiar—and stored as a personally relevant auditory object—or unfamiliar and handled perceptually in terms of stereotypes or generalized templates.

One final difference between familiar voice recognition and unfamiliar voice discrimination is that familiar voice patterns are remarkably robust, so that we can recognize a familiar voice in noise, based on very short samples (often just the word "Hi" on a band-limited telephone line), even when the voice has not been heard for years or even decades and has changed with time (voices appear to change less with age than do faces). In contrast, virtually anything will disrupt efforts to match an unfamiliar voice to a decaying memory trace. Studies (primarily focusing on forensic situations) have shown that identification scores fluctuate as a function of a wide range of factors characterizing the speaker, the listener, and the circumstances surrounding originally hearing and subsequently identifying the voice, (Table 2; see Bricker and Pruzansky, 1976, or Kreiman and Sidtis, 2011, for review). It appears that the greater the reliance on featural extraction, comparison, and analysis, the worse we are at the task.

Table 3: A few examples of terms for voice quality, from a long history of interest in such descriptors.

Julius Pollux (Second century AD; cited in Austin, 1806)	**Moore (1964)**	**Gelfer (1988)**
Clear (claram)	Clear, light, white	Clear
Deep (gravam)	Deep	Resonant, low
Brilliant (splendidam)	Bright, brilliant	Bright, vibrant
Smooth (suavam)	Cool, smooth, velvety	Smooth
Attractive (illecebrosam)	Pleasing	Pleasant
Dull (fuscam)	Dead, dull, heavy	Dull, heavy, thick
Thin (angustam)	Constricted, heady, pinched, reedy, shallow, thin	Thin
Harsh (asperam)	Harsh, strident, twangy	Harsh, gravelly
Unsound, hoarse (infirmam, raucam)	Faulty, hoarse, poor, raucous, rough	Hoarse, rough, labored, noisy
Brassy (aeneam)	Buzzy, clangy, metallic	Metallic

Features and patterns: A "fox and hedgehog" model for voice recognition

Taking an idea from the essay of Isaiah Berlin (1953) on Archilochus' fable about a fox and a hedgehog (Fig. 2), we have proposed a model of voice perception that suggests voices can be recognized by varying applications of featural and pattern recognition processes. In the fable, the fox knows many little things while the hedgehog knows one big thing. There are many versions of the bipolarity expressed in this adage: empiricism contrasted with rationalism, Aristotle meets Plato, behaviorism compared with the sweeping ideologies of cognitive science, agility of thought versus persistence (Gould, 2003). In our model of voice perception, the aphorism is meant to represent the interplay between features and patterns in the speaker-listener interface. Some voices and some voice perception tasks draw more heavily on features (many little things), while other voices and other tasks utilize pattern recognition abilities more heavily. This counterpoint helps elucidate the respective roles of unfamiliar and familiar voices, in that featural elements figure importantly in the discrimination of unfamiliar voices (in the sense of matching to generalized templates), while overall pattern recognition predominates for familiar voices (in accessing unique auditory percepts).

Measuring voice quality

We have argued thus far that humans are good at familiar voice recognition because we have inherited this ability through our evolutionary past, and that familiar voices are best treated as integral patterns. Nevertheless, most approaches to voice quality assessment depend on the use of perceptual or acoustic features for quality, or both—in other words, on approaches that use processing strategies that resemble those we apply to unfamiliar voices, with which we are considerably less adept. For example, many authors have proposed lists of descriptive terms to assess quality, and listeners typically measure quality by indicating the extent to which a voice possesses each feature (Voiers, 1964; Gelfer, 1988; Isshiki *et al.*, 1969; Kempster *et al.*, 2009). This approach (the only one currently available for quantifying quality), replete with redundancies and ambiguities, arises from 2000 years of tradition rather than from theory. Many of the features commonly in use today—for example, harsh, breathy, clear, bright, smooth, weak, shrill, deep, dull, and hoarse—can be traced to Roman writings on oratory (Table 3; Laver, 1981; Austin, 1806). Because assessing voices on such rating scales requires listeners to analyze a vocal pattern into component features, we might expect listeners to have a great deal of difficulty using such quality measurement protocols, and in fact many studies have shown quite low levels of interrater agreement, as predicted (see Kreiman *et al.*, 1993, for review).

Nevertheless, quantifying voice quality is essential to many endeavors, including studying the efficacy of treatments for voice disorders or the acceptability of speech synthesis efforts. This leaves us with the following problem: How do we quantify an unanalyzable pattern? One solution under investigation (Gerratt and Kreiman, 2001; Kreiman *et al.*, 2007) is the use of an analysis-by-synthesis approach in which voices are copied using a voice synthesizer specialized for replicating variations in voice quality. Because the complete voice pattern is copied exactly, the synthesizer parameters explicitly link a range of selected features of the acoustic signal to the overall, integral pattern, and can thus be used validly as *objective* acoustic indices of *subjective* perceptual responses. Because this method allows us to study how listeners manage the interplay between features and patterns, it allows for applicability to both familiar and unfamiliar voices and holds the promise of elucidating their distinctive dynamic processing characteristics.

The larger universe of perceptual judgments

Speakers make judgments regarding physical, psychological and social characteristics from voice that go well beyond mere speaker identity, and we are only beginning to understand the range of information conveyed and the manner in which such information is extracted and exploited. For example, the emotional and attitudinal nuances conveyed by voice may well number in the thousands; and many animals (including possibly humans) are adept at extracting information related to reproductive fitness from vocal signals (e.g., Hardouin *et al.*, 2009; Charlton *et al.*, 2007; Apicella and Feinberg, 2008). Thoughtful examination of everyday talk reveals an immense set of possible judgments listeners may make (Table 4). This is not an exhaustive list, but is intended to point to the potentially large constellation of characteristics that underlie functional voice perception. It becomes clear that a systematic reductionist approach to the study of voice perception in the face of these many variables is unre-

Table 4: A partial list of vocal characteristics potentially contributing to voice recognition in humans

Source and filter characteristics
F0 mean, variability and range
Noise excitation
Intensity mean, variability and range
Resonances (formants) and anti-resonances (frequencies,variability, bandwidths, relative spacing of formants
Mode of vocal fold vibration
"Modal"
Murmured
Glottal attack
Creaky voice
Breathy voice
Temporal characteristics
Rate
Mean syllable length
Words per minute
Pausing
Phrase length
Rhythm or meter
Vowel and consonant lengths
Stress patterns
Articulatory setting
Oral
Pharyngealized
Lip protrusion
Articulatory characteristics
Degree of hyperarticulation
Degree of slurring
Coarticulation
Vowel reduction
Diphthongization or reduction of diphthongs
Degree of Nasality
Hypernasal/hyponasal
Prosodic line
Predominately falling, rising
Monotonic, hypermelodic
Syllable structure
Broad
Clipped
Slurred
Other
Dialect/accent
Idiolectal variants (of great variety)

alistic; yet dismissal or rejection of all but a few characteristics holds little promise of explaining voice perception. It has become obvious in inspecting the array of potential cues that not all will pertain to the successful perception of a given voice pattern. Instead, some emerge as decisive to perception of a pattern, and most will be irrelevant.

Drawing on the perspective that individual voice patterns are singular and unique, we propose a model of voice perception that allows for interplay between characteristics or features and the signature voice pattern. Our model is based in the interactivity of voices and listeners in all of voice perception, and takes into account three continua—the relative contributions of feature and pattern recognition processes to recognition or perception of different sorts of voice patterns; differences in the neurological and psychological status of familiar and unfamiliar voices; and left versus right cerebral hemisphere processing and the contributions of subcortical systems in the brain. Perception of the myriad vocal characteristics communicating physical and personality cues, mood, emotion, attitude, background and so on is likely to differ significantly with the relationship of the voice to the listener—that is, its status as familiar or unfamiliar. While deconstruction of neutral voice samples will yield fascinating details about acoustic structure, it is taking on the challenge of the talker-listener interaction with a personally familiar voice pattern and its complex indices of information that will lead to fruitful studies of this immense natural endowment.

Acknowledgments

Some of the research described in this paper was supported by grant DC01797 from the National Institute on Deafness and Other Communication Disorders. Software used in analysis-by-synthesis can be freely downloaded from www.surgery.medsch.ucla.edu/glottalaffairs/software.htm.**AT**

References

Apicella, C. L. and Feinberg, D. R. (**2008**). "Voice pitch alters mate-choice-relevant perception in hunter-gatherers," Proc. Biol. Sci. **276**, 1077–1082.

Austin, G. (**1806**) *Chironomia* (Cadell and Davies, London). Reprinted by Southern Illinois University Press, Carbondale, IL, 1996.

Bahrick, H., Bahrick, P., and Wittlinger, R. (**1975**). "Fifty years of memory for names and faces: A cross-sectional approach," J. Experimental Psychol.: General **104**, 54–75.

Bee, M. A., and Gerhardt, H. C. (**2002**). "Individual voice recognition in a territorial frog (*Rana catesbeiana*)," Proc. Royal Society of London B: Biological Sci. 269, 1443–1448.

Belin, P., and Grosbras, M. H. (**2010**). "Before speech: Cerebral voice processing in infants," Neuron **65**, 733–735.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (**2000**). "Voice-selective areas in human auditory cortex," Nature **403**, 309–312.

Berg, K. S., Delgado, S., Okawa, R., Beissinger, S. R., and Bradbury, J. W. (**2011**). "Contact calls are used for individual mate recognition in free-ranging green-rumped parrotlets, *Forpus passerinus*," Animal Behaviour 81, 241–248.

Berlin, I. (**1953; 1994**). *The Hedgehog and the Fox: An Essay on Tolstoy's View of History* (Weidenfeld & Nicolson, London). Reprinted in Russian Thinkers (Penguin, Oxford).

Bestelmeyer, P., Rouger, J., DeBruine, L. M., and Belin, P. (**2010**). "Auditory adaptation in vocal affect perception," Cognition **117**, 217–223.

Bricker, P. D., and. and Pruzansky, S. (**1976**). "Speaker recognition," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass, (Academic, New York), pp. 295–326.

Burke, E. J., and Murphy, C. G. (**2007**). "How female barking tree frogs, *Hyla gratios*a, use multiple call characteristics to select a mate," Animal Behaviour **74**, 1463–1472.

Charlton, B. D., Reby, D., and McComb, K. (**2007**). "Female perception of size-related formant shifts in red deer, *Cervus elaphus*," Animal Behaviour **74**, 707–714.

Charrier, I., Mathevon, N., and Jouventin, P. (**2001**). "Mother's voice recognition by seal pups," Nature **412**, 873.

Cheney, D. L., and Seyfarth, R. M. (**1980**). "Vocal recognition in free-ranging vervet monkeys," Animal Behaviour **28**, 362–367.

Cheney, D. L., and Seyfarth, R. M. (**1999**). "Recognition of other individuals' social relationships by female baboons," Animal Behaviour **58**, 67–75.

Fenn, K. M., Shintel, H., Atkins, A. S., Skipper, J. I., Bond, V. C., and Nusbaum, H. C. (**2011**). "When less is heard than meets the ear: Change deafness in a telephone conversation," Quarterly J. Experimental Psychol. **64**, 1442–1456.

Fuchs, T., Iacobucci, P., MacKinnon, K. M., and Panksepp, J. (**2010**). "Infant-mother recognition in a social rodent (*Octodon degust*)," J. Comparative Psychol. **124**, 166–175.

Gainotti, G. (**2011**). "What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition," Neuropsychologia **49**, 2273–2282.

Gelfer, M. P. (**1988**). "Perceptual attributes of voice: Development and use of rating scales," J. Voice **2**, 320–326.

Gerratt, B. R., and Kreiman, J. (**2001**). "Measuring vocal quality with speech synthesis," J. Acoust. Soc. Am. **110**, 2560–2566.

Gerratt, B. R., Kreiman, J., Antoñanzas-Barroso, N., and Berke, G. (**1993**). "Comparing internal and external standards in voice quality judgments," J. Speech and Hearing Res. **36**, 14–20.

Goldman, J. A., Phillips, D. P., and Fentress, J. C. (**1995**). "An acoustic basis for maternal recognition in timber wolves (*Canis lupus*)?," J. Acoust. Soc. Am. **97**, 1970–1973.

Gould, S. J. (**2003**). *The Hedgehog, the Fox, and the Magister's Pox: Mending the Gap Between Science and the Humanities* (Harmony, New York).

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (**2010**). "The developmental origins of voice processing in the human brain," Neuron **65**, 852–858.

Hardouin, L. A., Bretagnolle, V., Tabel, P., Bavoux, C., Burneleau, G., and Reby, D. (**2009**). "Acoustic cues to reproductive success in male owl hoots," Animal Behaviour **78**, 907–913.

Insley, S. J. (**2001**). "Mother-offspring vocal recognition in northern fur seals is mutual but asymmetrical," Animal Behaviour **61**, 129–137.

Isshiki, N., Okamura, H., Tanabe, M., and Morimoto, M. (**1969**). "Differential diagnosis of hoarseness," Folia Phoniatrica **21**, 9–19.

Jouventin, P., and Aubin, T. (**2002**). "Acoustic systems are adapted to breeding ecologies: Individual recognition in nesting penguins," Animal Behaviour **64**, 747–757.

Kempster, G. B., Gerratt, B. R., Verdolini-Abbott, K., Barkmeier-Kraemer, J. M., and Hillman, R. E. (**2009**). "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech Language Pathol. **18**, 124–132.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., and Wang, Z. (**2003**). "Effects of experience on fetal voice recognition," Psychol. Sci. **14**, 220–224.

Kreiman, J., and Gerratt, B. R. (**in press**). "Interactions between harmonic and inharmonic components in voice quality perception," to appear in J. Acoust. Soc. Am..

Kreiman, J., Gabelman, B., and Gerratt, B. R. (**2003**). "Perception of vocal tremor," J. Speech, Language, and Hearing Res. **46**, 203–214.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," J. Speech and Hearing Res. **36**, 21–40.

Kreiman, J., Gerratt, B. R., and Ito, M. (**2007**). "When and why listeners disagree in voice quality assessment tasks," J. Acoust. Soc. Am. **122**, 2354–2364.

Kreiman, J., and Sidtis, D. (**2011**). *Foundations of Voice Studies: Interdisciplinary Approaches to Voice Production and Perception* (Wiley-Blackwell, Boston).

Ladefoged, P., and Ladefoged, J. (**1980**). "The ability of listeners to identify voices," UCLA Working Papers in Phonetics **49**, 43–51.

Latinus, M., and Belin, P. (**2011a**). "Anti-voice adaptation suggests prototype-based coding of voice identity," Frontiers in Psychol. **2**, 1–12.

Latinus, M., and Belin, P. (**2011b**). "Human voice perception," Current Biol. **21**, R143–R145.

Laver, J. (**1981**). "The analysis of vocal quality: From the classical period to the 20th century," in *Toward a History of Phonetics*, edited by R. Asher and E. Henderson (Edinburgh University Press, Edinburgh), pp. 79–99.

Li, X., and Pastore , R. E. (**1995**). "Perceptual constancy of a global spectral property: Spectral slope discrimination," J. Acoust. Soc. Am. **98**, 1956–1968.

McComb, K., Moss, C., Sayialel, S., and Baker, L. (**2002**). "Unusually extensive networks of vocal recognition in African elephants," Animal Behaviour **59**, 1103–1109.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Perception and Psychophysics **48**, 169–178.

Neuner, F., and Schweinberger, S. R. (**2000**). "Neuropsychological impairments in the recognition of faces, voices, and personal names," Brain and Cognition **44**, 342–366.

Papcun, G., Kreiman, J., and Davis, A. (**1989**). "Long-term memory for unfamiliar voices," J. Acoust. Soc. Am. **85**, 913–925.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (**2008**). "A voice region in the monkey brain," Nature Neurosci. **11**, 367–374.

Petkov, C. I., Logothetis, N. K., and Obleser, J. (**2009**). "Where are the human speech and voice regions, and do other animals have anything like them?," The Neuroscientist **15**, 419–429.

Pitcher, B. J., Harcourt, R. G., and Charrier, I. (**2010**). "Rapid onset of maternal vocal recognition in a colonially breeding animal, the Australian sea lion," PloS ONE 5, e12195.

Scherrer, J. A., and Wilkinson, G. S. (**1993**). "Evening bat isolation calls provide evidence for heritable signatures," Animal Behaviour **46**, 847–860.

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., Robertson, D. M., Simpson, A. P., and Zäske, R. (**2008**). "Auditory adaptation in voice perception," Current Biol. **18**, 684–688.

Schweinberger, S. R., Herholz, A., and Sommer, W. (**1997a**). "Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues," J. Speech, Language, and Hearing Res. **40**, 453–463.

Schweinberger, S. R., Herholz, A., and Stief, V. (**1997b**). "Auditory long-term memory: Repetition priming of voice recognition," Quarterly J. Experimental Psychol. Section A—Human Experimental Psychol. **50**, 498–517.

Sebastian, R. J., and Ryan, E. B. (**1985**). "Speech cues and social evaluation: Markers of ethnicity, social class, and age," in H. Giles and R. N. St. Clair (Eds.), *Recent Advances in Language, Communication, and Social Psychology* (Erlbaum, London), pp. 112–143.

Sebe, F., Duboscq, J., Aubin, T., Ligout, S., and Poindron, P. (**2010**). "Early vocal recognition of mother by lambs: Contribution of low- and high-frequency vocalizations," Animal Behaviour **79**, 1055–1066.

Sidtis, D., and Kreiman, J. (in press, **2011**). "In the beginning was the familiar voice: Personally familiar voices in the evolutionary and contemporary biology of communication," to appear in J. Integrative Psychol. and Behavioral Sci..

Simmons, A. M. (**2004**). "Call recognition in the bullfrog, *Rana catesbeiana*: Generalization along the duration continuum," J. Acoust. Soc. Am. **115**, 1345–1355.

Torriani, M. V. G., Vannoni, E., and McElligott, A. G. (**2006**). "Mother-young recognition in an ungulate hider species: A unidirectional process," Am. Naturalist **168**, 412–420.

Van Lancker, D., and Kreiman, J. (**1987**). "Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities," Neuropsychologia **25**, 829–834.

Van Lancker, D., Kreiman, J., and Cummings, J. (**1989**). "Voice perception deficits: Neuroanatomic correlates of phonagnosia," J. Clinical and Experimental Neuropsychol. **11**, 665–674.

Van Lancker, D., Kreiman, J., and Wickens, T. D. (**1985**). "Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices," J. Phonetics **13**, 39–52.

Vitousek, M. N., Adelman, J. S., Gregory, N. C., and St Clair, J. J. H. (**2007**). "Heterospecific alarm call recognition in a non-vocal reptile," Biological Lett. **3**, 632–634.

Voiers, W. D. (**1964**). "Perceptual bases of speaker identity," J. Acoust. Soc. Am. **36**, 1065–1073.

Voigt-Heucke, S. L., Taborsky, M., and Dechmann, D. K. N. (**2010**). "A dual function of echolocation: Bats use echolocation calls to identify familiar and unfamiliar individuals," Animal Behaviour **80**, 59–67.

Wolskia, T. R., Houpta, K. A. and Aronson, R. (**1980**). "The role of the senses in mare–foal recognition," Applied Animal Ethology **6**, 121–138.

Zäske, R. and Schweinberger, S. R. (**2011**). "You are only as old as you sound: Auditory aftereffects in vocal age perception," Hearing Res. **282**, 283–288.

Zäske, R., Schweinberger, S. R., and Kawahara, H. (**2010**). "Voice aftereffects of adaptation to speaker identity," Hearing Research **268**, 38–45.

Diana Sidtis (formerly Van Lancker) received an MA in English Language from the University of Chicago and a PhD in Linguistics from Brown University, followed by an National Institutes of Health postdoctoral fellowship at Northwestern University and clinical certification in Speech Pathology from California State University at Los Angeles. She is the former chair and currently Professor of Communicative Sciences and Disorders at New York University. She performs research in neurolinguistics, motor speech disorders and voice at the Nathan Kline Institute for Psychiatric Research.

Jody Kreiman received her PhD in Linguistics from the University of Chicago. She is currently Professor of Head and Neck Surgery at the University of California, Los Angeles (UCLA), and is also affiliated with UCLA's Departments of Communication Studies and Biomedical Engineering. Her research focuses on perception of voice quality, on the relationships between voice production and acoustics, and on the uses of voice quality in natural languages. She is a Fellow of the Acoustical Society of America and is the Editor for Speech of the *Journal of Speech, Language, and Hearing Research*.

Only custom made Quiet Curtains provide laboratory tested (WEAL) metrics that allow you to recommend our products with confidence. **Sound Blocking STC Quiet Curtains use specialized, proprietary vinyl linings rated 15, 17 and 20 STC.** These linings are sewn into beautiful curtains and drapes for use as window and wall treatments, room opening covers and room dividers.

Sound Absorbing Acoustic Quiet Curtains provide tested NRC values from .4 NRC to 1.0 NRC. Years of testing various fabrics and constructions allow us to "dial in" precise values for your client's requirements. From performing arts venues to high tech laboratories, our Acoustic Quiet Curtains and draperies are proving their value.
We also make combination curtain systems with useful STC and NRC values.

Here's what experts say to us:
"Complete Soundproofing has the best, well-engineered sound blocking and sound absorbing curtain systems in the United States… not only do you understand the complexity of acoustical curtain materials, but you have taken the painstaking diligence to test your products and correctly apply their physical properties to your remarkable curtains. I truly feel that this effort has made my job easier by allowing me the opportunity to peruse your ever-growing product line for satisfying a variety of my client's particular needs"

Michael Burrill, INCE
Director / Senior Acoustical Scientist, ARCADIS U.S., Inc

QUIETCURTAINS

www.QuietCurtains.com *858.272.3615*

IMITATION IN SPEECH

Molly Babel
Department of Linguistics
University of British Columbia
Vancouver, British Columbia V6T 1Z4, Canada

> *"Social preferences and liking modulate the process of spontaneous phonetic imitation."*

Why do we sound the way we do? As people learn to speak, they acquire the language and dialect spoken around them. Sentence structure, word choice, and pronunciation are all determined by the patterns used in the ambient language to which we are exposed. Having grown up in Minnesota, I did not learn how to speak with a British accent, but with a Minnesotan one. As a language-learning child, the language input I received determined the general shape of my language output. This article focuses on the spontaneous or natural imitation of speech acoustics. In this article I use the terms imitation, convergence, and accommodation interchangeably. I use all of these terms to describe the unintentional process by which exposure to a speech stimulus causes an observer to display characteristics of the stimulus in their own productions.

This phenomenon of imitating the input is not limited to the acoustic signal that we use to transmit language. Let's begin with a straightforward example from the literature on syntactic priming and word order of how recent linguistic exposure modifies our subsequent speech behavior. Imagine a picture of a man holding a cake and facing a woman. The orientation of the image suggests the man intends to pass the cake to the woman. Participants who have been exposed to the sentence *The boy gave the toy to the teacher* prior to viewing this image are more likely to describe the cake picture as *The man gave the cake to the woman* as opposed to *The man gave the woman the cake*. The second description is a completely grammatical utterance that accurately conveys what is going on in the image, but having been previously exposed to the construction *give X to Y* biases the future use of that construction over *give Y X* (Bock, 1986). Bock's seminal finding reveals quite convincingly that what we say is highly influenced by what we have just heard.

The notion that children learn the speech variety to which they are exposed seems intuitive, but the situation becomes a little more complicated when we shift our attention to the acquisition patterns of adults. What happens when adults who have already acquired a particular speech variety move to a new dialect area? As a young adult, I moved to California and, with time, my speech lost many of its original Minnesotan features. To native Californian ears, I might never have sounded truly Californian, but I eventually sounded much less Minnesotan. Several recent studies have documented this personal anecdote on a larger scale (Evans and Iverson, 2006; Munro *et al.*, 1999). Interestingly, the Minnesotan features of my speech return when I am interacting with old friends and family who have retained our native dialect. This indicates that the acquisition of a new dialect or the adoption of new speech features serve as an update to and expansion upon my linguistic system, as opposed to wholly replacing a previous system. The fact that we do imitate the ambient language tells us that our linguistic categories are malleable and easily influenced by new information.

Inherent variability in speech production

When one considers how speech is produced, it becomes apparent that physiological and anatomical variation across talkers will inevitably be reflected in the spectral characteristics of speech sounds. Let's start with the sex of the talker. Men's voices typically pattern together in having lower pitch and lower resonant frequencies than women's voices. These differences are due in part to sexual dimorphism: men's vocal tracts and vocal folds are generally larger than women's. Age is another key factor in cross-talker variability. Aging is accompanied by various physiological changes. For example, the extrinsic muscles that support the larynx become slack with age, and the mucosal tissue covering the vocal folds loses its elasticity. These changes lead to alterations in voice quality, along with lower pitch and lower resonant frequencies for a talker's voice. These factors highlight the fact that some of the variation in speech is due to anatomical and physiological age- and sex-based variation.

That being said, the extent to which physiological factors determine speech characteristics is frequently oversold. A closer inspection of large datasets reveals that physiological variation does not account for all of the observed differences between groups divided by gender or age. While it is true that classic studies describing vocalic resonant frequencies of men and women show that women produce higher resonant frequencies than men (Peterson and Barney, 1952), a recent examination of gender differences across languages illustrated that gender-based differences in vowel production vary across languages, even when population height is controlled (Johnson, 2006). Such a finding indicates that some of the gender-based differences in resonant frequencies are the result of learned social norms. In addition, despite the fact that many significant anatomical differences do not emerge between males and females until puberty, children acquire gender-specific speech patterns starting in toddlerhood (Sachs *et al.*, 1973; Perry *et al.*, 2001). This suggests a strong socio-cultural component to language production.

Armed with this information, we can hedge a response to the question of why we sound the way we do by acknowledging that a portion of one's speech acoustics is determined by the size and shape of the vocal tract. There is clearly more at issue, however. As mentioned above, we acquire the speech

SPEECH COMMUNICATION 101

You say tomato, I say tomato

The types of changes to which I refer in this paper involve sometimes subtle, sometimes not-so-subtle, changes in the pronunciation of particular sounds. It is important to familiarize ourselves with the ways in which linguists and speech scientists talk about speech sounds. Speech production boils down to a manipulation of the airstream. For example, say the word *tomato*; do this with your hand in front of your mouth and you will have a tactile impression of this airstream manipulation, in addition to the auditory one. Figure 1 presents a spectrogram and waveform from two speakers' pronunciations of *tomato*. My production is on the left, and a male's production of this word is on the right. We return to some key differences in female and male productions later.

Tomato, like all words, is made up of a series of speech sounds. To produce the word *tomato*, your tongue tip moves up and makes contact in the region behind your front teeth, sometimes making contact with your teeth themselves, to make the /t/. If you put your hand in front of your mouth, you will feel a rather strong puff of air as you release the /t/; this is called *aspiration*. From the /t/, your mouth changes its configuration seamlessly as it moves towards a more neutral configuration for the initial vowel, which is a shortened, indistinct schwa-like vowel. Then, to produce an /m/, you close your lips and open the passageway to your nasal cavity, allowing the air to flow through your nasal cavity and sinuses on the way to the open atmosphere. Following the /m/, the vowel you produce will vary considerably depending on the variety of English you speak. Most speakers from North America will produce the vowel sound which also occurs in *bake* and *cake*: /e͡ɪ/. If you speak a variety of British English, you will likely produce this sound as an /ɑ/, which is more similar, although not identical, to the vowel that North American English speakers use in the word *father*. Next we come to a sound that is like /t/, but which is produced much more quickly and with a different movement trajectory in natural speech; this is called a *flap*: /ɾ/. Finally, there is an /o/ sound, which involves rounding your lips a little bit. Note that this last vowel sound is produced and sounds quite different from the first vowel in the word, despite the fact they are both spelled with an "o." It is rather amusing how long it takes to describe *how* to produce a word, compared to how long it takes to simply say the word. Our mouths do some impressive articulatory gymnastics at incredible speed in speech production, and we do not even give it much, if any, thought.

These articulatory movements modify the airstream, making constrictions in the oral cavity to varying degrees. In making a /t/, there is a complete constriction to the point where air is trapped inside the oral cavity. To produce a vowel sound, the oral cavity is left relatively unconstricted, but the oral cavity and tongue are shaped in particular ways. These constrictions and configurations result in particular acoustic consequences when the air column in your vocal tract is excited. You can move your articulators to your heart's content, but without the excitation of the air column in your vocal tract, no sound will be emitted.

A crash course on the acoustics of speech production

The acoustics of the speech signal are determined by two main factors: the sound source and the filter through which that sound passes. The role of the filter is to modify the spectral shape of what was produced by the sound source. In the production of voiced sounds with a relatively open vocal tract—sounds like vowels, /r/, and /l/—the

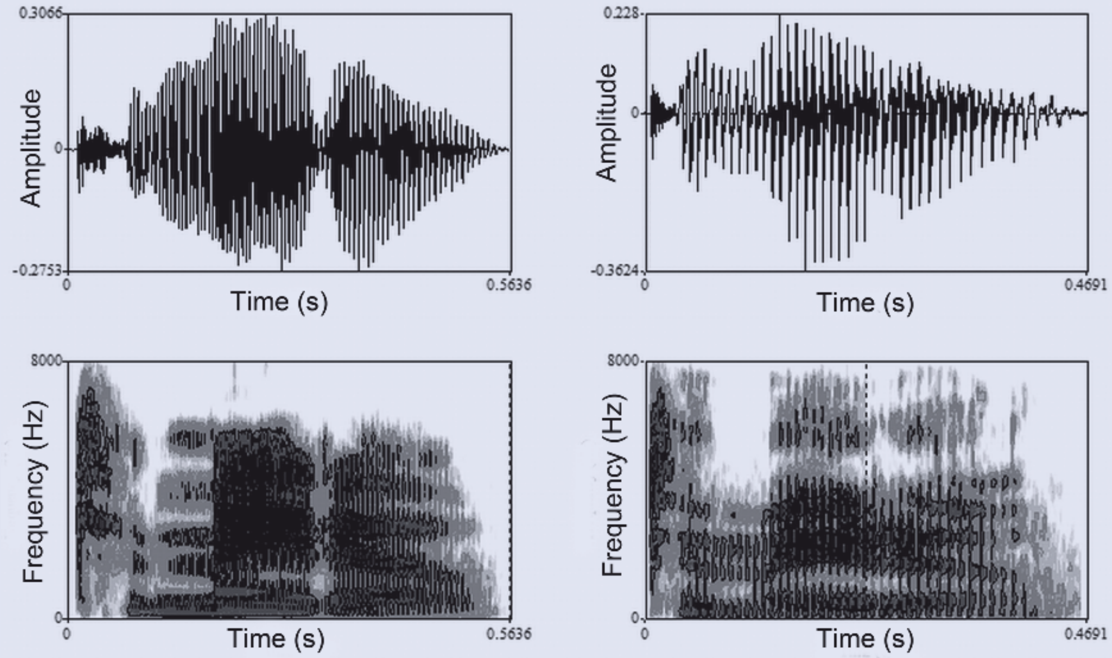


Fig. 1. A waveform (top) and spectrogram (bottom) of the word "tomato" produced by a female speaker (left) and a male speaker (right) of American English.

vibration of the vocal folds serves as the sound source. The supraglottal cavity, which is the part of the vocal tract that extends above the vocal folds, is the filter which shapes the sound generated at the vocal folds. Differences in the size and shape of the vocal folds contribute to inter-speaker differences in pitch and voice quality. Larger vocal folds with more mass will vibrate more slowly, producing a voice that is lower in pitch than that of a talker with smaller vocal folds. The oscillation of the vocal folds provides the fundamental frequency, which listeners perceive as the pitch of the voice, and harmonics, which occur at multiples of the fundamental frequency. "Voice quality" refers to the variations in the sound of a talker's voice, ranging, for example, from breathy to modal to creaky. It is determined largely by the relative speed and duration of vocal fold closure in the course of vibration, and by whether full vocal-fold closure is achieved. (The vibration of the vocal folds is a complex process and we are glossing over many details here.) The size and morphology of the supraglottal cavity determines the resonant properties of the filter. The larger this oral cavity, the lower the resonant frequencies produced by the vocal tract. Manipulating the configuration of the vocal tract modifies the resonant frequencies it produces. These resonant frequencies provide valuable information about a speaker's vowels within the acous-tic-phonetic space of human vocalizations.

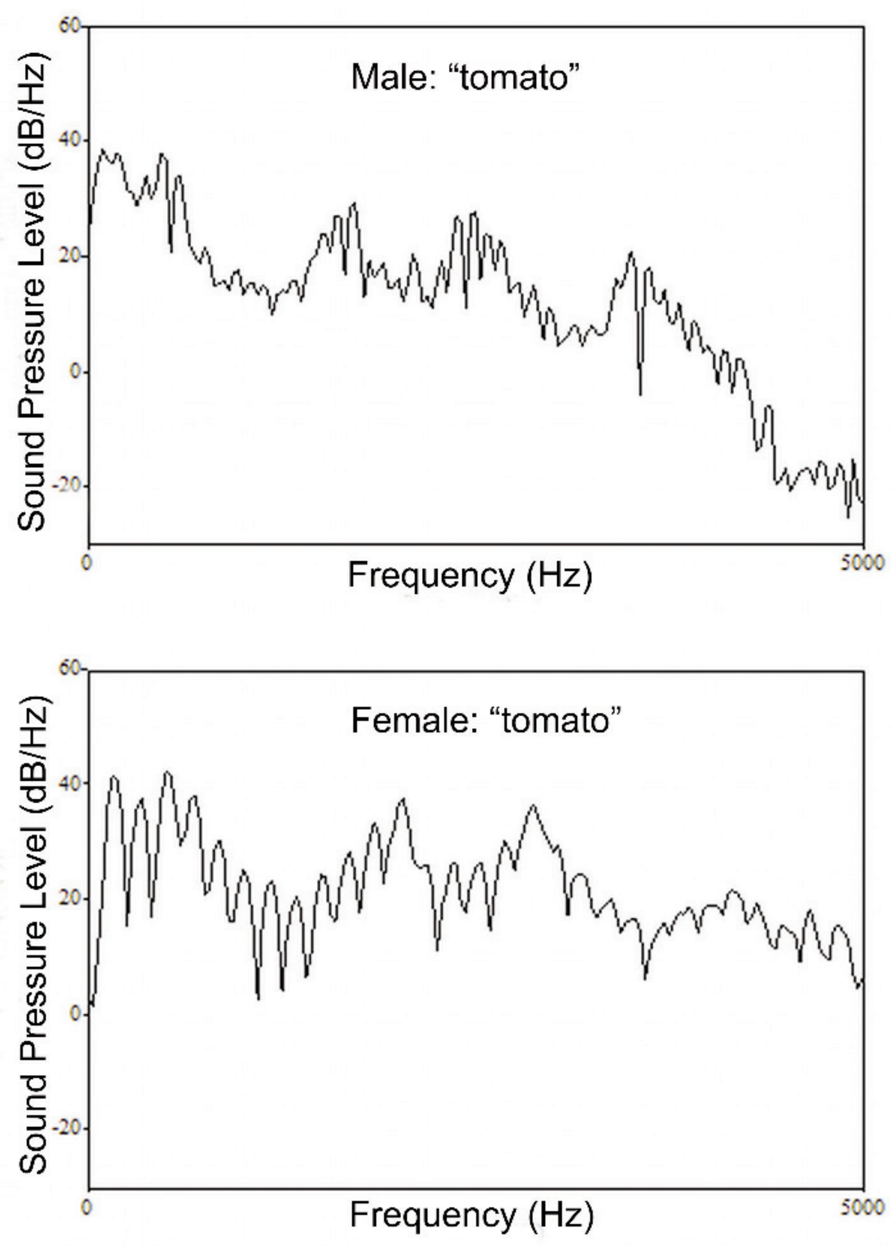


Fig. 2. Power spectra from the first third of the final vowel in the word tomato from a male (top) and a female (bottom) speaker.

Power spectra made from the first third of the final vowel in *tomato* from a male (top) and female (bottom) speaker are shown in Fig. 2. There are several differences between the female and male productions. Let's first turn our attention to the lowest frequency component of the spectra. This is the fundamental frequency. The fundamental frequency is lower in the male's voice than in that of the female. The components going up in frequency are the harmonics. The fundamental and the harmonics are produced by the vibrating vocal folds. Note also that the high amplitude resonance peaks in the female-produced spectrum are generally higher in frequency than the high amplitude resonance peaks in the male-produced spectrum. One intuitively expects to find these predictable differences in speech produced by males and females, due to the relative differences of their vocal tract sizes. Differences in filter shape and source vibrations, however, would arise even when comparing spectra for two female speakers, or two productions from the same speaker. The number of potential acoustic parameters is limitless, and speakers do not have complete control over all of these parameters. Speech, simply, is immensely variable.

characteristics of those around us, so let's consider for a moment regional and dialect variation in speech production. Imagine modeling the pronunciation variants of the vowel /u/, as in the word *food* or *dude*, for which there is considerable variation in regional vernacular production. In Minnesota, this vowel is produced with the tongue in a very high and very back position; the lips are also considerably rounded. These articulatory movements cause this vowel to have a very low second resonant frequency. Now, picture to yourself a Californian surfer saying *dude*. In this stereotypical version, production is quite different from the Minnesotan *dude*. This type of /u/ pronunciation has an extremely high second resonant frequency. Minnesotans and Californians will not globally produce these vowels in exactly these caricatured ways; this increase in the second resonant frequency of /u/ is part of a sound-change-in-progress, leaving individuals in all communities at different stages of the change.

Talk show hosts imitating their guests

The acquisition of regional variation can be simply considered part of what we acquire based on what we hear around us. However, language use does not only vary according to region, and we also find systematic variation based around other macro-sociological categories like class and ethnicity. Our use of language does not reflect a monochromatic mirroring of what we acquired as children, but rather a flexible matching process of sorts that is largely influenced by who we are speaking with or what we are talking about. Let's take as an example the speech patterns of two celebrities. Consider first the pronunciation patterns of Ms. Oprah Winfrey from the *Oprah Winfrey Show*, which were analyzed in Hay *et al.* (1999). Under analysis was the degree of monophthongization of /ɑ͡ɪ/ in Ms. Winfrey's speech; that is, did she pronounce a word like *time* as /tʰɑ͡ɪm/ or /tʰam/? Examples of these two variants are shown in Fig. 3. The primary acoustic differences between these two variants relate to the first and second resonant frequencies of the vocal tract. Note, for example, the dynamic trajectories in /tʰɑ͡ɪm/ on the left compared to the more stable resonant frequencies in /tʰam/ on the right. Given Ms. Winfrey's linguistic biography, we might predict her speech would make use of both variants. Ms. Winfrey grew up in rural Mississippi, a region where /ɑ͡ɪ/ monopthongization is common. This is a typical feature of African American English-speaking speech communities as well. Given the racial division in the Southern US during her childhood, we can infer that Ms. Winfrey grew up in a speech community where this monophthongization was common. As an adult, Ms. Winfrey lives and interacts in speech communities where this sort of monophthongization is not common, and where most talkers use a diphthongal pronunciation.

Hay and colleagues demonstrated that the way Ms. Winfrey pronounced this word varied as a function of both how frequently the word was used and the racial identity of her upcoming guest. Words were defined as high frequency if used five or more times in the corpus under study, and low frequency if used fewer than five times. The researchers found words used more frequently were more likely be produced with the monophthong /tʰam/: 30% of the frequent words were monophongized, as opposed to only 14% of the infrequent words. The racial identity of the upcoming guest largely influenced Ms. Winfrey's pronunciation of this vowel as well. Ms. Winfrey was three times more likely to use a monophthongal pronunciation when she was introducing or discussing an upcoming African American guest on her television program than when she was talking about a non-African American guest. We can interpret her behavior as an accommodation process where she uses a particular variable based on the predicted pronunciation patterns of her guests. This is a process of a sort of global speech style imitation or accommodation. The fact that lexical frequency influences the pattern is also important. The pronunciation variability is not wholly determined by social and interpersonal context; language internal factors, such as lexical frequency, also affect how sounds are produced.

A second celebrity example of acoustic imitation takes us to another talk show: *Larry King Live*. On the *Larry King Live* show, Mr. King interviewed a range of guests, including celebrities, politicians, and others. Gregory and Webster

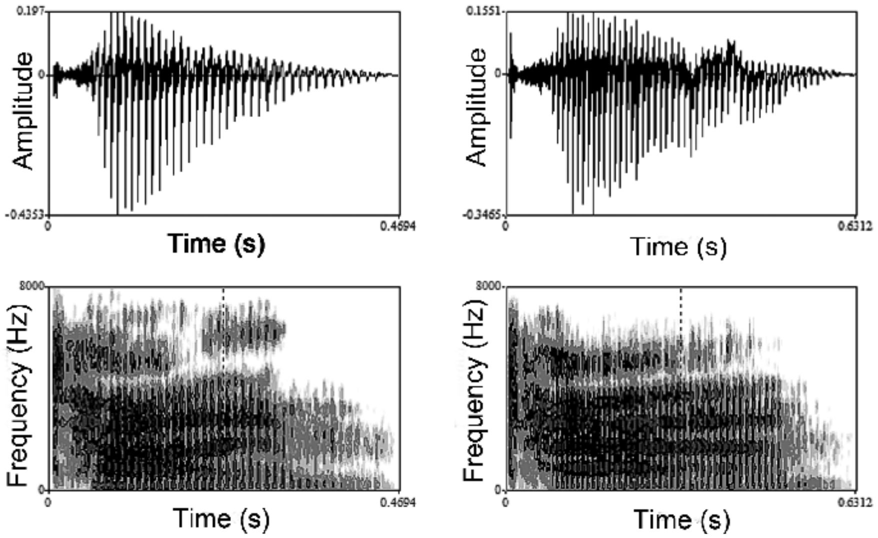


Fig. 3. Pronunciation variants of the word time. The variant on the left is the diphthongal /tʰɑ͡ɪm/ variant and the token on the right is the monophthongal /tʰam/ variant.

(1996) took speech samples from Mr. King and twenty-five of his guests, measuring the long-term average spectra (LTAS) in a low frequency band-pass filter from each speech sample. Using these LTAS measures, correlation coefficients of the actual conversations were compared to those of pseudo-conversation pairings. Actual conversations had significantly higher correlation coefficients, suggesting that interacting talkers accommodated their spectral patterns. The researchers used a measure of LTAS variability and established that Mr. King's interviewees fell into a dominating, low deference group and a high deference group. Mr. King's LTAS measures indicated he took a deferent stance toward the dominating group and a more dominating position in interviews with those in the high deference group. Influential politicians of the time, like former President George H. W. Bush and former President Bill Clinton, fell into the dominating group, triggering a more deferring response from Mr. King. On the other hand, former Vice President Dan Quayle was a member of the high deference group who accommodated more to the speech patterns of Mr. King. (These interviews were taken from broadcasts from April 1992 through July 1993; it may help to keep the social and political context of that era in mind.) Undergraduate students completed surveys to rate the social status of the interviewees and to evaluate how perceived social status affected Mr. King's accommodative behavior toward the interviewees. These subjective student-elicited measures echoed Mr. King's patterns of phonetic accommodation: for example, former President Clinton was at the top of the social status ranking, while former Vice President Quayle was at the bottom. Simply, the researchers found that Mr. King accommodated to the speech patterns of his guests who were of higher social status, while the lower status guests accommodated low frequency spectral characteristics of their speech toward those of Mr. King.

Phonetic imitation in the speech laboratory

Talk show hosts are, of course, not the only ones who imitate and accommodate to interlocutors during spoken language interaction. In recent years, phonetic imitation has been a hot topic in laboratory-based studies of speech. These studies often take one of two forms: a task guised as a sort of guided spontaneous conversation or an auditory naming task. Let's discuss these in turn, starting with the guided spontaneous conversations. These are often map tasks or spot-the-difference tasks, typically involving two participants. They are guided in the sense that they are centered around cooperative activities dictated by the experimenter, but are spontaneous in that the detail of the conversation is freely determined by the natural interaction. In one recent and influential study on phonetic convergence, Pardo (2006) examined phonetic convergence in same-gender dyads involved in jointly completing a map task, where one member of the dyad was the giver of map directions and one was the receiver whose task it was to navigate the described path. Dyads were found to have converged on 62% of the experiment trials. Female dyads were found to converge toward the speaker who was receiving instructions, whereas male dyads patterned oppositely; they converged toward the speech of the talker giving instructions. Pardo concluded that particular social factors dependent on the situational context of a conversation—factors such as gender and the power dynamic of a giving-receiving interaction—determine the direction of phonetic accommodation. Another recent study examined convergence between dyads completing a spot-the-difference task (Kim *et al.*, 2011). The conversational dyads in this study were pairs of native English speakers and native Korean speakers who either did or did not speak the same dialect, and dyads of native and non-native speakers of English. More accommodation was found in the same-dialect dyads than in the different-dialect or the cross-language pairs. Kim and colleagues concluded the process of convergence is facilitated when members of a dyad share a language background, indicating that convergence is easier when the target of the convergent behavior is within an individual's pre-existing phonetic repertoire.

The second design frequently employed in the literature is an auditory naming task. An auditory naming task consists of a listener hearing a model talker produce a word over headphones or loudspeakers, and the listener's task is to identify the word by saying it out loud: that is, to *name* the auditory object. While this method eliminates the natural social context for imitative behaviors to emerge, it offers a more controlled environment for speech researchers to query particular aspects of what might facilitate or inhibit the imitation process. Using this methodology, Goldinger (1998) established that less common words are imitated more than words that are used more frequently. This finding suggests that phonetic imitation may play a role in how we learn about our native languages. For example, the number of times you have heard the word *potato* uttered around you is likely many times more than the number of times you have heard *kohlrabi*. Exposure to variation in how *potato* is pronounced is unlikely to sway your production of the word: in a sense, you confidently know how to say *potato*. Hearing a slightly different pronunciation of the word *kohlrabi*, on the other hand, may cause you to cast your own pronunciation of the word into doubt. On some level, this is an inaccurate way to describe the role of lexical frequency in phonetic imitation; an auditory naming task is too fast-paced to allow for personal reflection on pronunciation insecurities. However, if you have amassed fewer experiences with a particular word, you have fewer memories about how to pronounce it, making a single new exposure to *kohlrabi* all the more prominent a perceptual experience.

Other research using an auditory naming paradigm has explored the interaction of social and phonetic factors in speech imitation. In my own work, I have examined how social biases and preferences moderate phonetic imitation. In an auditory naming task where New Zealanders were presented with an Australian model talker, implicit social biases in New Zealanders' positive or negative views about Australia were measured using an Implicit Association Task, a standard social psychology tool (Babel, 2010). A strong relationship was found: while overall New Zealanders imitated the Australian model, the more positive the New Zealanders' implicit social biases toward Australia, the more they imitated. In another study, I found that the more attractive female participants rated a male model talker, the more they imitated his vowels (Babel, 2012). Social preferences and liking thus modulate the process of spontaneous phonetic imitation. This suggests that the relationship between speech perception and production may be somewhat labile in nature. In a study of gender bias in imitation, Namy and colleagues found that women imitated more than men, but that women's imitative behaviors were focused on a particular male voice used in the experiment (Namy *et al.*, 2002). This indicates there was a particular aspect of this male model's voice which encouraged females to imitate it more than the other voices used in the task.

It is clear that social factors play a role in phonetic imitation. Language-specific internal factors are also involved in the process of imitating speech, just as lexical frequency played a role in Ms. Winfrey's variable pronunciation of /ɑ͡ɪ/ or /a/. Using a modified version of an auditory naming task, Nielsen (2011) presented listeners with a block of model productions that had been digitally modified such that the aspiratory puff of air that accompanies the /pʰ/ sound in *pint* or *pound* was longer than it typically is in natural speech production. (Note that such a puff does not occur in /b/ initial words like *beer* or *baseball*.) Nielsen found exposure to these modified words caused participants to not only increase the duration of the aspiration of /pʰ/, but also to generalize this increase in aspiration to /kʰ/ initial words like *canoe* and *kite*. This indicates that imitation can be abstracted and generalized across one's linguistic system.

Measuring imitation

How do researchers determine whether imitation took place? The complexity of the speech signal makes the method of measuring imitation an important topic. There are two primary ways to gauge or measure phonetic imitation—acoustic or perceptual—each with its positive points and drawbacks. The choice between them is guided by the goals of the study. Let's say a researcher would simply like to demonstrate that the speech signal was imitated in some way. The most common way to accomplish this is to have naive listeners rate perceptual similarity using an AXB task, in which listeners are presented with three tokens in a trial and are asked to judge whether the A or B token is more similar to the X token. The X token is a speech sample from Talker 1, the model talker who was presumably imitated. The A and B tokens would be speech samples from Talker 2; one token would be a "baseline" sample recorded before exposure to or interaction with Talker 1, and the other token would be a sample recorded during or after exposure to or interaction with Talker 1. When listeners consistently choose the post-exposure token as more similar-sounding to the X production of the same word, there is evidence for phonetic imitation. There are two primary virtues to using an AXB similarity task to assess phonetic imitation: (1) it is a holistic measure that allows for imitation of any part of the acoustic signal to contribute to listener judgments of perceived similarity, and (2) if imitative behaviors serve as a seed to sound change, as has been argued (Garrett and Johnson, in press), then these behaviors must be perceptible to listeners.

Oddly enough, this issue of sound change is critical in choosing to measure imitation acoustically as well. It has been argued that phonetic imitation is the means by which sound changes spread through communities. Sound change tends to affect particular aspects of the speech signal. Let's return to the earlier example of the pronunciation of /u/. It was noted above that the pronunciation of this vowel in words like *dude* varies as a function of region. In addition to this regional variation, there is a sound change in progress with this vowel across most varieties of North American English. The sound change is not random: the second resonant frequency of this vowel is becoming higher due to talkers adopting a more fronted tongue position. Acquiring specific acoustic evidence about what was imitated has the potential to address the issue of whether phonetic imitation is the seed by which sound change is spread. In imitating words with /u/, is this increase in the second resonant frequency one of the acoustic features that listeners imitate? If imitation studies demonstrate that what is imitated corresponds to the acoustic phonetic details involved in sound change, then researchers have promising evidence for how sound changes might spread through speech communities. While specific acoustic measures of imitation offer valuable insight into *what* is imitated, perceptual measures of imitation attenuate a potential experimenter bias: in acoustic measures of imitation, the researcher must predict which

aspects of the signal are worth measuring, whereas with a holistic perceptual approach, all features of the signal can be taken into consideration. Recent work comparing a single acoustic measure of imitation and perceptual measures of imitation demonstrated that even when both measures reveal significant effects of imitation, the acoustic and perceptual measures are not correlated (Babel and Bulatov, 2011). This finding underscores the fact while phonetic imitation may interact with individual acoustic-phonetic features on the macro-level in terms of the diffusion of sound change across communities, phonetic imitation on an individual level does not involve singular features. Furthermore, this result indicates that listeners naturally evaluate perceptual similarity from a more holistic perspective. Of course, more holistic acoustic measures are also possible. Recent work, for example, has used mel-frequency cepstral coefficients as a measure of phonetic imitation (Delvaux and Soquet, 2007).

Concluding remarks

The research on phonetic imitation allows several important conclusions about speech communication. First, it highlights an important feature about speech perception and speech production. For imitation to occur, listeners have to perceive a certain amount of subtle acoustic-phonetic detail in the speech signal. This underscores listener sensitivity to the details of the signal. From there, the listener-turned-talker must map the acoustic-phonetic detail onto their own subsequent speech productions. This observation from the imitation literature indicates a relationship between speech perception and production. Importantly, this relationship cannot be a one-to-one mapping because we find many cases where phonetic imitation does not occur (see Vallabha and Tuller (2003) for a clear example). Second, work on phonetic imitation brings us to an interesting conclusion with respect to social cognition and language. The data indicate that, at least in laboratory contexts where social meaning is comparatively void, the default behavior seems to be imitation. Outside of the laboratory, we can imagine that imitation would be crucial for creating and developing social cohesion. The fact that socio-cultural factors moderate even low-level laboratory-based speech behavior strongly suggests that speech production is never without social influence. Lastly, the syntactic analogue of phonetic imitation, known as syntactic alignment or priming (previewed in the introduction), alludes to the important observation that imitative behaviors are pervasive across the language system. This suggests that imitation may serve as a fundamental component in the process of language acquisition and language learning.**AT**

References

Babel, M. (**2012**). "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," J. Phonetics, **40**, 177–189.

Babel, M. (**2010**). "Dialect divergence and convergence in New Zealand English," Language in Society **39**, 437–456.

Babel, M., and Bulatov, D. (**2011**). "The role of fundamental frequency in phonetic accommodation," to appear in Language and Speech, 1–18, doi:10.1177/0023830911417695.

Bock, K. (**1986**). "Syntactic persistence in language production," Cognitive Psychology **18**, 355–387.

Delvaux, V., and Soquet, A. (**2007**). "The influence of ambient speech on adult speech productions through unintentional imitation," Phonetica **64**,145–173.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent change in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Garrett, A. and Johnson, K. (**in press**). "Phonetic bias in sound change," in A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (The Oxford University Press, Oxford).

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychological Rev. **105**, 251–279.

Gregory, S.W., and Webster, S. (**1996**). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions," J. Personality and Social Psychology **70**, 1231–1240.

Hay, J., Jannedy, S., and Mendoza-Denton, N. (**1999**). "Oprah and /ay/: Lexical Frequency, Referee Design and Style," *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, CA.

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phonetics **43**, 485–499.

Kim, M., Horton, W. S., and Bradlow, A. R. (**2011**). "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance," J. Laboratory Phonology **2**, 125–156.

Munro, M. J., Derwing, T. M., and Flege, J. E. (**1999**). "Canadians in Alabama: A perceptual study of dialect acquisition in adults," J. Phonetics **27**, 385–403.

Namy, L. L., Nygaard, L. C., and Sauerteig, D. (**2002**). "Gender differences in vocal accommodation: The role of perception," J. Lang. and Social Psychology **21**, 422–432.

Nielsen, K. (**2011**). "Specificity and abstractness of VOT imitation," J. Phonetics **39**, 132–142.

Pardo, J. S. (**2006**). "On phonetic convergence during conversational interaction," J. Acoust. Soc. Am. **119**, 2382–2393.

Perry, T. L., Ohde, R., and Ashmead, D. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**, 2988–2998.

Peterson, G., and Barney, H. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Sachs, J., Lieberman, P., and Erickson, D. (**1973**). "Anatomical and cultural determinants in male and female speech," in *Language Attitudes*, edited by, R. W. Shuy and R. W. Fasold, 74–83 (Georgetown University Press, Washington, DC).

Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.

Molly Babel is an Assistant Professor in the Department of Linguistics at the University of British Columbia. She teaches courses in phonetics, laboratory phonology, and social factors in language use. Her research examines variation in the speech signal with particular attention to the relationship between speech perception and production and how social cognition modulates speech processing. She maintains multiple lines of research which investigate several sides of these issues including imitation and accommodation, audio-visual and cross-linguistic speech perception, and the effects of talker typicality in speech processing. Babel received her BA in Linguistics, Anthropology, and Spanish from the University of Minnesota, Twin Cities, and her MA and PhD degrees in Linguistics from the University of California, Berkeley.

HUMAN VOICE IN EVOLUTIONARY PERSPECTIVE

Michael J. Owren
Department of Psychology, Georgia State University
Atlanta, Georgia 30302

Introduction

The human voice is a remarkable, multi-faceted instrument that has been studied and discussed by scholars throughout recorded history. Modern scientific study has revealed much about its fundamental properties, such as the physics and physiology of vocal-fold action, the causes and consequences of vocal impairment, and the rich, varied articulatory maneuvers used among the world's many languages. While inquiry has typically been prompted by issues concerning speech communication or vocal performance, work on vocalization in nonhumans is inspiring new questions and insights about the voice from an evolutionary perspective. A major goal in this approach is to understand how and why the human voice has come to have its current, particular form. The premise is that the basic biological forces shaping vocalization in other species have also been important in humans—creating basic commonalities that arguably transcend the many obvious differences that exist between human and nonhuman communication.

> *"…a flurry of recent, exciting studies connecting cues from pitch and resonance to vocalizer fitness and reproductive success."*

This article is intended as an introduction to some of the issues that arise in understanding the voice in evolutionary terms. The *source-filter model* of vocalization will be central throughout, explaining vocal production as a combination of laryngeal energy and vocal-tract resonance. While originally developed in speech science, it is now widely applied to nonhuman vocalization as well. *Indexical cuing* is a second underlying theme, referring to acoustic aspects of the voice and vocal signals that are correlated with important vocalizer characteristics such as sex, identity, age, and emotional state. Both source-filter production and indexical cuing are deeply rooted in the phylogeny of human vocalization, which becomes clear in reviewing our species' mammalian and primate pasts. Commonalities are especially clear in sex and identity cuing, with sex differences in vocal anatomy and acoustics in particular having inspired a flurry of recent, exciting studies connecting cues from pitch and resonance to vocalizer fitness and reproductive success.

Source-filter theory

Understanding the voice in comparative perspective begins by examining the physical characteristics of the vocal tract, important features of which are illustrated for humans and nonhuman primates in Fig. 1. Two critical components can be distinguished. First, the *source energy* of vocalization is derived from laryngeal, vocal-fold vibration driven by air flowing from the lungs (*phonation*), or by creating turbulence in the flow by forcing it through a constriction or onto a surface within the tract. In both cases, this source energy excites cavities located above the larynx, which make up the *supralaryngeal vocal tract*. Resonances of these cavities are referred to as *formants*, and shape the spectral characteristics of the source energy in accordance with their input-output relation. The overall effect is often referred to as *vocal-tract filtering*, and has long been fundamental to understanding human speech production (Chiba and Kajiyama, 1941; Fant, 1960; Stevens, 2000). Over the last two decades, however, this two-component, source-filter approach to vocalization has been applied to an ever-increasing range of nonhuman species as well (Taylor and Reby, 2010).

The process involved in producing a complex, tonal sound is also illustrated in the figure using naturally occurring vocalizations from a human male and a female rhesus monkey (*Macaca mulatta*). Each sound is produced by putting the vocal folds in regular, or quasi-periodic, vibratory motion. As the folds are forced apart and come back together, bursts of air emanate from the *glottis*, which is the opening between the folds. The frequency spectrum of glottal airflow exhibits most energy at the *fundamental frequency* (F_0), or base rate of vibration, with energy at corresponding higher harmonics declining exponentially with increasing frequency. The cavities and tissues of each species' supralaryngeal vocal tract can strongly shape glottal waveform components through resonance and anti-resonance effects, which respectively reinforce or damp energy in corresponding frequency regions. The filtering that results mirrors the sizes and shapes of the vocalizer's supralaryngeal vocal-tract cavities. In an adult human male, a relaxed, "neutral" vocal tract is modeled as a uniform, straight tube closed at the glottal end. It is composed of approximately equal-length pharyngeal and oral cavities, with an overall vocal-tract length of about 17 cm measured from glottis to lips. The characteristic frequency spectra of resulting phonated sounds are marked by 4 to 5 prominent spectral peaks in the 0- to 5-kHz range, each of which reflects a formant. In a rhesus monkey, smaller vocal folds and a much shorter supralaryngeal vocal tract produce higher F_0 values and formant frequencies, respectively.

The pattern formed by these peaks can play a major role in determining the auditory quality of a given vocalization. Corresponding effects are routinely evident in many mammals, taking into account differences in overall vocal-tract length and characteristics of individual supralaryngeal cavities. Due to coincidental resemblance to humans in F_0 and vocal-tract length, for example, the chacma baboon (*Papio cynocephalus ursinus*) "grunt" call bears a remarkable resemblance to an unarticulated, human vowel sound (Owren *et* *al.*, 1997; Rendall *et al.*, 2005). While vocal anatomy can be specialized in particular species, basic principles of production are importantly similar across all mammals. The most important point is that, at least for larger-bodied animals, vocal quality reflects characteristics of both source energy and subsequent vocal-tract filtering. Critical perceptual attributes like pitch, tonality, and other aspects of timbre can all be understood based the combined effects of these two components.

Origins

Reptiles and mammals—Probing the evolutionary history of source-filter production, one might ask if dinosaurs also vocalized using such a system. Films like *Jurassic Park* (1993) and *The Land That Time Forgot* (2009) show them doing exactly that, inasmuch as their sounds are remarkably mammal-like. Such portrayals are only weakly grounded in scientific evidence, however, which consists of little more than finding that certain duck-billed, *Parasaurolophus* dinosaurs had elongated nasal passages forming hollow crests (reviewed by Weishampel, 1997; Isles, 2009). Having ruled out other possible functions, paleontologists have concluded that these crests must have acted as acoustic resonators for vocalization. Unfortunately, there is no evidence as to what the source energy used to excite those cavities might have been.

This intriguing example from dinosaurs does, however, underscore the broader point that, as a group, reptiles have a purely valve-like larynx that cannot also produce sound. Some modern crocodilians, geckos, and tortoises and turtles do vocalize, but these species represent the exception rather than the rule for reptiles as a whole. In contrast, a sound-producing larynx is ubiquitous among the more than 4,500 extant mammal species. Given that all the current major mammalian groups had already emerged by about 93 million years ago (Binenda-Edmonds *et al.*, 2007), laryngeal vocalization must have arisen even earlier—but nonetheless after divergence from the reptile line. When mammals underwent rapid proliferation after the disappearance of dinosaurs about 65 million years ago, they carried that vocalizing larynx along. In fact, one could argue that vocalization is as fundamental to being a mammal as having three middle-ear bones or being homeothermic.

Primates as mammals—In spite of fundamental commonalities, mammals do exhibit significant variability in vocal production as well. Differences can occur in both source and filter components, depending on factors such as overall body-size, hearing range, and niche-specific adaptations (Fitch, 2006, Brudzynski, 2010). Vibration frequencies vary widely across species, ranging from infra- to ultra-sonic. Extra-laryngeal vocal sacs can dramatically amplify or attenuate particular frequency regions, and some species even have a mobile larynx that can dramatically increase effective supralaryngeal tract length during sound production.

There is an additional, cross-species similarity to point out, however, which is that all mammalian larynges can evidently produce a range of phonated sounds, including both harmonically structured and noisy versions. Broadly speaking, this division reflects vocal-fold vibration patterns that are either stable and regular, or unstable and chaotic, respec-

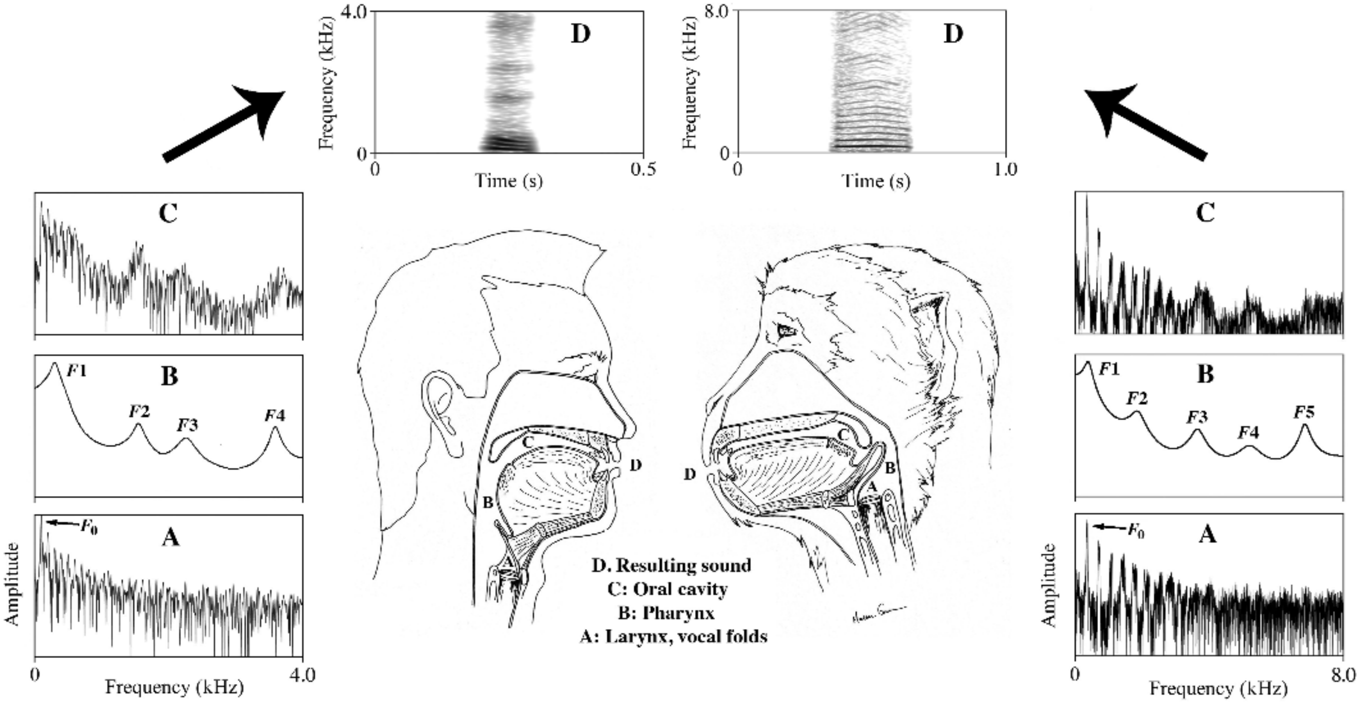


Fig. 1. Schematic views of a human male and a female rhesus monkey vocal tract illustrating the source-filter vocal production process. For both species, panels A and B illustrate source energy frequency spectrum and supralaryngeal transfer function, respectively. Panel C shows the spectrum resulting from combining source and filter, and panel D shows a narrowband spectrogram of the original sound. F_0 refers to the fundamental frequency of the sound, while F1–F5 refer to formants. Rhesus monkeys are significantly smaller relative to humans than indicated here, have significantly higher F_0 and formant frequency values. Note that the rhesus vocalization is shown over a wider frequency range. (Drawings by Michael Graham)

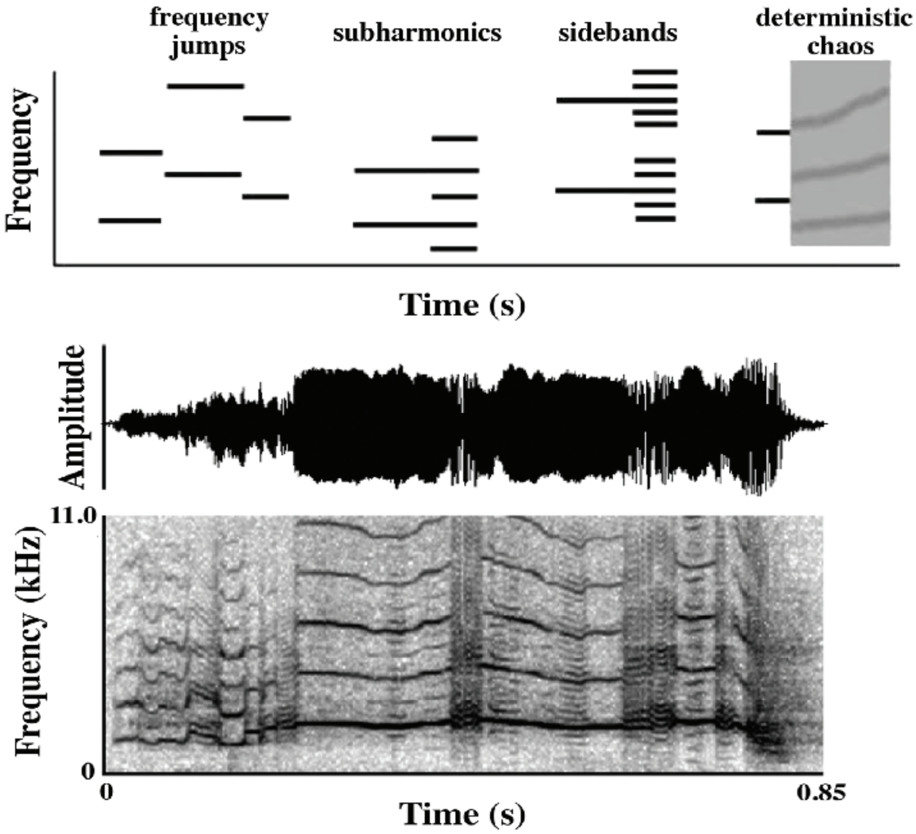


Fig. 2. (top) Schematic depictions of four kinds of "nonlinear phenomena". Each vocalization begins in stable, harmonic form, then undergoes bifurcation to a different vocal-fold vibration regime. (bottom) A rhesus monkey scream that includes each of the nonlinear phenomena illustrated above. While less than a second in duration, the scream includes at least 22 bifurcations among qualitatively distinct vibration regimes.

tively. A key point is that the two vocal folds influence one another when vibrating, and thereby constitute a coupled, nonlinear-dynamical system. All vocalizations are therefore technically "nonlinear" in nature, with the vocal folds exhibiting characteristic vibration regimes that represent attractor states familiar from classic chaos theory (Wilden *et al.*, 1998; Fitch *et al.*, 2002). It is nonetheless useful to differentiate between *harmonic* vocalizations and *nonlinear phenomena*, illustrated in Fig. 2. The former reflect regular, well-synchronized vibration, while the latter include abrupt frequency jumps, perceptually jarring spectral sidebands believed to be produced by laryngeal amplitude-modulation effects, and viscerally grating deterministic chaos (Riede *et al.*, 2004).

While not yet systematically documented, nonlinear phenomena are likely present in every mammalian vocal repertoire, specifically including primates. A critical implication is that the biomechanics of the larynx itself can be primary in determining the qualitatively distinct vocal-types a given species produces (Brown *et al.*, 2003). In other words, whereas the vocalizer's central nervous system determines global "system parameters" such as sub-glottal air pressure and laryngeal muscle tensions, the larynx itself is the ultimate arbiter of vocal-fold behavior. As in other nonlinear systems, the coupled vocal folds show "exquisite sensitivity" to minor changes in global parameters, with even very small changes potentially producing near-instantaneous bifurcation into qualitatively different vibratory regimes and associated acoustics.

Humans as primates—Overall, it is clear that the human voice has ancient phylogenetic roots. Vocal-tract design is fundamentally similar across mammals, including humans, with corresponding operating principles. As in primate and non-primate mammals alike, the human larynx is a nonlinear-dynamical system whose vibration regimes represent attractor states that give rise to a range of qualitatively different source signals. Any such energy is subsequently shaped by supralaryngeal cavities, including when the source is simply turbulence in the airflow. In the absence of species-specific modifications, supralaryngeal filtering effects are expected to be similar in humans and larger-bodied mammals. Humans are also clearly mammal-like in being endowed with a repertoire of highly heritable, emotion-triggered signals such as spontaneous crying and laughter (Owren and Goldstein, 2008). These sounds emerge in recognizable form very early in life, without apparent need for practice or even to first hear the sounds from others (Owren *et al.*, 2011). Infant crying in particular is marked by chaotic vibration (Mende *et al.*, 1990) resembling that observed in nonhuman primate screaming (Tokuda *et al.*, 2002). Spontaneous, emotion-triggered vocalizations remain important even as the child gains increasing volitional control over sound production and begins to speak.

Humans do have their own specializations, of course, including a thick, highly mobile tongue used to flexibly alter supralaryngeal resonances, and an exceptional degree of volitional control over sound production (Owren *et al.*, 2011). Because supralaryngeal filtering is largely static in nonhuman primates (although see Riede and Zuberbühler, 2003), their vocalizations can be characterized as fundamentally "laryngeal" in nature. In other words, vocal quality is primarily determined by the laryngeal vibration regime involved, which is also the case for spontaneous crying and laughter in humans. In contrast, human speech is marked by a relative paucity of source-energy types—essentially, quasi-periodic phonation versus turbulent noise. In other words, production is importantly "supralaryngeal," with the tongue, mandible, and lips used to flexibly and dynamically create the many sounds of each different language.

Human vocal-fold structure and response also show important developmental changes (Schweinfurth and Thibeault, 2008; Hartnick *et al.*, 2005). One evident consequence is that the vibration regimes underlying the psyche-shattering shrieks and screams characteristic of young children become difficult, if not impossible, for adults to produce. Instead, vocal-fold behavior appears to become more stable, centered on regular, synchronized vibration and associated harmonically-structured sounds. In fact, the vocal gymnastics of infants and children would constitute vocal abuse in adults, for whom chronic shouting or screaming can induce vocal-fold nodules and other pathologies (Stemple *et al.*, 2009). Suggestive evidence along these lines is also provided by a recent comparison of tickle-induced laughter in great apes and humans. While all five species produced distinctive-sounding laughter sounds, humans stood out from the others in showing significantly greater regularity in underlying vocal-fold action (Davila Ross *et al.*, 2009). A speculative but logical inference is that human vocal folds show evolutionary modification for more stable response across a range of air pressures and muscle tensions. While arguably losing some flexibility in laryngeal response, adult human voices have become less prone to nonlinear phenomena. That change has created a requisitely higher proportion of regular, well-synchronized phonation, which in turn may have promoted the effectiveness of source-filter-based indexical cuing.

Indexical cuing in the voice

Source-filter theory, laryngeal nonlinearity, and the similarities as well as differences between humans and other mammals create the foundation for understanding vocal indexical cuing. In a sense, all vocalizations must be considered inherently indexical, for instance in simply showing that a vocalizer is present. However, the more important consideration is how indexical cues are affected by the acoustics of a given vocalization. The indexical potency of harmonically structured sounds, in particular, is clearly evident from everyday experience alone. Here, the pitch and timbre of phonated speech allow listeners to immediately discern a talker's sex, identity, approximate age, and other personal characteristics. These capabilities are traceable to inherent differences in vocal-tract characteristics both among age-sex classes—such as adults versus children and males versus females—and among individuals within each group. For example, phonation allows even potentially subtle differences in vocal-fold size, shape, and tissue properties to be revealed in features such as F_0, relative noisiness of the glottal signal, and cycle-to-cycle variation in vibration. Thus, humans tested with male versus female voices require fewer than two waveform cycles—each corresponding to a single opening and closing of the glottis—to hear the difference (Owren *et al.*, 2007). Supralaryngeal filtering also contributes strongly to indexical cuing, even as talkers are dynamically altering the pharyngeal and oral cavities for linguistic purposes. Even brief segments of recorded vowel sounds show that details of formant patterning can provide important potential cues to both sex and individual identity (Bachorowski and Owren, 1999).

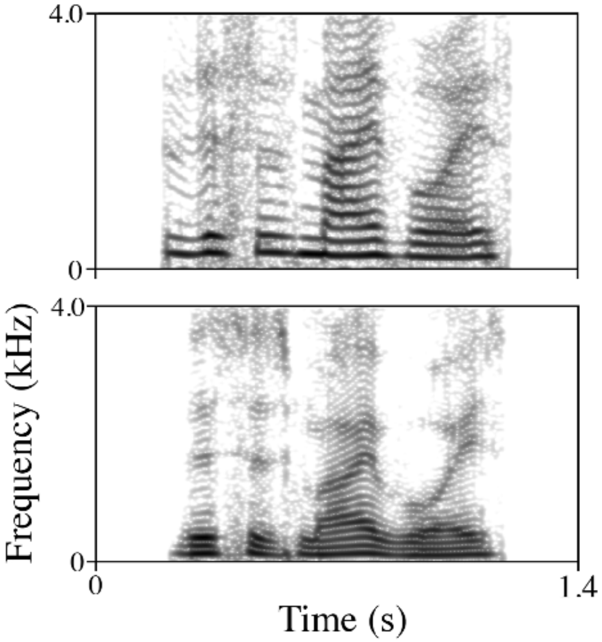


Fig. 3. Narrowband spectrograms of a human female (top) and male (bottom) saying the words "this is my voice." The lower pitch and resonance in male voices makes formants more distinct and easier to measure than in female voices.

However, indexical cuing can be strongly affected by the nature of the source energy involved. As shown in Fig. 3 for male and female speech, for example, supralaryngeal cues become less evident as F_0 increases. This effect occurs because harmonics occurs at integer multiples of F_0 and raising this basic rate of vibration spaces them further apart. The source spectrum thereby becomes more sparsely populated, with less opportunity for supralaryngeal resonances to create a distinct imprint. Another way to understand this outcome is that formants become less well "sampled" by the source signal, giving the listener less to go on in recovering details of frequency, bandwidth, and amplitude. Some formants may not be sampled at all when F_0s become very high. Adding some noisiness to otherwise stable vocal-fold vibration can improve the situation, for instance by "filling out" the source spectrum. That effect occurs in breathy phonation in human talkers, as well as in the noisy, but nonetheless regularly phonated "roars" of red deer (*Cervus elaphus*) and other mammals (Taylor and Reby, 2010).

But too much noisiness becomes a liability. Reducing the source energy of speech to noise alone—as in whispering—makes both phonetic and indexical cuing less effective (Tartter, 1991; Katz and Assman, 2001). Deterministic chaos is nonetheless by far the greatest challenge to supralaryngeal cuing. As a general phenomenon, the occurrence of nonlinearity in a voice has been suggested contribute to individual identity signaling (Fitch *et al.*, 2002). Such events might, for example, occur idiosyncratically in particular vocalizers and thereby become compelling cues to their respective identities. Nonlinear vocal phenomena are by nature unstable, however, and therefore not likely to provide as consistent a substrate for indexical cuing as vocalizer-specific vocal-fold properties or supralaryngeal filtering (Rendall, 1996; Owren and Rendall, 2001). Furthermore, informal examination of a variety of chaos-based screams suggests that virtually no source- or resonance-related indexical cuing occurs in such sounds—no matter what species they are from (see Fig. 4). Empirically, direct comparisons of identity signaling in rhesus monkey and baboon vocalizations have shown that harmonically structured sounds are a markedly better vehicle.

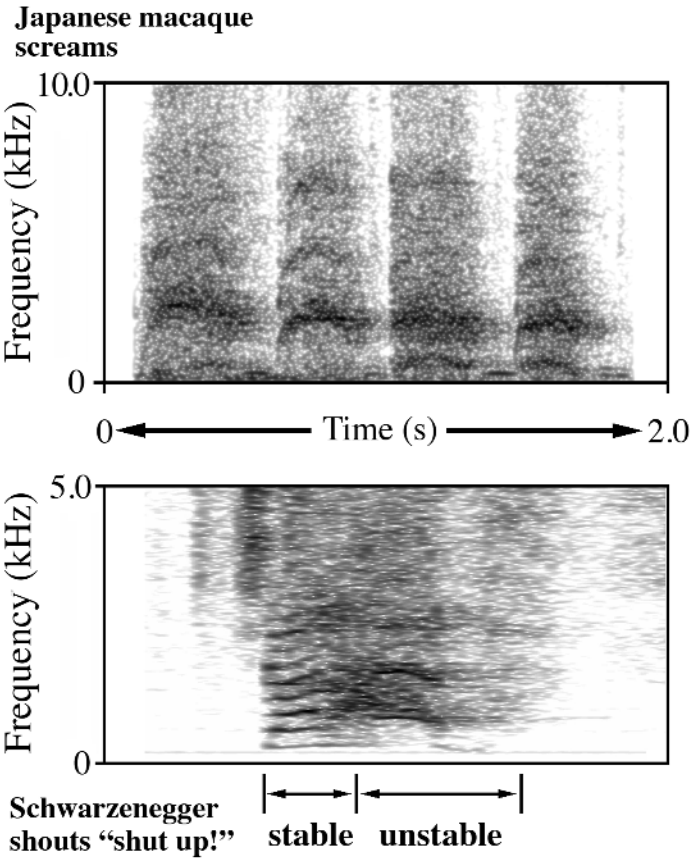


Fig. 4. Narrowband spectrograms of high-arousal, screams produced by an adult female macaque (top), and Arnold Schwarzenegger shouting the words "shut up," as recorded from the movie Kindergarten Cop (1990). The macaque screams deterministic chaos throughout, with no apparent evidence of formant patterning. Schwarzenegger's shout initially shows regular vocal-fold vibration (the "uh" sound from "shut"), but then gives way to unstable, irregular, and likely chaotic action. Schwarzenegger's initially distinctive voice quality is readily apparent in the stable portion, but disappears when the source energy becomes unstable.

Both species have been tested in playback experiments under naturalistic conditions, with adult females hearing either harmonic calls or chaotic screams (Rendall *et al.*, 1996; Rendall *et al.*, 1998; Rendall *et al.*, 2009). These listeners heard sounds from either their own or others' offspring, or from adult females that were either biological kin or non-kin. Outcomes were unequivocal. When listeners heard harmonically structured calls, their responses clearly depended on the caller's relationship to them. However, the subjects showed little or no evidence of differentiating among vocalizers when hearing screams. Naïve human listeners tested with rhesus calls in a lab setting were similarly significantly better at discriminating among individual callers when hearing harmonically versus chaotically structured vocalizations (Owren and Rendall, 2003). As yet, there is no ready explanation for the absence of filtering effects in these screams, a puzzle that begs for further investigation.

Sex differences and sexual selection

Examining the possible impact of sexual selection on male and female voices has become an active and exciting area of research. Sex differences in human vocal characteristics are, of course, so familiar from everyday experience that they are almost taken for granted. However, when working from an evolutionary perspective, noticing such differences almost reflexively triggers questions about their origin and possible function. In general, sexual selection is proposed to occur when individuals compete for access to opposite-sex mates (*intrasex competition*), or compete to be selected as a mate by members of the opposite sex (*mate-choice competition*). In both cases, one sex may acquire distinctive and unique features that need not have direct counterparts in the other. In humans, examples of these kinds of dimorphisms include body-fat distribution, facial morphology, and beard growth (Boyd and Silk, 2011). In such cases, sexual selection is suspected when differences cannot be readily explained as an artifact of more global dimorphisms, such as in body-size. The next step then becomes to show that the exaggerated features found in one sex or the other play a significant role in intrasex competition, mate-choice competition, or both.

Possible effects of body-size on the voice become important in that primate males are, in fact, larger than females in many species, including humans and all four great ape species. Furthermore, male-female differences in F_0 are common without necessarily exceeding overall dimorphism (Mitani and Gros-Louis, 1995; Ey *et al.*, 2007). However, vocal dimorphisms can be disproportionate as well, which is the case for both F_0 and formants in baboons (Rendall *et al.*, 2004). Outcomes for humans are similar, with adult males being approximately 8% taller and 15-20% heavier than females (Puts, 2010). Laryngeal dimorphism is quite disproportionate, with the vibrating segments of the adult vocal folds being about 60% longer in males than in females, which lower speaking F_0 by approximately 50% (Titze, 1994). Dimorphism in vocal tract length is also disproportionate to height, being about 15–20% greater in males (Fant, 1960; Goldstein, 1980).

In humans, vocal-tract development proceeds along similar trajectories in males and females until puberty, at which point boys famously show marked laryngeal growth (Titze, 1994; Harries *et al.*, 1998). Physical changes include lengthening and thickening of the vocal folds, effects triggered by increases in circulating sex steroid levels—particularly testosterone. Both masculinizing and feminizing effects are classically hormone-related, with dimorphism resulting from differential tissue growth in one sex or the other (Dixson, 2009). In the male voice, the process can occur in as little as a year, but can also take up to five years. The larynx also shows a pubertal growth spurt in girls, but much more modestly. The vocal-tract also grows longer during this period, with male puberty being associated with a process of secondary laryngeal descent. This laryngeal lowering thereby lengthens the pharynx, ultimately positioning the male larynx a full vertebra's distance below its female counterpart (Fitch and Giedd, 1999). Overall, then, evidence from both male and female anatomy and vocal acoustics are indicative of sexual selection in human vocal production.

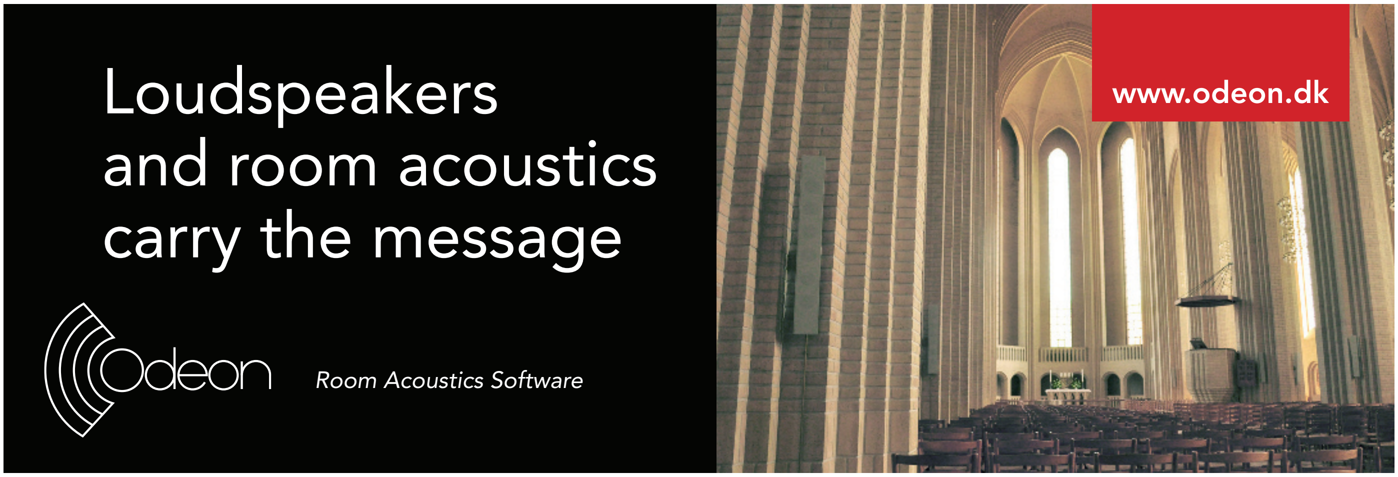
Intrasex competition

Within-sex competition is common among mammals, most frequently between males (Puts, 2010). As a rule, the larger individual wins in male-male contests, with many encounters being resolved before escalation to violence. Vocalizations often play a key role in such cases, with intimidation through vocal signaling of size believed to be a critical factor (Bradbury and Vehrencamp, 2011). Results from a number of nonhuman species are consistent with this view, for instance demonstrating correlations among vocalizer body-size, vocal-tract length, and formant frequencies, as well as listeners' sensitivity to vocalizer resonance cues (Fitch, 1997; Fitch and Fritz, 2005; Harris *et al.*, 2006; Reby and McComb, 2003; Riede and Fitch, 1999). Using signaling to influence contest outcomes necessarily creates selection pressure for exaggeration, however, in this case of apparent size. In one extreme case, red deer males have been shown to lower their larynx more than 30 cm when vocalizing (Fitch and Reby, 2001). Females are indeed affected by resulting resonance cues, but do not show this effect themselves.

If voice-related intrasex competition also occurs in humans, it is thus reasonable to expect that males will be most affected. Male vocal characteristics in particular should be correlated with overall body-size, but vocalizers may also exaggerate those cues. Between age-sex classes, at least, it is clear that key vocal characteristics are significantly correlated with body-size. Both F_0 and formants are lower in adults than in children (Hirano *et al.*, 1983; Hollien *et al.*, 1994), and in adult males than in adult females (Hillenbrand *et al.*, 1995; Rendall *et al.*, 2005). Human listeners are also sensitive to these differences, and use them to identify vocalizers as men, women, or children (Coleman, 1976; Owren *et al.*, 2007). However, correlations between voice and body-size are much weaker within age-sex class—including in adult males. In fact, there may be no relationship between F_0 and body-size in either males or females (Rendall *et al.*, 2007). There is stronger evidence of a reliable correlation between vocal-tract length and body-size, but the degree of correlation is again modest, and not entirely consistent across studies. The picture is also complicated by the fact that human listeners are not very good at judging vocalizer body size (Collins, 2000; van Dommelen and Moxness, 1993; González, 2004). Furthermore, judgments tend to be based on F_0 differences, which is the less-reliable cue (Rendall *et al.*, 2007). Formants do predominate when stimuli are equated for discriminability on the two dimensions (Pisanski and Rendall, 2011), but with the caveat that naturally occurring resonance differences between the sexes are significantly smaller than pitch differences. In other words, equating for discriminability means presenting formant cues that are arguably proportionately larger than the F_0 cues.

Overall, then, results concerning intrasex competition based on body-size signaling are mixed. On the one hand, it is clear that disproportionate sex differences do exist for both F_0 and formants. Furthermore, F_0 cues sway listener judgments for both male and female vocalizers, while it is specifically male versions that are exaggerated. A smaller, but detectable effect is also present for formants, most often in male voices. On the other hand, within-group correlations between vocal characteristics and body-size are uncertain for F_0 and modest for formants. Furthermore, listeners are paradoxically more swayed by vocal pitch, which is almost entirely unreliable.

One possible explanation for these seemingly contradictory outcomes is that reliable body-size cues are unnecessary—vocalizers may instead be capitalizing on the strong, global relationship between physical size and both pitch and resonance that exists in the world at large. In other words, because a strong relationship exists between the size of an object or animal and associated pitch and resonance cues in the world at large, listeners are swayed even by unreliable vocal cues (Rendall *et al.*, 2004). F_0 cues may also be easiest to exaggerate, as the human larynx grows more or less perpendicularly to the body axis, and can protrude from the neck without disturbing other tissue (Fitch, 2000; Fitch and Hauser, 2002). As the pharynx grows along the body axis and oral cavity length is likely constrained by mandible and tooth geometry, vocal-tract length remains more proportional to body-size as a whole. One might even argue that F_0 cues have become exaggerated to the point of unreliability in human males, with formant cues differing only in being affected to a lesser degree. However, that account leaves unexplained why listeners would be differentially sensitive to the less reliable cue. A second, quite different argument is that F_0 cuing is more accurate than hitherto believed. In this view, relying on height and weight differences importantly underestimates male-female differences. Specifically, human males have 60% more lean muscle-mass than females, and 80% greater mus-



cle mass in the arms (Puts, 2010). To test that idea, researchers estimated male "threat potential" by measuring height, bicep size, hand strength, salivary testosterone level, and inherent aggressiveness. Outcomes showed stronger correlations between F_0 and formants, and size, strength, and testosterone level than previously reported for either height or weight (Puts *et al.*, 2011; see also Sell *et al.*, 2010). The resulting argument is that the male voice does provide important, reliable cues to vocalizer competitive capabilities, and that listeners are responding reasonably to those cues.

Mate-choice competition

Recent studies have also addressed the related question of whether vocal characteristics play an important role in mate-choice competition—here expecting both sexes to show such effects. The basic approach has been to ask listeners to rate the relative attractiveness of a variety of male and female voices, with pitch and resonance again being the critical variables. Testing females, it is common to find a preference for masculinized voices—meaning those with lower vocal pitch and resonances (Feinberg, 2008; Jones *et al.*, 2010; Pisanski and Rendall, 2011). From a mate-choice perspective, these characteristics may represent hormone-related ornamentation that has emerged precisely due to being attractive to females (Feinberg, 2008). Although the relationship is modest, male salivary testosterone levels have indeed been found to be inversely correlated with both F_0 and formant frequencies (Dabbs and Mallinger, 1999; Bruckert *et al.*, 2006; Evans *et al.*, 2008). Other evidence supporting an influence of mate-choice selection includes a statistical correlation between these vocal characteristics and both number of children fathered (Apicella *et al.*, 2007) and number of sexual encounters reported (Hodges-Simeon *et al.*, 2011). Finally, listeners are more likely to expect infidelity from males with masculinized voices (O'Connor, 2011), which are also preferred more by females when approaching ovulation than at non-fertile times within the menstrual cycle (Feinberg *et al.*, 2006).

Males show approximately converse preferences, as might be expected. For instance, many studies have revealed a preference for higher-pitched female voices (Apicella and Feinberg, 2009; Pisanski and Rendall, 2011). This effect may be traceable to pitch as a fertility cue, with females being most fertile and having highest speaking F_0 values in early adulthood (Stathopoulos *et al.*, 2011). Males have also been found to prefer voices of females recorded when close to ovulation (Pipitone and Gallup, 2008), a point in the menstrual cycle that is also associated with increased vocal pitch (Bryant and Haselton, 2009)—although not uniquely so (Fischer *et al.*, 2011). Other evidence includes increased pitch among females when believing they are communicating with more masculinized and attractive males (Fraccaro *et al.*, 2011), and women with higher-pitched voices are deemed more likely to exhibit infidelity (O'Connor, 2011). There are again complications, of course, but perhaps fewer than for intrasex competition in voice. For example, not all studies have found more masculine or feminine voices to be the most attractive. In at least one case, mid-range or average voices in the opposite sex have been the most attractive for both male and female listeners (Hughes *et al.*, 2010). The same work reported that all participants tended to speak at lower pitches when interacting with an attractive partner. While consistent with previously reported female preferences, the result is inconsistent with other findings for males. There is also disagreement as to whether F_0 and formants work separately (Hodges-Simeon *et al.*, 2010; Jones *et al.*, 2010) or synergistically (Feinberg *et al.*, 2008; Feinberg *et al.*, 2011).

Conclusions

This review has moved quickly and lightly over a variety of topics, each of which deserves much more thorough treatment. Nonetheless, the evidence covered underscores the fact that the human voice does have a long evolutionary history and has been importantly shaped through shared phylogeny with other species. Vocal-fold action and vocal-tract resonance have emerged as recurring themes, equally applicable to vocal production in humans and nonhuman mammals and creating substantive evolutionary connections between the two. However, it has also become apparent that hominin evolution also brought important changes. Understanding those changes raises questions that comparisons to other primates and mammals alone may not fully address. Yet, combining clues from other species with evidence of novel human vocal characteristics may ultimately prove to be an effective means of shedding further light on hominin evolution overall. Three issues will be briefly followed up in closing, including the evident weakness of correlations between human vocal characteristics and physical features such as overall body-size, possible changes in vocal-fold stability over hominin evolution, and the intertwining of indexical and phonetic cuing in speech—the most unique of human vocalizations.

Sexual selection and the voice—Understanding the role of sexual selection on the voice is a recent undertaking, and progress has been rapid. The overall approach has been validated not only by finding evidence of sexual-selected vocal effects in both sexes, but also by the fact that outcomes are predictably somewhat different in males versus females. However, it is also difficult to avoid the feeling that an important piece of the puzzle is still missing. Correlations between vocal and physical characteristics are too weak, for example, and it is not satisfying to invoke global correlations from the world at large to explain an apparently illusory relationship between pitch and size. Another possibility is that those human mating decisions have become sufficiently complex over evolutionary time that vocal characteristics have lost an important link to physical characteristics that they once had. However, a more compelling explanation may emerge though more substantive recasting of key vocalizer traits as a combination of physical and psychological characteristics, such as threat potential. Overall, understanding sexual selection effects in the human voice has some surprising, but interesting, complexities that may require imagination and re-thinking to untangle.

Vocal-fold stability—The question of whether vocal-fold response characteristics changed over the course of hominin evolution has broader potential implications than one might first imagine. For example, available comparative data indicate that vocal-fold composition can vary across primates and it furthermore appears that developmental modifications known to occur in humans are correlated with changes in acoustic output. Combining these two kinds of information can help illuminate relationships between vocal-fold morphology and how vocalizations were being used, thereby shedding new light on the adaptive changes occurring in hominins. The specific suggestion made here is that vocal action became more stable, especially in adults. This change would be natural to connect to increasing reliance on vowel-like sounds, for instance in association with the evolution of speech. However, greater vocal-fold stability may have emerged earlier to facilitate indexical cuing in the context of increasingly complex hominin social groups and relationships. While not directly connected to the emergence of speech, such changes may have helped set the stage for this development. Detailed knowledge of the relationship between vocal-fold structure and communicative function could be key in resolving such questions.

Indexical and phonetic cuing—A final note concerns the interplay of indexical and phonetic cuing. Historically, speech scientists have struggled to separate the two in seeking to find invariant physical features that distinguish individual phonemes in the face of the acoustical variability occurring across talkers (as well as other factors). The upshot has been that cues to the phonetic content of speech may not exist independently of a given talker's personal, vocal characteristics. This conflation of the phonetic and indexical is understandable in that both flow simultaneously from the same source and filter system during speech. If so, however, then understanding phonetic cuing requires getting a handle on indexical cuing as well—which in turn brings the evolutionary perspective into the picture. In a sense, sexual selection effects in the voice have already had a notable effect on how phonetic features are understood. In the early years of modern acoustic phonetics, researchers focused mainly on adult male talkers, importantly because their speech revealed prominent and easily measured formants. Formant measurement was notably more difficult in speech from adult females or children, so much so that some came to consider the sound-spectrographic technology being used to be inherently "sexist." There is some truth to that charge, particularly as it was later found that female speech is actually the more intelligible. As discussed earlier, finding prominent, well-defined formants in males versus other talkers is straightforwardly due to the forces of sexual selection lengthening their vocal folds and supralaryngeal vocal tracts. One can only wonder if history might have unfolded differently in the study of speech had the role of evolutionary forces on the voice itself been realized from the beginning.**AT**

References

Apicella, C. L., and Feinberg, D. R. (**2009**). "Voice pitch alters mate-choice-relevant perception in hunter–gatherers," Proc. Royal Soc. B **276**, 1077–1082.

Apicella, C. L., Feinberg, D. R., and Marlowe, F. W. (**2007**). "Voice pitch predicts reproductive success in male hunter–gatherers," Biology Lett. **3**, 682–684.

Bachorowski, J.-A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.

Binenda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., and Purvis, A. (**2007**). "The delayed rise of present-day mammals," Nature **446**, 507–512.

Boyd, R., and Silk, J. (**2011**). *How Humans Evolved* (W. W. Norton, New York).

Bradbury, J. W., and Vehrencamp, S. L. (**2011**). *Principles of Animal Communication* (Sinauer Associates, New York).

Brown, C. H., Alipour, F., Berry, D. A., and Montequin, D. (**2003**). "Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*)," J. Acoust. Soc. Am. **113**, 2114–2126.

Bruckert, L., Liénard, J.-S., Lacrois, A., Kreutzer, M., and Leboucher, G. (**2006**). "Women use voice parameters to assess men's characteristics," Proc. Royal Soc. B **273**, 83–89.

Brudzynski, S. (editor) (**2010**), *Handbook of Mammalian Vocalization, Volume 19: An Integrative Neuroscience Approach* (Elsever Science, Oxford, UK).

Bryant, G. A., and Haselton, M. G. (**2009**). "Vocal cues of ovulation in human females," Current Biol. **5**, 12–15.

Chiba, T., and Kajiyama, J. (**1941**). *The Vowel: Its Nature and Structure* (Tokyo: Tokyo-Kaiseikan).

Coleman, R. O. (**1976**). "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness of the voice," J. Speech and Hearing Res. **19**, 168–180.

Collins S. (**2000**). "Men's voices and women's choices," Animal Behaviour **60**, 773–780.

Dabbs, Jr. J. M., and Mallinger, A. (**1999**). "High testosterone levels predict low voice pitch among men," Personality and Individual Differences **27**, 801–804.

Davila Ross, M., Owren, M. J., and Zimmermann, E. (**2009**). "Reconstructing the phylogeny of laughter in great apes and humans," Current Biol. **19**, 1106–1111.

Dixson, A. F. (**2009**). *Sexual Selection and the Origins of Human Mating Systems* (Oxford University Press, New York).

Evans, S., Neave, N., Wakelin, D., and Hamilton, C. (**2008**). "The relationship between testosterone and vocal frequencies in human males," Physiology & Behavior **93**, 783–788.

Ey, E., Pfefferle, D., and Fischer, J. (**2007**). "Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review," Primates **48**, 253–267.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Feinberg, D. R. (**2008**). "Are human faces and voices ornaments signaling common underlying cues," Evolutionary Anthropol. **17**, 112–118.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Little, A. C. (**2008**). "Correlated preferences for men's facial and vocal masculinity," Evolution and Human Behavior **29**, 233–241.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., O'Connor, J. J. M., Tigue, C. C., and Borak, D. J. (**2011**). "Integrating fundamental and formant frequencies in women's preferences for men's voices," Behavioral Ecology **22**, 1320–1325.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., and Perrett, D. I. (**2006**). "Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice," Hormones and Behavior **49**, 215–222.

Fischer, J., Semple, S., Fickenscher, G., Jurgens, R., Kruse, E., Heistermann, M., and Amir, O. (**2011**). "Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime," PLoS One **6**, e24490.

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T. (**2000**). "The evolution of speech: A comparative review," Trends in Cognitive Sciences **4**, 258–267.

Fitch, W. T. (**2006**). "Production of vocalizations in mammals," in *Encyclopedia of Language and Linguistics, 2nd Ed., Vol. 10*, edited by K. Brown (Elsevier Science, Oxford, UK), pp. 115–121.

Fitch, W. T., and Fritz, W. B. (**2006**). "Rhesus macaques spontaneously perceive formants in conspecific vocalizations," J. Acoust. Soc. Am. **120**, 2132–2141.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract. A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Fitch, W. T., and Hauser, M. D. (**2002**). "Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. R. Fay, and A. N. Popper (Springer Verlag, New York), pp. 65–137.

Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," Animal Behaviour **63**, 407–418.

Fitch, W. T., and Reby, D. I. (**2001**). "The descended larynx is not uniquely human," Proc. Royal Soc. B **268**, 1669–1675.

Fraccaro, P. J., Jones, B C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., Little, A. C., and DeBruine, L. M. (**2011**). "Experimental evidence that women speak in a higher voice pitch to men they find attractive," J. Evolutionary Psychol. **9**, 57–67.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

González, J. (**2004**). "Formant frequencies and body size of a speaker: A weak relationship in adult humans," J. Phonetics **32**, 277–287.

Harries, M., Hawkins, S., Hacking, J., and Hughes, I. (**1998**). "Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice," J. Laryngol. and Otology **112**, 451–454.

Harris, T. R., Fitch, W. T., Goldstein, L. M., and Fashing, P. I. (**2006**). "Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass," Ethology **112**, 911–920.

Hartnick, C. J., Rehbar, R., and Prasad, V. (**2005**). "Development and maturation of the pediatric human vocal fold lamina propria," The Laryngoscope **115**, 4–15.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hirano, M., Kurita, S., and Nakashima, T. (**1983**). "Growth, development and aging of human vocal folds," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 22–43.

Hodges-Simeon, C. R., Gaulin, S. J. C., and Puts, D. A. (**2010**). "Different vocal parameters predict perceptions of dominance and attractiveness," Human Nature **21**, 406–427.

Hodges-Simeon, C. R., Gaulin, S. J. C., and Puts, D. A. (**2011**). "Voice correlates of mating success in men: Examining 'Contests' versus 'mate choice' modes of sexual selection," Archives of Sexual Behavior **40**, 551–557.

Hollien, H., Green, R., and Massey, K. (**1994**). "Longitudinal research on adolescent voice change in males," J. Acoust. Soc. Am. **96**, 2646–2654.

Hughes, S. M., Farley, S. D., and Rhodes, B. C. (**2010**). "Vocal and physiological changes in response to the physical attractiveness of conversational partners," J. Nonverbal Behavior **34**, 155–167.

Isles, T. E. (**2009**). "The socio-sexual behaviour of extant archosaurs: Implications for understanding dinosaur behaviour," History of Biol. **21**, 139–214.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. (**2010**). "A domain-specific opposite-sex bias in human preferences for manipulated voice pitch," Animal Behaviour **79**, 57–62.

Katz, W. F., and Assmann, P. F. (**2001**). "Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing," J. Phonetics **29**, 23–51.

Mende, W., Herzel, H., and Wermke, K. (**1990**). "Bifurcation and chaos in newborn infant cries," Physics Lett. A **145**, 418–425.

Mitani, J. C., and Gros-Louis, J. (**1995**). "Species and sex differences in the screams of chimpanzees and bonobos," Int. J. Primatology **16**, 393–411.

O'Connor, J. J. M. (**2011**). "Voice pitch influences perceptions of sexual infidelity," Evolutionary Psychol. **9**, 64–78.

Owren, M. J., Amoss, R. T., and Rendall, D. (**2011**). "Two organizing principles of vocal production: Implications for nonhuman and human primates," Am. J. of Primatology, **73**, 530–544.

Owren, M. J., Berkowitz, M., and Bachorowski, J.-A. (**2007**). "Listeners judge talker sex more efficiently from male than from female vowels," Perception & Psychophysics **69**, 930–941.

Owren, M. J., and Goldstein, M. H. (**2008**). "The babbling-scaffold hypothesis: Subcortical primate-like circuitry helps teach the human cortex how to talk," in *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*, edited by D. K. Oller and U. Griebel (MIT Press, Cambridge, MA) pp. 169–192.

Owren, M. J., and Rendall, D. (**2001**). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," Evolutionary Anthropol. **10**, 58–71.

Owren, M. J., and Rendall, D. (**2003**). "Salience of caller identity in rhesus monkey (*Macaca mulatta*) coo and screams: Perceptual experiments with human listeners," J. Comparative Psychol. **117**, 380–390.

Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951–2963.

Pipitone, R. N., and Gallup Jr., G. G. (**2008**). "Women's voice attractiveness varies across the menstrual cycle," Evolution and Human Behavior **29**, 268–274.

Pisanski, K., and Rendall, D. (**2011**). "The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness," J. Acoust. Soc. Am. **129**, 2201–2212.

Puts, D. A. (**2010**). "Beauty and the beast: mechanisms of sexual selection in humans," J. Human Evolution **31**, 157–175.

Puts, D. A., Apicella, C. L., and Cárdenas, R. A. (**2011**). "Masculine voices signal men's threat potential in forager and industrial societies," Proc. Royal Soc. B. doi: 10.1098/rspb.2011.0829.

Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer," Animal Behaviour **65**, 519–530.

Rendall, D. (**1996**). *Social Communication and Vocal Recognition in Free-Ranging Rhesus Monkeys*. unpublished doctoral dissertation, University of California-Davis.

Rendall, D., Emond, R. E., and Rodman, P. S. (**1996**). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Animal Behaviour **51**, 1007–1015.

Rendall, D., Lloyd, P., Kollias, S., and Ney, C. (**2005**). "Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**, 944–955.

Rendall, D., Notman, H., and Owren, M. J. (**2009**). "Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons," J. Acoust. Soc. Am. **125**, 1792–1805.

Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," J. Acoust. Soc. Am. **103**, 602–614.

Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. D. (**2004**). "Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners," J. Acoust. Soc. Am. **115**, 411–421.

Rendall, D., Vokey, J. R., and Nemeth, C. (**2007**). "Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size," J. of Experimental Psychol.: Human Perception and Performance **33**, 1208–1219.

Riede, T., Owren, M. J., and Clark Arcadi, A. (**2004**). "Nonlinear acoustics in the pant-hoot vocalizations of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," Am. J. Primatology **64**, 277–291.

Riede, T., and Fitch, W. T. (**1999**). "Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*)," J. Experimental Biol. **202**, 2859–2867.

Riede, T., and Zuberbühler, K. (**2003**). "The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization," J. Acoust. Soc. Am. **114**, 1132–1142.

Schweinfurth, J. M., and S. L. Thibeault, S. L. (**2008**). "Does hyaluronic acid distribution in the larynx relate to the newborn's capacity for crying?" Laryngoscope **118**, 1692–1699.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., and Gurven. M. (**2010**). "Adaptations in humans for assessing physical strength from the voice," Proc. Royal Soc. B **277**, 3509–3518.

Stathopoulos, E. T., Huber, J. E., and Sussman, J. E. (**2011**). "Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age," J. Speech, Language, and Hearing Res. **54**, 1011–1021.

Stemple, J. C., Glaze, L. E., and Klaben, B. G. (**2009**). *Clinical Voice Pathology: Theory and Practice, 4th Ed.* (Plural Publishing, San Diego).

Stevens, K. N. (**2000**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Perception & Psychophysics **49**, 365–371.

Taylor, A. M., and Reby, D. (**2010**). "The contribution of source–filter theory to mammal vocal communication research," J. Zoology **280**, 221–236.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., and Herzel, H. (**2002**). "Nonlinear prediction of irregular animal vocalizations," J. Acoust. Soc. Am. **111**, 2908–2919.

van Dommelen, W. A., and Moxness, B. H. (**1993**). "Acoustic parameters in speaker height and weight identification: Sex-specific behaviour," Language and Speech **38**, 267–287.

Weishampel, D. B. (**1997**). "Dinosaurian cacophony," Bioscience **47**, 150–159.

Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). "Subharmonics, biphonation, and deterministic chaos in mammalian vocalization," Bioacoustics **9**, 171–196.

Michael J. Owren received his B.A. in Psychology from Reed College, and his Ph.D. in Animal Behavior from Indiana University, Bloomington. He is a Fellow of the Acoustical Society of America (2011), as well as the Association for Psychological Science (2011). Following a National Research Service Award (NRSA) postdoctoral fellowship jointly sponsored at the University of Pennsylvania and the University of California-Davis, Michael held faculty positions at the University of Colorado at Denver, Reed College, and Cornell University. He is currently an Associate Professor in Psychology at Georgia State University, where he is also Chair of the Cognitive Sciences program. Broadly stated, his research focuses on the role of acoustic structure and communicative function in the evolution of both human and nonhuman vocalizations, particularly including primate calls, emotion-triggered human vocalizations such as spontaneous laughter, and human speech. Michael has been a Consulting Editor for *Psychological Science* and the *Journal of Comparative Psychology*, and is currently an Associate Editor for *Emotion Review* and the *Journal of the Acoustical Society of America* (Bioacoustics).

PHONETICS OF ENDANGERED LANGUAGES

D. H. Whalen
Speech-Language-Hearing Program
Graduate Center of the City University of New York
365 Fifth Avenue
New York, New York 10016
and
Haskins Laboratories
300 George Street, Suite 900
New Haven, Connecticut 06511
and
Endangered Language Fund

Christian T. DiCanio
Haskins Laboratories
300 George Street, Suite 900
New Haven, Connecticut 06511

Patricia A. Shaw
First Nations Languages Program
University of British Columbia
1866 Main Mall
Vancouver, British Columbia V6T 1Z1, Canada

The world is filled with an astounding array of languages, 6,909, by the count of the *Ethnologue* (Lewis, 2009). Most of these use an acoustic signal as the main element in signal transmission, though vision affects speech even for typically hearing individuals (e.g., McGurk and MacDonald, 1976); sign languages (126 are listed in Lewis, 2009) use the visual channel almost exclusively. The acoustic signal for speech is powered mostly by the larynx and shaped by the vocal tract. Because human populations have essentially the same anatomy, there is a great deal of similarity in the sounds that languages use. However, there is an impressive range of variability as well. The largest survey of sound systems (Maddieson, 1984), for example, lists no sound that occurs in all languages, even though broad patterns are seen. The number of significant sounds, or phonemes, ranges from about 12 (Pirahã, Rotokas) to over a hundred (!Xóõ), and the mechanisms used vary greatly as well.

> *"Documenting differences among the world's most disparate languages is of central importance to the field of linguistics and to the language community's heritage."*

The acoustics of speech have proven to be extraordinarily complex. Early estimates that simple acoustic pattern matching would make automatic speech recognition practical (e.g., Juang and Furui, 2000) proved to be wrong. Current high levels of recognition are founded on acoustic analysis of huge amounts of data combined with statistical inference about common co-occurrences among words and sounds (e.g., Jelinek, 2009). Understanding what it is that listeners are sensitive to in this complex acoustic signal has been fruitfully guided by examining how those sounds are generated in the vocal tract (e.g., Iskarous *et al.*, 2010). For sounds in lesser-studied languages, having articulatory data to help interpret the acoustic signal is even more valuable.

A few of the world's languages have been well-studied, but most of them have yet to be explored in any detail, if at all. Pioneering efforts by Peter Ladefoged and Ian Maddieson to record the sounds of the world's languages resulted in descriptions of many of the less typical sounds used (Ladefoged and Maddieson, 1996) and phonetic sketches of several languages (e.g., Ladefoged *et al.*, 1998; McDonough *et al.*, 1993; Silverman *et al.*, 1995; Taff, *et al.*, 2001). Funded for many years by the National Science Foundation, this work has provided an invaluable basis for further phonetic work, given that it provides an initial understanding of the expected production and acoustic bases of virtually all the sounds that are used in language.

These descriptions are far from complete, however, as can be seen from two trends. First, we are still learning new and important facts about even the best-known languages, including English, as evidenced by the continuing appearance of phonetic studies in the pages of scientific journals. Second, even though the "same" sound may be described in phonetic studies from different languages, we can not assume that it shares the same characteristics across languages. For instance, ejective fricatives occur in several languages (Maddieson *et al.*, 2001), but they vary widely in how they are produced. Such wide variability calls into question the validity of these phonetic categories across languages. There is much left to learn.

Phoneticians have begun stepping up their efforts to study endangered languages while there are still fluent speakers left, especially those who acquired the language as their first. Language documentation efforts have been on the upswing in recent years, but phonetic studies have not been as obviously useful as the collection of texts and the making of dictionaries. As more communities try to revive their languages from documentary sources, it is becoming increasingly clear that phonetic documentation can contribute in valuable ways to describing the pronunciation of the ancestral language. Furthermore, many endangered languages lack writing systems, and good phonetic descriptions can often help guide their development. Contributing to community literacy is a frequent concrete aim. The scientific goals of the academic community often overlap with the revitalization goals of native communities.

Table 1: Tahltan consonants. Highlighted columns participate in the three-way harmony; those left unshaded are transparent to harmony. The orthography used draws on three characters commonly used in Americanist traditions. These differ from the International Phonetic Alphabet as follows: ž for ʒ, š for ʃ, and y for j.

b	d	dl	dθ	dz	dž	g	gʷ	G	
	t	tl	tθ	ts	tš	k	kʷ	q	
	t'	tɬ'	tθ'	ts'	tš'	k'	kʷ'	q'	ʔ
		ɬ	θ	s	š	x	xʷ	X	h
		l	ð	z	ž	ɣ	ɣʷ	ʁ	
m	n				y		w		
	n'								

In this paper, we will discuss two phonetic studies of endangered languages. Both studies use articulatory and acoustic data to examine specific scientific questions. The first study uses ultrasound, while the second uses acoustic data coupled with electroglottography (EGG).

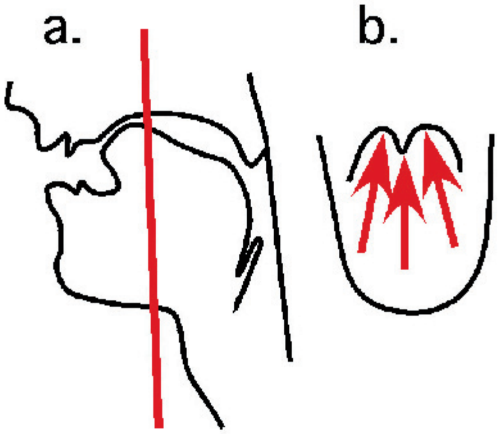
Ultrasound study of tongue shape in Tahltan

Tahltan (ISO 639 code tht) is an Athapaskan language of northern British Columbia, spoken by fewer than 20 elders as their first language. Some younger community members are learning the language as a second language, and the elders are hopeful that the language can be revived. There are three main dialects, and one of them, Telegraph Creek (see Fig. 1), has one of the world's few three-way consonant harmony systems. Harmony is a linguistic process in which sounds in a word have to agree on, or "harmonize with," certain dimensions. Vowel harmony is fairly common across languages, large and small. Some forms are rather limited, such as the umlaut process of German, while systems that affect all vowels can be found in some languages like Turkish. In Turkish, all vowels in a word should have the same frontness (the vowels of "bee" and "bay" are front, those of "boo" and "though" are back), although there are exceptions. Some affixes agree in rounding as well. (The English front vowels are unrounded; the back ones are rounded; Turkish has front rounded and back unrounded vowels.) Less common is consonant harmony, in which one or more features of the non-vowel sounds have to agree. Most such systems have two sounds that have to agree (say an "s" and an "sh"), but a handful have a three-way system of agreement.

All verb stems in Tahltan that contain fricatives (sounds like "s" and "sh") have to come from the same "series"—one that has "th" sounds (like in "thing"), "s" sounds or "sh" sounds (Shaw, 1991). It sounds simple, but Tahltan has 46 consonants, 15 of which participate in the harmony while the others are "transparent" to it in that they neither change nor stop the harmony (Table 1). When vowel harmony is at issue, it is easy to imagine that the consonants are overlaid on top of vowels and that the vowels are really adjacent underneath. With consonants, it seems instead that these segments are acting "at a distance," reaching across vowels that make use of the same articulator—the tongue—that the consonants

Fig. 1. View of the Tuya River crossing heading into Telegraph Creek, British Columbia, Canada.

Fig. 2. Data collection from speaker Margery Inkster with a portable ultrasound machine. The probe is held under the chin, giving a view across the tongue.



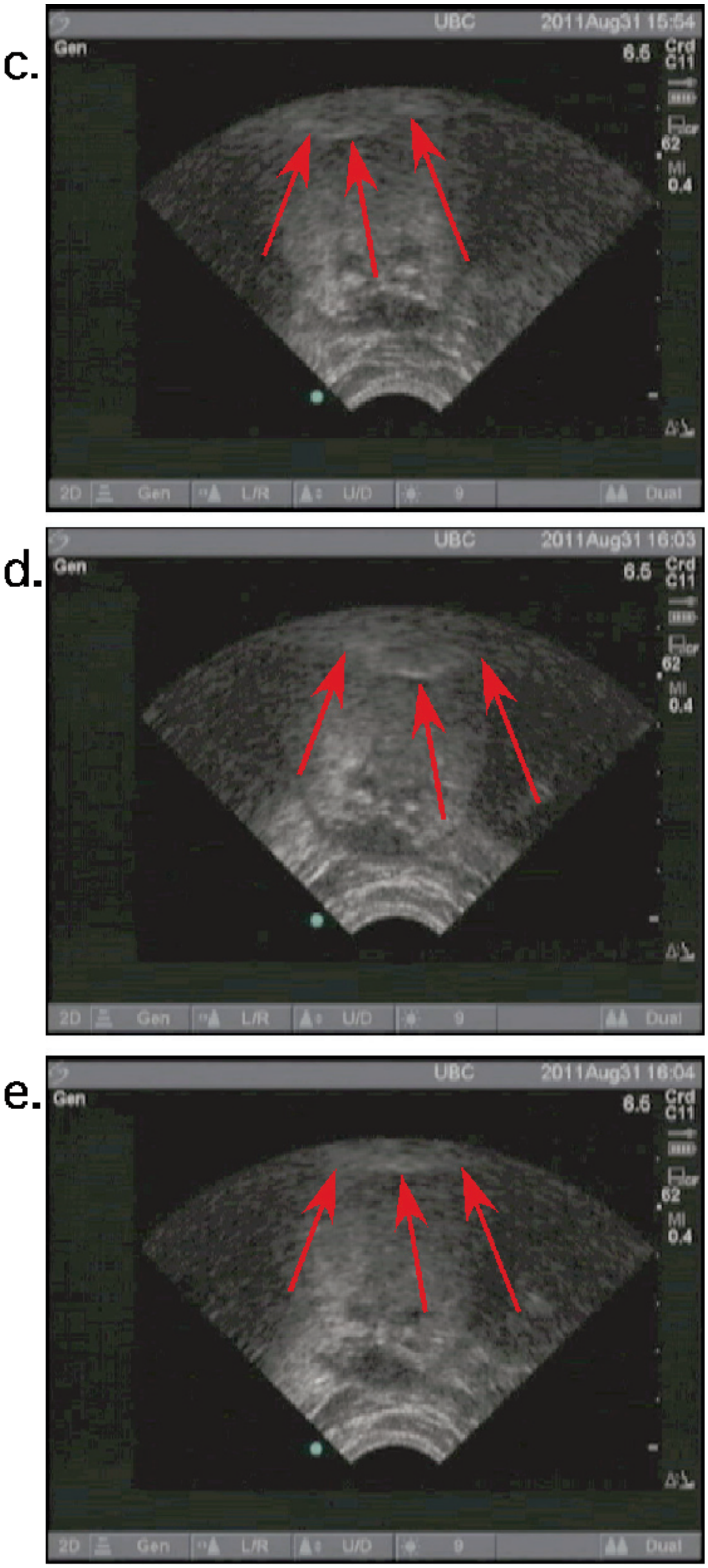


Fig. 3. Ultrasound imaging of the tongue. a: Schematic midsagittal vocal tract for an [š]-like articulation. The red line shows the approximate location of the plane of the ultrasound image. b: Schematic of the tongue shape in the coronal plane shown by the line in (a). The arrows mark the peaks and trough of the tongue groove. c-e: Three ultrasound images of the tongue during c) [θ] d) [s] e) [š]. The speaker's left is on the left and the speaker's right on the right. Selection of the three points was guided by repeated viewing of the video image, where the coherent structure of the tongue is more obvious than it is in any single frame.

depend on. Explaining this long-distance effect has been challenging.

The operation of this harmony can be seen in the following forms, where the underlined letter corresponds to the word segment "-s-", the marker for the first person singular subject:

εsk'ɑ:	'I'm gutting fish'
εšdžini	'I'm singing'
εθdu:θ	'I whipped him'

As can be seen from this last form, this agreement can work across intervening segments (here, /d/ and /u/) that neither promote nor block harmony. It looks like "action at a distance."

What if there is something that makes this process local after all? Gafos (1996) proposed just such a solution. He claimed that the differences in the amount of tongue grooving seen in these fricatives could extend through the intervening vowels and non-participating consonants, thus making the agreement local. Our efforts, supported by the National Science Foundation's Office of Polar Programs, have begun to test this hypothesis using ultrasound images of the tongue.

Ultrasound allows us to look at the tongue during speech (e.g., Stone and Lundberg, 1996) in a minimally invasive way that is appropriate for a broad range of speakers, including the elderly, and in non-laboratory conditions as well (Gick *et al.*, 2005). We generally take a "sagittal" view of the tongue, showing a two-dimensional image from near the tip to the back of the tongue near the uvula. However, we can just as easily take "coronal" sections that go across the tongue. This is ideal for measuring tongue grooves, and it is what we did in our study.

Figure 2 shows one of our speakers, Margery Inkster, with the ultrasound probe under her chin. Figure 3 shows three of the cross-sections of her tongue during the fricatives [θ s ʃ] The groove depth differs among the three, and this same groove persists through the vowel. This is in accord with Gafos's prediction. (See Figs. 2 and 3)

To understand where this pattern might have come from, however, we need to know whether the same kind of persistence appears in languages without the consonant harmony.

Fig. 4. Young Trique women dancing in traditional huipiles (dresses), Oaxaca, Mexico.

For this, we studied a language close at hand, English. Our preliminary results showed that, indeed, English also allows these three different groove depths to persist through the vowel (Whalen *et al*., 2011). Although no language is completely neutral as a comparison to harmony processes, it is useful to compare such patterns to languages for which we have greater phonetic knowledge, like English. So it is plausible that the foundations for a three-way harmony system could be seen in a language that has not yet shown any use of such a process.

We hope to extend this work by measuring images from additional Tahltan speakers and by making a more thorough comparison with English data. Nevertheless, these results are already helping to settle an important issue in linguistic theory: Consonant harmony can be seen as a local process after all.

Electroglottographic study of devoicing in Itunyoso Trique

Itunyoso Trique (ISO 639 trq) is an Oto-Manguean language spoken in Oaxaca, Mexico by approximately 1,400 people (Lewis, 2009). It is one of three Trique dialects, each of which has a unique sound structure and grammar. Like all Oto-Manguean languages, Itunyoso Trique is tonal. This means that the level and direction of pitch in the voice may distinguish the meaning of words. The Itunyoso dialect has nine different tonal melodies that words can carry (DiCanio, 2010). As a comparison, Mandarin Chinese is also tonal, but with only four possible tonal melodies. Most Trique speakers are bilingual, speaking Spanish as a second language. In recent years, children have begun to use Spanish more among their peers (IGABE, 2011). While there are currently many Trique speakers, the language is danger of being replaced entirely by Spanish (See Fig. 4).

The large inventory of tonal melodies is certainly a feature of the Itunyoso Trique language that is seldom found elsewhere, but it is not the only such feature. Itunyoso Trique also has a contrast between long and short consonants found only in the initial position of words. While many languages in the world have consonant length contrasts, such as Japanese *katta* 'bought' and *kata* 'shoulder', this distinction is often restricted to the middle of the word (Muller, 2001). Itunyoso Trique is one of only two languages in the world known to restrict this contrast between long and short consonants to the *beginning* of a word. The other language is Nhaheun (ISO 639 nev), an Austroasiatic language spoken in Laos (Muller, 2001).

One of the intriguing things about this rare contrast in Trique is how voicing functions in the short and long stop consonants. Stops are sounds with a closure in the mouth followed by a sudden release. Many languages distinguish between voiceless stops, like for instance, French "p", "t", and "k," and voiced stops, like French "b", "d", and "g." Voiced stops are produced when the vocal folds are vibrating, voiceless stops when the vocal folds are not vibrating. Yet, in Itunyoso Trique, the long consonants are often preceded by a short puff of air, called *preaspiration*. The short consonants are rarely produced this way, but vary in their production. They may be voiceless, like a "p", or voiced, like a "b." This variability has led researchers to misclassify length contrasts like the one in Itunyoso Trique as strength contrasts, i.e. “fortis” and “lenis” stops.

DiCanio (2012) examined the timing of vocal fold vibration in Trique consonants using electroglottography (EGG) in order to determine exactly what accounted for this variability in voicing in short stops and to investigate if another explanation could account for the pattern. With EGG, sensors are placed on opposite sides of the speaker’s neck, just over the thyroid cartilage (below the Adam’s apple) through which a weak electrical current is passed. When the vocal folds are closed, more of the current can pass from one side to the other. When the vocal folds are open, less current passes through. EGG maxima correspond to the moment of maximum contact between the vocal folds while minima correspond to the moment of minimum contact between the vocal folds (Childers and Krishnamurthy, 1985; Childers and Lee, 1991; Heinrich *et al.*, 2004). The presence of EGG maxima and minima indicates that there is vocal fold vibration. EGG data are typically collected along with acoustic recordings for the identification of acoustic-phonetic boundaries.

The advantage to using an electroglottograph is that the EGG signal is unaffected by acoustic disturbances in field recordings. In rural villages, recording is often done in quiet spaces in private homes. Typically, houses are constructed by community members without the help of an electrician. Thus, in addition to the external noise found in these communities, there are often ground loops due to the use of low current wiring. Ground loops in AC power lines can produce an unwanted signal in acoustic recordings, with a fundamental frequency of 60 Hz and its associated harmonics. To adjust for this effect, a low stop filter can be applied to the recordings; yet, this filter also eliminates low frequency voicing from the signal. Devoicing typically involves low frequency and low amplitude glottal pulses. EGG is ideal for examining devoicing in these less-than-ideal recording conditions because it accurately captures low amplitude voicing.

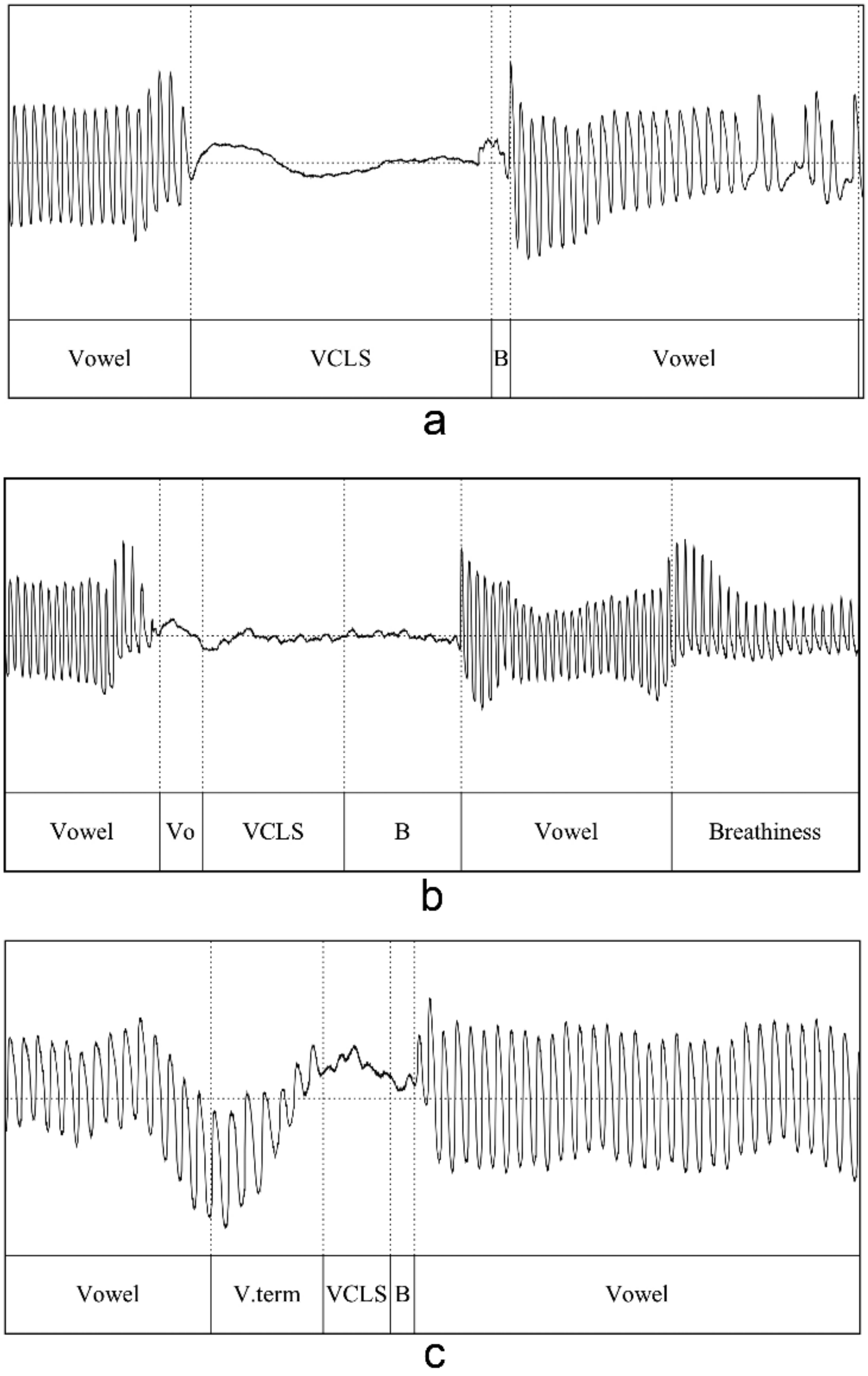


Fig. 5. Oral-glottal alignment configurations from EGG signal. VCLS represents voicelessness during closure. B represents the voiceless burst duration following closure release. Figures from DiCanio (in press). Reprinted with permission.

There are three ways in which vocal fold vibration may be aligned to the closure of a stop in running speech. If the sound preceding the stop is voiced, voicing may cease coincidental with the closure of the following stop. This is simultaneous oral-glottal alignment. Voicing may also cease during the vowel duration prior to stop closure. The time between devoicing and closure is called *voice offset time (Vo)*. This is the gestural configuration used for

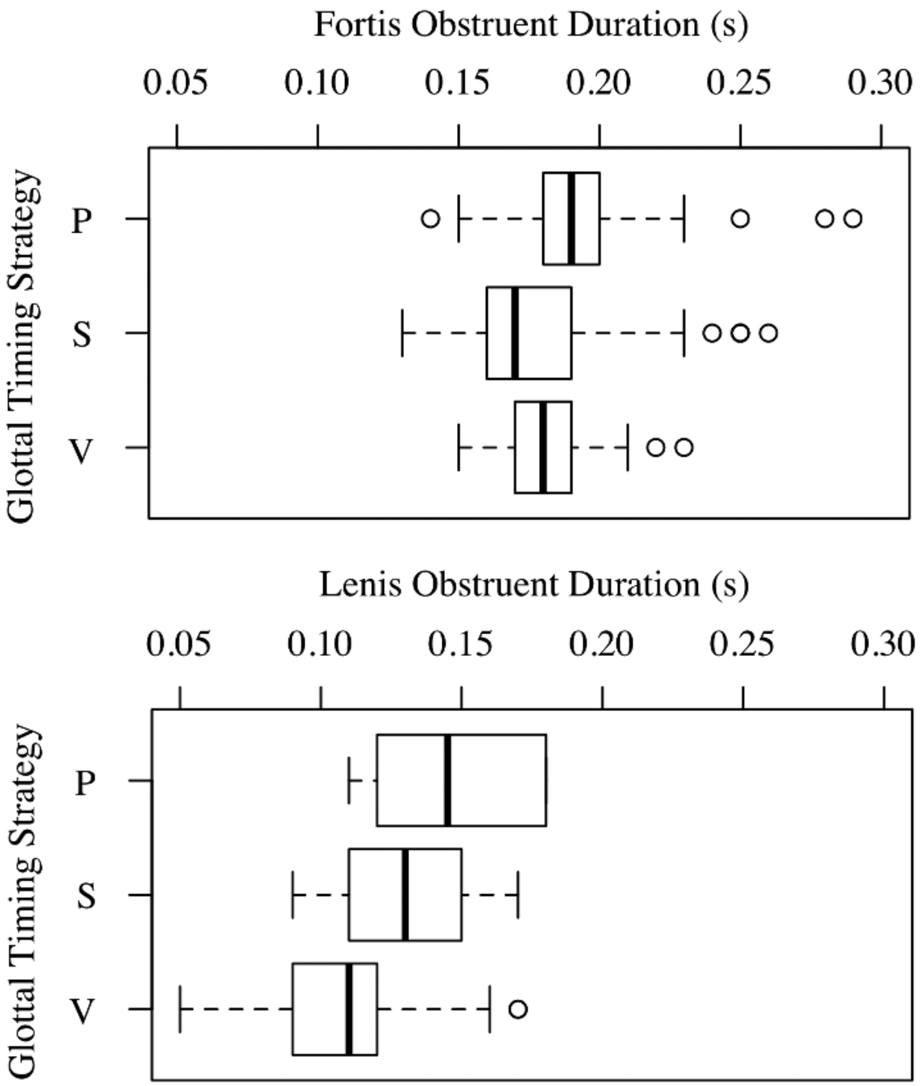


Fig. 6. Effect of total stop duration on glottal timing strategy. Long stop data (left), Short stop data (right). (P = preaspiration, S = simultaneous timing, V = voicing during closure). Figure from DiCanio (in press). Reprinted with permission.

preaspirated stops, as in Icelandic (Helgason, 2002). Finally, voicing may cease during the stop closure duration. This is called *partial voicing* or *voice termination time (V.term)*. Figures 5a-5c show these three oral-glottal alignment configurations, respectively.

In DiCanio (2012), acoustic and EGG recordings were made from four speakers producing short and long stops in Itunyoso Trique. These words were presented in contexts so that the duration of stop closure could be examined from the acoustic signal along with the alignment of devoicing from the EGG signal. The results found significant differences between the long and short stops in relation to the timing of devoicing. Long stops are produced with either simultaneous glottal timing (49.4%), as shown in Fig. 5a, or with devoicing prior to closure (43.8%), as shown in Fig. 5b. Short stops are typically produced with voicing which extends into the stop closure (84%), as shown in Figure 5c, and rarely with simultaneous oral-glottal alignment (14.1%). The amount of voicing during closure for the short (lenis) stops varied in relation to the overall duration of the stop, shown in Fig. 6. No such variation was observed for long (fortis) stops.

Research on related languages like Zapotec is inconclusive as to whether short stops are voiced or voiceless (Avelino, 2001; Nellis and Hollenbach, 1980). These results suggest that such variability is conditioned by within-category changes in consonant duration. The effect of such variability is that, for some short stops with particularly short duration, voicing may extend through the entire stop. This pattern is called passive voicing (Jansen, 2004; Westbury and Keating, 1986). Long stops show a different pattern. They are actively devoiced by an abrupt glottal spreading gesture, the timing of which does not vary with consonant duration.

This work addresses a general descriptive question within Itunyoso Trique: what is the phonetic realization of short and long stops in the language? However, the answer to this question also informs a more general theory of speech production. Voicing is generally considered to be a discrete category within the phonology of a language. Sounds are typically classified as either "voiced," with vocal fold vibration, or "voiceless," without vocal fold vibration. The Trique data suggest that for certain sound types, there is a continuum of voicing that varies due to durational differences. Findings like this illustrate one of the ways in which descriptive phonetic work on endangered languages has a broader impact on the linguistic sciences.

Summary

We have presented just two of the many phonetic investigations currently under way that examine the world's extensive, but shrinking, variety of languages. Documenting differences between the world's most disparate languages is of central importance to the field of linguistics and to the language community's heritage. Such efforts are funded by such governmental agencies as the National Science Foundation and the Administration for Native Americans, and non-profits like the Endangered Language Fund. While the work presented here focuses mainly on the production of consonants, many more aspects of speech acoustics need to be investigated. For instance, prosody (speech timing, intonation) varies substantially across languages; yet, this topic is rarely addressed in studies of endangered languages. As more communities attempt to revive their heritage languages or seek the help of linguists in the development of writing systems, phonetic detail becomes more important. The techniques that are now available for doing phonetic research in field locations are much better than those available even a decade ago, and they continue to improve. The full range of phonetic diversity in human languages should become clearer in the coming years as many more intriguing patterns are discovered.

Acknowledgments

This research was funded by National Science Foundation grants NSF-1128159 and NSF-0966411 to Haskins Laboratories. Reethee Antony, Colleen Leung and Aude Noiray provided crucial support in the collection and analysis of the Tahltan data, as did Bryan Gick and Donald Derrick. Acknowledgments are given to the community of San Martín Itunyoso.**AT**

References

Avelino, H. (**2001**). “Phonetic correlates of fortis-lenis in Yalálag Zapotec, ” unpublished Master’s thesis, University of California, Los Angeles.

Childers, D. G., and Krishnamurthy, A. K. (**1985**). “A critical review of electroglottography,” CRC Critical Reviews in Biomedical Eng. **12**, 131–161.

Childers, D. G., and Lee, C. K. (**1991**). “Vocal quality factors: Analysis, synthesis, and perception,” J. Acoust. Soc. Am. **90**, 2394–2410.

DiCanio, C. T. (**2010**). “Itunyoso Trique (Illustrations of the IPA),” J. Intl. Phonetic Assn. **40**, 227–238.

DiCanio, C. T. (**2012**). “The phonetics of fortis and lenis consonants in Itunyoso Trique,” Intl. J. Am. Linguistics **78**, 239–272.

Gafos, A. (**1996**). “The articulary basis of locality in phonology,” unpublished Ph.D. dissertation, Johns Hopkins University, Baltimore, MD.

Gick, B., Bird, S., and Wilson, I. (**2005**). “Techniques for field application of lingual ultrasound imaging,” Clinical Linguistics and Phonetics **19**, 503–514.

Heinrich, N., D’Alessandro, C., Doval, B., and Castellengo, M. (**2004**). “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” J. Acoust. Soc. Am. **115**, 1321–1332.

Helgason, P. (**2002**). “Preaspiration in the Nordic Languages: Synchronic and diachronic aspects,” unpublished Ph.D. thesis, Stockholm University.

IGABE (News Agency) (**2011**, March 9). “Resisten niños triquis a usar lengua materna,” Noticias Voz e Imagen.

Iskarous, K., Nam, H., and Whalen, D. H. (**2010**). Perception of articulatory dynamics from acoustic signatures. J. Acoust. Soc. Am., **127**, 3717-3728.

Jansen, W. (**2004**). “Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch,” unpublished Ph.D. dissertation, Groningen University.

Jelinek, F. (**2009**). “The dawn of statistical ASR and MT,” Computational Linguistics **35**, 483–494.

Juang, B. H., and Furui, S. (**2000**). “Automatic recognition and understanding of spoken language: A first step toward natural

human-machine communication," Proc. of the IEEE **88**, 1142-1165.

Ladefoged, P., Ladefoged, J., Turk, A., Hind, K., and Skilton, S. J. (**1998**). "Phonetic Structures of Scottish Gaelic," J. Phonetics **28**, 1–41.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages*. (Blackwell, Oxford, UK; Cambridge, MA) .

Lewis, M. P. (Ed.). (**2009**). *Ethnologue: Languages of the World, Sixteenth edition.* (SIL International, Dallas).

Maddieson, I. (**1984**). *Patterns of Sounds* (Cambridge University Press, New York).

Maddieson, I., Smith, C. L., and Bessell, N. (**2001**). "Aspects of the phonetics of Tlingit," Anthropological Linguistics **43**, 135–176.

McDonough, J., Ladefoged, P., and George, H. (**1993**). "Navajo vowels and phonetic universal tendencies," UCLA Working Papers in Phonetics **84**, 143–164.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature **264**, 746–748.

Muller, J. S. (**2001**). "The phonology and phonetics of word-initial geminates," unpublished Ph.D. thesis, The Ohio State University.

Nellis, D. G., and Hollenbach, B. E. (**1980**). "Fortis versus Lenis in Cajonos Zapotec Phonology," Intl. J. Am. Linguistics **46**, 92–105.

Shaw, P. A. (**1991**). "Consonant harmony systems: The special status of coronal harmony," in C. Paradis and J.-F. Prunet (Eds.), Phonetics and Phonology 2: The Special Status of Coronals (pp. 125–157) (Academic Press, San Diego).

Silverman, D., Blankenship, B., Kirk, P., and Ladefoged, P. (**1995**). "Phonetic structures in Jalapa Mazatec," Anthropological Linguistics **37**, 70–88.

Stone, M. L., and Lundberg, A. (**1996**). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. **99**, 3728–3737.

Taff, A., Rozelle, L., Cho, T., Ladefoged, P., Dirks, M., and Wegelin, J. (**2001**). "Phonetic structures of Aleut," J. Phonetics **29**, 231–271.

Westbury, J. R., and Keating, P. A. (**1986**). "On the naturalness of stop consonant voicing," J. Linguistics **22**, 145–166.

Whalen, D. H., Shaw, P. A., Noiray, A., and Antony, R. (**2011**). "Analogs of Tahltan consonant harmony in English CVC syllables," in W.-S. Lee and E. Zee (Eds.), *Proc. of the 17th International Congress of Phonetic Sciences* (pp. 2129–2132). (City University of Hong Kong, Hong Kong.)

Christian T. DiCanio is a research associate at Haskins Laboratories. He received his Ph.D. in Linguistics from the University of California, Berkeley, in 2008. Prior to his employment at Haskins, he was a Fyssen Foundation postdoctoral fellow at Université Lumière in Lyon, France. His research encompasses both descriptive phonetic work on endangered languages and experimental work in speech production and perception. He has done field work on a variety of endangered languages in Mexico, including Trique, Ixcatec, and Mixtec. Apart from his descriptive linguistic work, his areas of research include tone perception, voice quality perception, voice quality production, and the coarticulatory dynamics of tone. Together with colleagues at Haskins Labs, he is investigating how automatic speech recognition software can be used to align text to speech in corpus data from different endangered languages.

Patricia A. Shaw received her Ph.D. in Linguistics from the University of Toronto, and is the Founder and Chair of the First Nations Languages Program at the University of British Columbia in Vancouver, BC, where she provides training for language documentation and revitalization and engages in research on and teaching of Kwak'wala, Musqueam and Cree. She has worked in close collaboration with several critically endangered language communities (Salish, Wakashan, Siouan, Athapaskan, Tsimshian) to record and analyze extant grammatical knowledge, to teach research skills and archiving methodologies, and to develop pedagogical materials for language revitalization. She is currently President of the Society for the Study of the Indigenous Languages of the Americas (SSILA), an organization that brings linguists and community members together to work on languages of North, Central, and South America.

D. H. Whalen received his B.A. from Rice University and his Ph.D. from Yale University. He is currently Distinguished Professor of Speech-Language-Hearing Sciences at the Graduate Center of the City University of New York. He also serves as Vice President of Research at Haskins Laboratories, a private non-profit research institute in New Haven, CT. He is Founder and President of the Endangered Language Fund, a nonprofit organization dedicated to the description and revitalization of languages at risk of falling silent. You can learn more at www.endangeredlanguagefund.org. He is a Fellow of the Acoustical Society of America.

APPLICATIONS DUE 1 FEBRUARY 2012

SEABASS 2012

BioAcoustics Summer School

SeaBASS provides the opportunity for graduate students interested in pursuing careers in marine bioacoustics to develop a strong foundation of both marine animal biology and acoustics from distinguished lecturers in the field. The goals of SeaBASS are to discuss important topics in marine bioacoustics, foster technical communication across disciplines, and promote mentoring and collaboration. SeaBASS gives students an opportunity to learn from experts who will discuss a suite of topics not often offered at any one university.

WHEN:	**June 17-22, 2012**
WHERE:	**Penn Stater Conference Center Hotel, State College, PA**
COSTS:	**The basic costs of student attendance (room and board, based on double occupancy for students) will be funded from sponsor support. There is no registration fee. Full-time participation of all participants is required. Travel costs are the responsibility of the student, but travel assistance can be requested during the application process.**
APPLICATION:	**On-line application is available at www.arl.psu.edu/edu_seabass.php. Applications deadline is February 1, 2012. Direct questions to Dr. Jennifer Miksis-Olds via email at: seabass@arl.psu.edu.**

COURSE TOPICS:

Introduction to Underwater Sound, Sound Propagation, Marine Mammal Biology & Behavior, Sound Production, Fisheries Acoustics/Fish Behavior, Hearing and Masking, Marine Animal Acoustic Communication, Echolocation, Hot Topic: Passive Acoustic Monitoring, Hot Topic: Effects of Noise

PENNSTATE
Applied Research Laboratory

DEPARTMENT OF THE NAVY
ONR
Science & Technology

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

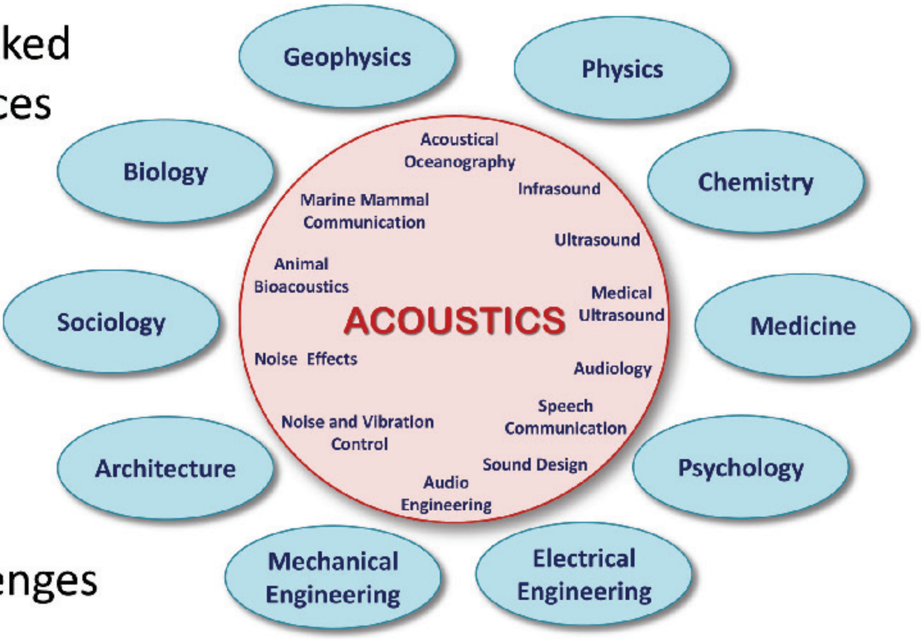
INTERNATIONAL COMMISSION FOR ACOUSTICS ICA

www.icacommission.org

Acoustics is linked to other sciences

… multi-disciplinary

… meeting global challenges



ICA promotes international development and collaboration in all fields of acoustics including research, development, education, and standardization

44 national member societies from around the world

The International Commission for Acoustics (ICA) was instituted in 1951 as a sub-committee to the International Union of Pure and Applied Physics (IUPAP). The ICA in its new statutes held its first General Assembly in 1998 during the 16th Congress in Seattle where the By-laws of the new organization were adopted by the Member Societies. The ICA is a Scientific Associate of the International Council for Science (ICSU) and an Affiliated Commission for both the International Union of Pure and Applied Physics (IUPAP) and the International Union of Theoretical and Applied Mechanics (IUTAM).

affiliated with

- European Acoustics Association
- Iberoamerican Federation of Acoustics
- Western Pacific Acoustics Commission
- International Institute of Noise Control Engineering
- International Institute for Acoustics and Vibration
- International Congress on Ultrasonics
- International Commission on Biological Effects of Noise
- Audio Engineering Society

ICA Board

President MICHAEL VORLÄNDER, Germany
Vice-President CHARLES SCHMID, U.S.A.
Secretary-General MARION BURGESS, Australia
Treasurer ANTONIO PEREZ-LOPEZ, Spain
Past-President SAMIR GERGES, Brazil

Canada MICHAEL STINSON
China JING TIAN
Denmark CLAUS MØLLER PETERSEN
France PHILLIPE BLANC-BENON
Italy ADRIANO ALIPPI
Japan SONOKO KUWANO
Korea JEONG-GUON IH
Poland EUGENIUSZ KOZACZKA
Russia VYACHESLAV MASLOV
United Kingdom YIU LAM

ICASecGen@icacommission.org

Next congress:

ICA 2013 Montréal

www.ica2013montreal.org

hosted by the
Acoustical Society of America and
Canadian Acoustical Association

The International Commission for Acoustics (ICA) is the international umbrella organization representing acoustics in the world. It was instituted in 1951 as a sub-committee to the International Union for Pure and Applied Physics, IUPAP. The ICA in its new statutes held its first General Assembly in 1998 during the 16th Congress in Seattle where the by-laws of the new organization were adopted by the Member Societies. The ICA today is a Scientific Associate of ICSU (International Council for Science, a non-governmental organization under UNESCO with a global membership of national scientific bodies (121 members) and international scientific unions (30 members)) and an Affiliated Commission for both IUPAP and IUTAM, the International Unions for Pure and Applied Physics and Theoretical and Applied Mechanics.

In contrast to a Union, however, a Commission is considered to represent a kind of sub-science. In this respect acoustics is apparently a part of physics or mechanics. The consequence for acoustics world wide is well known: Acoustic departments and institutes are scattered in various schools and faculties. Coordination of our activities is sometimes difficult, calling for better support in academia as well as for the profession itself. Sometimes it is difficult to make our voice heard inside physics, engineering, biology and many of the other disciplines.

Following-up the initiatives started by the pioneering work of the past ICA presidents Gilles Daigle and Phil Nelson, a major objective of the ICA board is to get acoustics established in ICSU as its own scientific discipline covering more than physics and mechanics, thus forming its own Union. We think that there will be benefit for all people working in or with acoustics and vibration. This process, however, is very tedious and can be continued only in small steps, year by year. We, as the executive officers, will continue this discussion in ICSU.

The ICA has 44 national member societies from around the world and 8 International Affiliates (I-INCE, ICBEN, IIAV, ICU, AES, EAA, FIA and WESPAC). The 2011-2013 Board includes the executive officers, President Michael Vorländer (Germany), Secretary General Marion Burgess (Australia), Vice-President Charles Schmid (USA), Treasurer Antonio Perez-Lopez (Spain) and Past-President Samir Gerges (Brazil), and ten more board members representing the acoustical societies of Canada, China, Denmark, France, Italy, Japan, Korea, Poland, Russia, and the United Kingdom.

One of the main activities of the ICA is the organization of the International Congress on Acoustics (Sydney 2010, Madrid 2007, Kyoto 2004, etc.). But in between this triennial conference planning, the ICA commission sponsors specialized symposia, helps emerging acoustic societies in their foundation, and maintains the meeting calendar of acoustic events throughout the world. The ICA-sponsored conferences are normally limited to a specialized topic with an anticipated small attendance typically no more than 300. Specialist regional meetings or national meetings are supported, especially in developing regions, but only considered if the conference has an international character. The amount of financial support is mainly provided by the ICA to pay travel expenses for distinguished speakers, young scientists and especially for scientists from developing countries.

With these initiatives we are promoting international development and collaboration in all fields of acoustics including research, development, education, and standardization. All member societies are invited to distribute communications from the ICA to their members. Individual members in member societies may apply for any of the grants and awards offered by the ICA.

The world family of acoustics will meet again at ICA 2013 in June 2013 in Montreal, Canada. We are looking forward to visiting this beautiful city and to meeting there to discuss progress in acoustics.

Aachen, Germany
Michael Vorländer
President ICA 2011-2013

Read more:
www.ica2013montreal.org
http://www.icacommission.org
http://www.icacommission.org/ICAarticle.pdf

Co-sponsored Meeting Reports

Dick Stern
1150 Linden Hall Rd.
Boalsburg Pennsylvania 16827

Acoustics Today welcomes contributions for “Co-sponsored Meeting Reports” There is no charge for this service. Submissions of no more than 1000 words that may be edited in MSWord and up to two photos should be e-mailed to <AcousticsToday@aip.org>.

ACOUSTIC COMMUNICATION BY ANIMALS SYMPOSIUM

Mary Bates
maryebates@gmail.com

In August, researchers from disparate fields with a common interest in animal bioacoustics met in Ithaca, N.Y., for the 3rd International Symposium on Acoustic Communication by Animals. The conference was hosted by Cornell University's Bioacoustics Research Program and sponsors included the Acoustical Society of America, Office of Naval Research, National Oceanic and Atmospheric Administration, and National Science Foundation.

The meeting began with a keynote talk from Peter Narins of the University of California, Los Angeles. Narins discussed the concave-eared torrent frog, an unusual amphibian that makes its home at the base of Mt. Huangshan in Anhui Province, China. These animals were found living in an environment full of intense, broad spectrum ambient noise from the rushing creek and nearby waterfalls. Recordings of their calls revealed significant energy in the ultrasonic range, and examination of the frogs' anatomy showed a recessed tympanic membrane and a mammalian-like ear canal. It is likely these frogs faced selection pressure from their noisy habitat to increase the frequency of their calls and hearing to communicate effectively. In fact, Narins and his colleagues discovered another frog species in Borneo with similar ultrasonic vocalizations and a depressed tympanic membrane. The two species are not closely related, suggesting they independently evolved those characteristics in response to similar environmental pressures.

Narins' talk highlighted what would be a major theme of the meeting—noise and its effects on animal communication. Rachele Malavasi of the University of Carlo Bo, Italy, presented data that revealed songbirds in stable communities coordinated their chorusing to avoid signal masking. Cornell University’s Aaron Rice analyzed automatic recordings of marine acoustic communities off the shore of the southeastern United States and found evidence for acoustic niche partitioning between species that share acoustic space. These animals, and Narins' ultrasonic frogs, have adjusted or evolved solutions so they can still communicate amidst naturally occurring noise.

For other animals, problems arise when the noise is anthropogenic. Sandra Blumenrath (University of Maryland, College Park) explained how reverberant environments compromise detection and discrimination of communication sounds in songbird networks, and Jenelle Dowling (Cornell University) discussed how urban development results in structures with hard, impervious surfaces that, to wildlife, have unfamiliar absorptive and reflective properties.

Other researchers are examining the effects of anthropogenic noise on marine mammals. Christopher Clark of Cornell presented analyses showing the acoustic footprint of large shipping vessels was enormous, effectively “bleaching” large areas of endangered right whale habitat and significantly limiting their communication opportunities. The consequences of perturbing an acoustic community are not fully understood, especially in harder-to-observe marine environments. We currently do not know how shipping noise affects right whale movements and communication, but noise of this magnitude has the potential to interfere with foraging efficiency, mating opportunities, and possibly even survival.

Leila Hatch (Stellwagen Bank National Marine Sanctuary) addressed the need for changes in the way we try to abate anthropogenic noise in the ocean. She put forward that current noise management focuses on short-term, transient noise and has a heavy emphasis on marine mammals. Future noise management plans must pay attention to the cumulative noise footprints from multiple sources, consider ecologically relevant scales in both space and time, address chronic lower intensity noise sources, and include all wildlife (fish, invertebrates, etc.). Hatch called for methods for quantifying anthropogenic noise over large spatial and long temporal scales and assessing the effects of this noise on behavior of many species of marine animals, especially movement and communication.

Back on land, the situation is not much better. Kurt Fristrup of the National Park Service demonstrated how even the places humans designate as wild and protected are not immune from noise. The Park Service was established to support goals such as leaving wild areas unimpaired, ensuring superb environmental quality, making certain natural processes predominate, and preserving authentic landscapes. However, transportation noise is a major problem in national parks. Evaluation of the transportation networks within parks revealed motorcycles have a greater than 600 km acoustic footprint. Many aircraft routes fall over national parks, adding to the pervasive problem. The specific and long-term consequences of increasing transportation noise in national parks remain unknown, but some of the losses are already apparent. The wild inhabitants are losing active space in which to send and receive communication signals. The parks themselves are losing some of their wilderness character. Noise can even affect human visitors to the parks, by interfering with speech or sleep and disrupting those who wish to connect with nature. Fristrup ended his talk with hope, saying noise pollution in national parks could be "turned off as soon as we have the will to do so, and the benefit will be immediate."

Narins' presentation, and those that followed, supported the importance of continuing research into animal sounds. The Symposium provided an opportunity for scientists to gather and discuss why the study of acoustic communication by animals is needed not only to assess the effects of noise and contribute to conservation efforts, but also to learn about animal behavior and physiology, study the distribution and movement of animals, and estimate the density or population of some species.

Mary Bates is a freelance science writer based out of Boston, MA. She received her Ph.D. in psychology from Brown University, where she studied bat echolocation. Her research has appeared in *Science, Proceedings of the National Academy of Science*, and other journals. Her writing for popular audiences has been published by the American Association for the Advancement of Science, the Howard Hughes Medical Institute, and Harvard's Focus. You can read more of her work at www.marybatessciencewriter.wordpress.com.

Acoustical News

Elaine Moran
Acoustical Society of America
Melville, New York 11747

Professor Bridget Shield (l), Trevor Cox (r)

RWB Stephens Medal awarded to Bridget Shield

Professor Bridget Shield has been awarded the RWB Stephens Medal by the Institute of Acoustics (IOA) at the IOA's 2011 conference in Glasgow. The RWB Stephens Medal, named after the first President of the IOA, is awarded in odd-numbered years for outstanding contributions to acoustics research or education. The Institute of Acoustics, formed in 1974, is the UK's professional body for those working in acoustics, noise and vibration.

Bridget Shield is Professor of Acoustics at London South Bank University and President-Elect of the Institute of Acoustics. In the past few years her research has focused on the effects of noise and poor acoustics on children and teachers in primary schools. Bridget has many years' experience of teaching, research and consultancy in environmental and architectural acoustics. She has received many government research grants, and is the author of over 100 published papers. Her research interests have included prediction of industrial noise, community response to railway noise, concert hall acoustics, and annoyance caused by low frequency noise.

Professor Shield is a member of the Acoustical Society of America and in 2007 was elected an Honorary Fellow of the Institute of Acoustics. In 2011 she was also awarded the John Connell Lifetime Achievement Award from the Noise Abatement Society, recognizing her outstanding contributions to raising the profile of noise pollution as a critical environmental issue throughout her career

Per Bruel Gold Medal awarded to Mardi Hastings

Mardi C. Hastings, Professor at the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, was awarded the Per Bruel Gold Medal for Noise Control and Acoustics by the ASME (American Society for Mechanical Engineers) in November 2011. Dr. Hastings was cited "For research and international leadership in marine bioacoustics, particularly the increased understanding of effects of underwater noise on marine life and for research efforts leading to the mitigation of anthropogenic sound in the ocean." The Per Bruel Gold Medal for Noise Control and Acoustics was established in 1987 in honor of Dr. Per Bruel, who pioneered the development of sophisticated noise and vibration measuring and processing equipment. The medal recognizes eminent achievement and extraordinary merit in the field, including useful applications of the principles of noise control and acoustics to the art and science of mechanical engineering.

Mardi C. Hastings

Mardi Hastings received B.S. and M.S. degrees in Mechanical Engineering from The Ohio State University and a Ph.D., also in Mechanical Engineering, from the Georgia Institute of Technology. Her current research interests include fluid-structure interactions, effects of sound on the marine environment, and marine bioacoustics. She is coauthor, with Whitlow Au, of *Principles of Marine Bioacoustics*, (Springer-Verlag, 2008) and author of over 50 other publications.

Dr. Hastings has served on the National Academy of Sciences Study Panel on Ocean Noise and Marine Mammals (2001-02) as well as on various committees and boards of scientific organizations. She was a member of the Institute of Noise Control Engineering Board of Directors (2007-10), ASME Noise Control & Acoustics Division Chair (1998-99), Acoustical Society of America (ASA) Executive Council (2003-06), and Chair of the ASA Animal Bioacoustics Technical Committee (2000-03). She currently serves as President of the Acoustical Society of America (2011-12).

Mardi Hastings is a Fellow of the Acoustical Society of America and has received several awards and distinctions including the National Science Foundation Presidential Young Investigator Award (1988), Society of Automotive Engineers Ralph R. Teetor Educational Award (1993), The Ohio State University College of Engineering Lumley Research Award (1996), and the U.S. Federal Highway Administration Environmental Excellence Award (2005).

ASME is a not-for-profit membership organization that enables collaboration, knowledge sharing, career enrichment, and skills development across all engineering disciplines. It includes more than 120,000 members in over 150 countries worldwide.

Adnan Akay Receives Humboldt Research Award

Professor Dr. Adnan Akay, vice president of the Board of Trustees and chair of the Department of Mechanical Engineering at Bilkent University in Ankara, Turkey, has recently received a Humboldt Research Award. The Alexander von Humboldt Foundation in Germany grants these prestigious awards to researchers "whose fundamental discoveries, new theories, or insights have had a significant impact on their own discipline and who are expected to continue producing cutting-edge achievements in the future." Prof. Akay was elected to receive the award in recognition of his accomplishments to date in research and teaching.

Adnan Akay received B.S., M.M.E, and Ph.D. degrees from North Carolina State University. His research interests are in the areas of applied mechanics, vibrations and acoustics, noise control, tribology, and friction-induced sounds. As a Humboldt Research Award recipient, he will be invited to undertake an extensive research project in collaboration with specialist colleagues in Germany.

Adnan Akay joined Bilkent University on in 2009 as Vice President and the founding head of Mechanical Engineering Department. He moved to Bilkent from the U.S. National Science Foundation where he was the director of the Division of Civil, Mechanical and Manufacturing Innovation. Between 1992 and 2005, Dr. Akay was the head of the Mechanical Engineering Department at Carnegie Mellon University where he currently holds the position of professor. He has been recognized with several awards including the Per Brüel Gold Medal in Acoustics and Noise Control in 2005 from ASME. He is a Fellow of the American Society of Mechanical Engineers and the Acoustical Society of America.

Adnan Akay

Martin Klein

Martin Klein Receives Arnold O. Beckman Founder Award

Martin Klein was named recipient of the Arnold O. Beckman Founder Award by the International Society of Automation (ISA). The award was presented at the ISA Honors and Awards Gala, held 17 October 2011 in Mobile, Alabama. Mr. Klein was cited for the invention and development of the dual channel side scan sonar instrumentation that has opened the world's oceans for exploration, safe navigation, and underwater recovery. The Arnold O. Beckman Founder Award recognizes a significant technological contribution to the conception and implementation of a new principle of instrument design, development or application.

Martin Klein is an inventor and developer of the first commercial side scan sonar utilized for detection and mapping of lake and river beds and the ocean floor to the full known (7 miles) depth of the sea. Klein began his work on side scan sonar instrumentation in 1961 while a student at the Massachusetts Institute of Technology (MIT) and in 1968 founded his own company, Klein Associates, Inc. The Klein side scan sonar technology has been utilized to find most of the significant shipwrecks and sunken aircraft in the world, including the Titanic, USS Monitor, and the Mary Rose, and remains of the Space Shuttle Challenger to name a few. Today, the side scan sonar instrumentation is used by the U.S. government, corporations, research institutions, and marine archaeologists around the world to map ocean floors, lakes and river beds and to find objects of great interest and value.

Klein is the author of numerous publications and holds several marine technology patents. He is a member of the Acoustical Society of America and a Senior Life Member of ISA. He received a Bachelor of Science degree in electrical engineering (BSEE) from the Massachusetts Institute of Technology (MIT).

Founded in 1945, the International Society of Automation is a leading, global, nonprofit organization that is setting the standard for automation by helping over 30,000 worldwide members and other professionals solve difficult technical problems, while enhancing their leadership and personal career capabilities. ISA develops standards, certifies industry professionals, provides education and training, publishes books and technical articles, and hosts conferences and exhibitions for automation professionals. ISA is the founding sponsor of the Automation Federation.

Calendar of Meetings and Congresses

Compiled by the Information Service of the International Commission for Acoustics

2012

22-23 Feb.	Berlin, Germany. Berlin Beamforming Conference 2012. http://bebec.eu/
25-30 Mar	Kyoto, Japan. IEEE International Conference on Acoustics, Speech, and Signal Processing. http://www.icassp2012.com
19-22 Mar	Darmstadt, Germany. 38th German Annual Conference on Acoustics (DAGA2012). http://daga2012.de/
18 -20 April	Senlis, France. International Conference on Fan Noise, Technology, and Numerical Methods (FAN2012). http://www.fan2012con ference.org/
21-24 April	Sorrento, Italy. Noise and Vibration: Emerging Methods (NOVEM2012). http://www.novem2012.unina.it
23-27 April	Nantes, France. ACOUSTICS 2012-NANTES. http://www.acoustics2012-nantes.org/
13-18 May	Hong Kong, China. Acoustics 2012 Hong Kong. Joint meeting of the 163rd meeting of the Acoustical Society of America, 8th meeting of the Acoustical Society of China, 11th meeting of Western Pacific Acoustical Conference, and Hong Kong Institute of Acoustics. http://acoustics2012hk.org
21-24 May	Tokyo, Japan. 19th International Symposium on Nonlinear Acoustics (ISNA2012). http://www.isna19.com/index
10-13 June	Prague, Czech Republic. Euronoise 2012. http://www.euronoise2012.cz/
02-06 July	Edinburgh, UK. 11th European Congress on Underwater Acoustics. http://www.acua2012 .com
08-12 July	Vilnius, Lithuania 18th International Congress on Sound and Vibration (ICSV19) http://www.icsv19.org
22-27 July,	Porto, Portugal Symposium on Vibration and Structural Acoustics measurement and analysis in conjunction with 15th International Conference on Experimental Mechanics (ICEM15) http://paginas.fe.up.pt/clme/icem15/
19-24 Aug	Beijing, China. 23rd International Congrss of Theoretical and Applied Mechanics (ICTAM2012). http://www.ictam2012.org/
19-22 Aug	New York, NY, USA. Internoise 2012. http://www.internoise2012.com.
09-13 Sept	Portland, OR. USA. Interspeech 2012. http://interspeech2012.org
12-15 Sept	Granada, Spain. 30th European Conference on Acoustic Emission Testing (EWGAE) and 7th International Conference on Acoustic Emission (ICAE). http://2012.ewgae.eu/
12-15 Sept	Petrcane, Zadar, Croatia. 5th Congress of Alps-Adria Acoustics Association & 2nd Congress of Acoustical Society of Croatis (AAA2012). http://www.akustika.hr/had/kongrss
17-19 Sept	Leuven, Belgium. ISMA International Conference on Noise and Vibration Engineering (ISMA 2012). http://www.isma-isaac.be/conf/
21–23 Nov	Perth, Western Australia. 2012 Conference of the Australian Acoustical Society. http://www.acoustics.asn.au/joomla/acoustics-2012.html

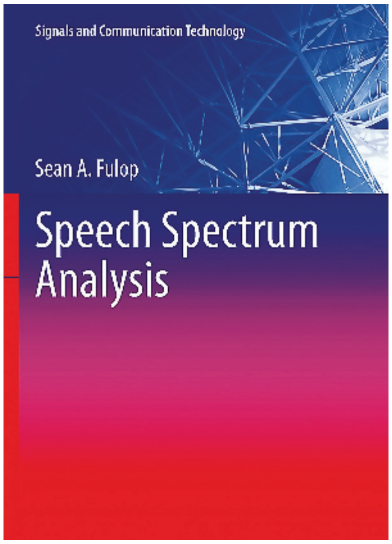
2013

26-31 Mar	Vancouver, Canada. 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). http://www.icassp2013.com
02-07 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013) http://www.ica2013montreal.org

Books and Publications

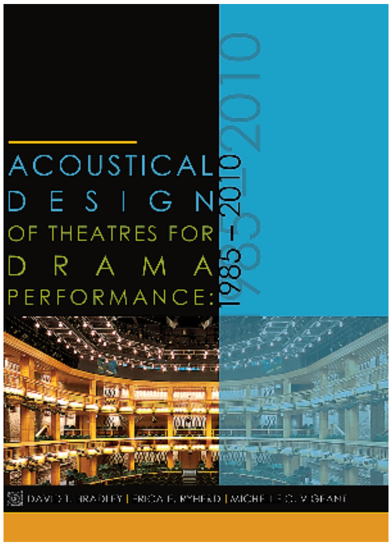
Dick Stern
1150 Linden Hall Road
Boalsburg, Pennsylvania 16827

Acoustics Today welcomes contributions for "Books and Publications." There is no charge for this service. Submissions of about 250 words that may be edited in MSWord or plain text files should be e-mailed to <acousticstoday@aip.org>. Cover graphics should accompany the text and must be at least 300 dpi. Please send the text and graphics in separate files.



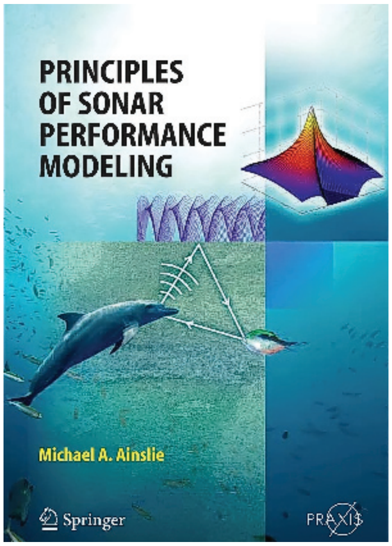
Title: *Speech Spectrum Analysis*
Author: Sean A. Fulop
Publisher: Springer-Verlag
ISBN: 978-3-642-17477-3
Pages: 219
Binding: Hardcover, ebook (PDF) available

The accurate determination of the speech spectrum, particularly for short frames, is commonly pursued in diverse areas including speech processing, recognition, and acoustic phonetics. With this book the author makes the subject of spectrum analysis understandable to a wide audience, including those with a solid background in digital signal processing and those without such background. In keeping with these goals, this is not a book that attempts to cover the material found in a general signal processing textbook. Some essentials of signal processing are presented in a preliminary chapter, but the concepts are presented in a generally understandable fashion as far as is possible. Throughout the book, the focus is on applications to speech analysis; mathematical theory is provided for completeness, but these developments are set off in boxes for optional reading by those with sufficient background. Other readers may proceed through the main text, where the key results and applications are presented in general heuristic terms, and illustrated with software routines and practical "show-and-tell" discussions of the results. Reference is made to the Praat speech analysis software to discuss many practical topics including spectrography and linear prediction. Additionally, MATLAB code to accompany the book is provided online, and if readers have the MATLAB package, they will be able to implement immediately routines for quadratic time-frequency distributions, reassigned spectrograms, and other procedures not normally included in turn-key software.



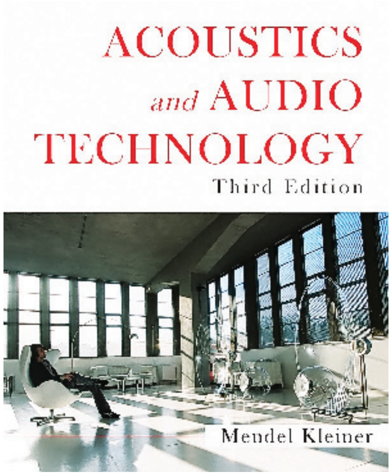
Title: *Acoustical Design of Theatres for Drama Performance: 1985–2010*
Editors: David T. Bradley, Erica E. Ryherd, and Michelle C. Vigeant
ISBN 978-0-9846084-5-4
Publisher: Acoustical Society of America
Binding: Hardcover
Pages: 334

The acoustic environment is paramount in any space that houses dramatic performance. Thornton Wilder said, "The unencumbered stage encourages the truth operative in everyone. The less seen, the more heard." This new book takes an inside look at the acoustical design of 130 drama theatres from around the world. It is a compilation of drama theatres that have been designed during the 25-year period from 1985 to 2010. Top acoustical consulting firms from around the world contributed examples of their work, including images, acoustical data, and descriptions of the theatres. The book is a valuable educational resource that provides introductions from leading theatre consultants and a prominent artistic director, an overview of key aspects involved in the acoustic design of drama theatres, and a comprehensive glossary of common theatre acoustics terminology. Further, the book is a useful reference, as the contributed theatres are categorized according to theatre type, and are indexed by consulting firm and by geographic location.



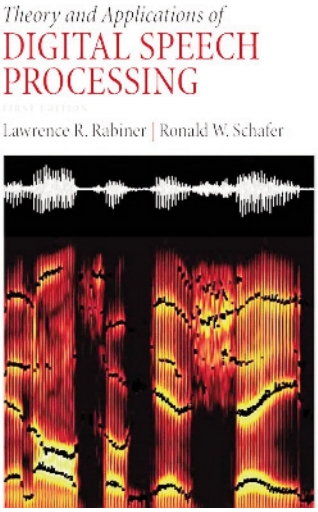
Title: *Principles of SONAR Performance Modeling*
Author: Michael A. Ainslie
Publisher: Springer
Binding: Hardcover
Pages: 828
ISBN-13: 978-3540876618

Human beings are used to using built-in optical sensors—our eyes—to build an accurate picture of our immediate surroundings, and when we wish to look beyond the visible horizon we turn to radio waves to do the same job. In water, neither visible light nor radio carries more than a few meters, whereas low frequency sound can travel tens or even hundreds of kilometers, making sonar the sensor of choice for underwater navigation, oceanography, or the detection of underwater objects. *Principles of SONAR Performance Modeling* opens with a description of the pioneering efforts of Pierre and Jacques Curie, who discovered piezoelectricity, Paul Langevin, who demonstrated underwater echolocation during World War I, and other giants of the twentieth century such as Ernest Rutherford, Léon Brillouin and Maurice Ewing, in making sonar and its applications a reality. Traditionally considered a branch of engineering, sonar performance modeling is treated here with a physicist's perspective, bringing together oceanography, acoustics, signal processing and detection theory in one volume. Separate chapters describe the characteristic physical, chemical and biological signature of the oceans, acoustic reflection from the oceans' boundaries and their contents, propagation, noise and reverberation modeling, beamforming and matched filter processing, the hearing capabilities of marine mammals, and the fundamentals of statistical detection theory for fluctuating and non-fluctuating signals. The cornerstone is a derivation from physical principles of the sonar equations, which are applied to examples of SONAR—both man-made and biological.



Title: *Acoustics and Audio Technology, Third Edition*
Author: Mendel Kleiner
Publisher: J. Ross Publishing, Inc.
ISBN: 978-1-60427-052-5
Pages: 480
Binding: Softcover

Acoustics and Audio Technology, Third Edition covers the physical background to and the mathematical treatment of propagation, generation, and radiation of sound as well as hearing, architectural acoustics, and audio. Examples from audio engineering are used to illustrate acoustics principles thereby drawing students into acoustics by using their interest in audio technology. *Acoustics and Audio Technology* starts with a chapter on one-dimensional and spherical waves, impedance, reflection, and power radiation fundamentals. Next, the physiology and psychoacoustics of hearing are explained. Room acoustics are covered from the viewpoint of geometrical, statistical, and physical acoustics as well as hearing in rooms. Sound radiation and sound isolation concepts are included as well. *Acoustics and Audio Technology* covers the electroacoustic topics of microphones, cartridges, loudspeakers, and headphones as well as the basics of digital sound reproduction—there is a chapter on audio systems and measurement. *Acoustics and Audio Technology* is an introductory text for students of sound and vibration as well as electrical and electronic engineering, civil and mechanical engineering, computer science, signals and systems, and engineering physics. A fundamental knowledge of basic engineering mathematics and physics is assumed. Problems are included at the end of the chapters and a solutions manual is available to instructors.



Title: *Theory and Applications of Digital Speech Processing*
Authors: Lawrence R. Rabiner and Ronald W. Schafer
Publisher: Pearson/Prentice Hall
ISBN: 978-0-13-603428-5
Pages: 1042
Binding: Hardcover

This book is a completely new and up-to-date treatment of digital speech processing by the authors of *Digital Processing of Speech Signals*, which has been widely used both as a course textbook and a technical reference for over 30 years. The earlier book had a long life due to its structure, which emphasized the fundamentals of speech production and the fundamentals of digital representations of speech signals. The present book continues this emphasis on fundamentals, but adds two more layers to the "speech stack," an organizing principle that also includes new emphasis on algorithms based on the fundamental principles and on current applications of digital speech processing (DSP). Chapters 1-5, which cover fundamentals of DSP, speech production, audition, and the acoustics of speech production, comprise the first layer. The core of the book is Chapters 6-9, which cover basic time-dependent digital representations including time-domain, frequency domain, cepstrum, and linear prediction representations of speech signals. Chapter 10 illustrates how the basic digital speech representations can be combined with statistical models and heuristics to create algorithms for extracting information such as pitch and formants from sampled speech signals. The fourth layer of the presentation is comprised of Chapters 11-14 which provide up-to-date discussions of speech and audio coding, text-to-speech synthesis, and automatic speech recognition. In each chapter, carefully explained examples, associated graphics, and an extensive set of homework problems aid in understanding both the mathematics and important concepts. The implementation of new speech processing concepts is illustrated by MATLAB® code and by an extensive collection of MATLAB-based exercises. The book also offers access to a website that provides MATLAB code and databases that support the exercises.

Instrumentation

Dick Stern
1150 Linden Hall Road
Boalsburg, Pennsylvania 16827

Acoustics Today welcomes contributions for "Instrumentation." There is no charge for this service. Submissions of about 250 words that may be edited in MSWord or plain text files should be e-mailed to <acousticstoday@aip.org>. Graphics must be at least 300 dpi. Please send the text and graphics in separate files.

G.R.A.S. Sound & Vibration, a globally renowned designer and manufacturer of precision acoustic measurement solutions, has announced the North American market introduction of the G.R.A.S. 45CB Acoustic Test Fixture According to ANSI S12.42.

Offering high-reliability performance over a wide dynamic range, the 45CB is expressly designed to meet a growing industry need for commercially available acoustic test fixtures (ATF) that can help meet or exceed the ANSI/ASA S12.42 standard, "Methods for the Measurement of Insertion Loss of Hearing Protection Devices in Continuous or Impulsive Noise Using Microphone-in-Real Ear or Acoustic Test Fixture Procedures."

The 45CB provides sufficient acoustic damping for objective measurements of both high-level continuous impulsive noise and the attenuation-related insertion loss encountered with active and passive hearing protection devices such as earplugs, earmuffs and safety helmets. The sturdy, high-temperature and humidity resistant construction of the 45CB is ideally suited for outdoor measurement environments, as well as simulated real-life conditions of test sites, vehicle interiors, aircraft and other areas. In addition, Ear Simulators with built-in ¼″ pressure microphones ensure a measurement system that can rapidly and correctly account for impulse peaks produced by heavy industrial and agricultural equipment and guns. They can therefore be used with any type of test signal or real-life noise source, including environments where use of human test subjects is simply not possible, due to high noise levels or expressed requirements for objective statistical data.

The 45CB fulfills requirements for real-life objective hearing protection measurements over a wide dynamic range, including self-insertion loss measurements of greater than 70 dB over a wide frequency range. A peak dynamic level of 174 dB allows for realistic testing levels. Levels of up to 190 dB can be measured and calculated accurately based on closed ear measurements, combined with measurement of the transfer function of the open ear (TFOE). Ear canal extension dimensions, rubber coating and appropriate shore hardness make it possible to further measure the insertion loss of insertion plugs and other elastomeric materials at actual human body temperatures.

When testing calls for the use of continuous noise, broadband random noise is recommended, as outlined in ANSI S12.42. Measurements can also be conducted using recorded signals of the actual usage environment for intended hearing protection devices, such as vehicle noise and similar sources. When measuring with continuous noise, a variety of G.R.A.S. free-field and random incidence measurement microphones may be used for reference.

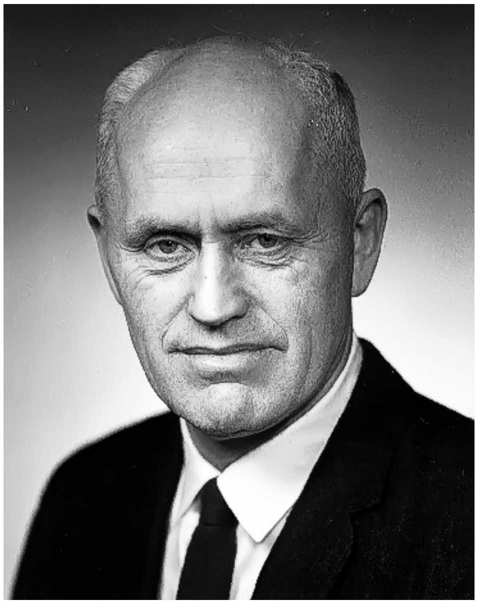
All G.R.A.S. products are made of high-quality materials that will ensure life-long stability and robustness over their useful service life. The G.R.A.S. 45CB Acoustic Test Fixture According to ANSI S12.42 is delivered fully assembled. An individual test certificate is included with each 45CB. All G.R.A.S. products are calibrated in a controlled laboratory environment using traceable calibration equipment, with subsequent annual calibrations recommended. In addition, the G.R.A.S. 45CB Acoustic Test Fixture According to ANSI S12.42 is offered with a two-year comprehensive product warranty. For more information, visit www.ansihead.com.

Passings

Dick Stern
1150 Linden Hall Road
Boalsburg, Pennsylvania 16827

Wesley L. Nyborg

1917 – 2011



Wesley L. Nyborg, a biophysicist who made profound contributions to our understanding of ultrasound, died on September 24, 2011 at age 94. He obtained his A.B. in 1941 from Luther College in Decorah, Iowa and his M.S. and Ph.D. in physics at Pennsylvania State University in 1944 and 1947, respectively. He was an Assistant and Associate Professor at Brown University from 1950 to 1960, and since 1960 was a Professor and then an Emeritus Professor of physics at the University of Vermont.

Wes's career was devoted to biophysical acoustics, first at low frequencies and since the early 50's primarily at ultrasonic frequencies of biomedical interest. His early work established a basis for much of our current knowledge of non-thermal mechanisms by which ultrasound interacts with biological materials. He developed the theoretical basis of acoustic streaming, and he extended his work to investigating ultrasonically induced fluid flow and particle movements. His research included important problems in sound propagation in scattering and absorbing media, heat production in an ultrasound beam, ultrasonic cavitation, and the mechanisms responsible for biological effects of ultrasound. Wes made many contributions to medical ultrasound safety and therapy.

During his career as an educator, Wes guided and inspired many students of physics and biophysics. He developed a unique course of study in biophysics which he taught for many years and which resulted in publication of the textbook *Intermediate Biophysical Mechanics* in 1975. Later, he co-edited a textbook with Marvin Ziskin in 1985 entitled *Biological Effects of Ultrasound*. In 1984 he was named "University Scholar in Physical Sciences" at the University of Vermont.

Wes was highly regarded by scientists and organizations. He was a Fellow of the Acoustical Society of America, the American Institute of Ultrasound in Medicine, and the American Association for the Advancement of Science. He served on the Editorial Boards of the *Journal of Biological Physics, Ultrasound in Medicine and Biology, Ultrasonics,* and *Clinics in Diagnostic Ultrasound*. He was the author of many chapters in scientific books and published many papers in scholarly journals.

Wes was active in the American Institute of Ultrasound (AIUM) where he was the Chairman of the Bioeffects Committee for a number of years. He played a major role in producing a number of official statements for AIUM pertaining to ultrasound safety. He continued the safety effort within the NIH Diagnostic Radiology Study Section during his tenure with that important federal advisory panel. He received the AIUM Presidential Recognition Award in 1977, its Pioneer Award in 1985, and its Fry Memorial Lecture Award in 1990. Wes was also active in the Acoustical Society of America and served on the Executive Council from 1965-1968. He was awarded the ASA Interdisciplinary Silver Medal in Physical Acoustics and Bioresponse to Vibration in 1990. From 1992-2002, Wes was chairman of ASA Standards Working Group 22, which produced the ANSI Technical Report on Bubble Detection and Cavitation Monitoring.

For 22 years Wes was the Chairman of the National

Acoustics Today accepts contributions for "Passings." Submissions of about 250 words that may be edited in MSWord or plain text files should be e-mailed to AcousticsToday@aip.org. Photographs may be informal, but must be at least 300 dpi. Please send the text and photographs in separate files.

Council of Radiation Protection Scientific Committee No. 66 devoted to establishing guidelines for the safe use of ultrasound in medicine. Three volumes were published by that committee: 1) Biological Effects of Ultrasound: Mechanisms and Clinical Implications, in 1983; 2) Exposure Criteria for Medical Diagnostic Ultrasound: I. Criteria Based on Thermal Mechanisms, in 1992, and 3) Exposure Criteria for Medical Diagnostic Ultrasound: II. Criteria Based on All Known Mechanisms, in 2002. These volumes are considered the most authoritative documents in this field.

Douglas L. Miller, PhD
University of Michigan
Ann Arbor, MI 48109

ASA has learned of the deaths of the following members:

P. E. Doak
Ronald L. McKay
Paul B. Oncley
Stannard M. Potter
Bertram Scharf
Maurice Sevik
Edgar M. Villchur
Jason T. (Tic) Weissenburger

Statement of Ownership, Management and Circulation

1. *Publication title:* ACOUSTICS TODAY
2. *Publication number:* ISSN1557-0215
3. Filing Date: 1 October 2011
4. *Issue frequency:* Quarterly
5. *No. of issues published annually:* 4
6. *Annual subscription price:* $5
7. *Complete mailing address of known office of publication:* 2 Huntington Quadrangle, Suite 1NO1, Melville, NY 11747-4502
8. *Complete mailing address of headquarters or general business offices of the publisher:* 2 Huntington Quadrangle, Suite 1NO1, Melville, NY 11747-4502
9. *Full names and complete mailing addresses of publisher, editor and managing editor:*
   Publisher: Acoustical Society of America, 2 Huntington Quadrangle, Suite 1NO1, Melville, NY 11747-4502
   Editor: Richard Stern, Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502
   Managing editor: none
10. *Owner (if the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a Partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address:* Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502
11. *Known bondholders, mortgagees and other security holders owning or holding 1 per cent or more of total amount of bonds, mortgages or other securities:* None
12. *The purpose, function and nonprofit status of this organization and the exempt status for Federal income tax purposes:* The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes has not changed during preceding 12 months.
13. *Publication name:* ACOUSTICS TODAY
14. *Issue date for circulation data below:* April 2011
15. *Extent and nature of circulation:*
    A. *Total number of copies (net press run)*
       Average 9,032 April 9,558
    B. *Paid Circulation(By mail and Outside the Mail)*
       1. Subscriptions Stated on PS Form 3541
          Average 8,583 April 9,088
       2. *Mailed In-County Paid Subscriptions*
          Average 0 April 0
       3. Paid Distribution Outside the Mails
          Average 0 April 0
       4. Paid Distribution by Other Classes of Mail Through the USPS
          Average 8,583 April 9,088

    C. *Total Paid Distribution (Sum of 15b. 1-4)*
       Average 8,583 April 9,088
    D. *Free or Nominal Rate Distribution (by Mail and Outside the Mail)*
       1. *Free or Nominal Rate Outside-County Copies included on PS Form 3541*
          Average 30 April 31
       2. *Free or Nominal Rate In-County Copies Included on PS Form 3541*
          Average 0 April 0
       3. *Free or Nominal Rate Copies Mailed at Other Classes Through the USPS*
          Average 0 April 0
       4. *Free or Nominal Rate Distribution Outside the Mail*
          Average 0 April 0
    E. *Total Free or Nominal Rate Distribution*
       Average 30 April 31
    F. *Total Distribution*
       Average 8,613 April 9,119
    G. *Copies Not Distributed*
       Average 419 April 439
    I. *Total*
       Average 9,032 April 9,558
    J. *Percent paid*
       Average 99.65% April 99.66%

I certify that all information furnished above is true and complete.
Dick Stern, Editor

Errata

Dick Stern

1150 Linden Hall Road
Boalsburg, Pennsylvania 16827

It has been brought to my attention that a number of typographical errors occurred in the July 2011 issue of *Acoustics Today*. In the Tiemann *et al.* article on Animal Bioacoustics (page 35), the word "censuring" was used to describe underwater acoustic methods in studies of animals. The word should have been "censusing" (taking a census). In addition, one of the figures was left out and its caption was used incorrectly on another figure. I apologize for the any confusion that might have occurred and listed below are the correct figures, with their correct captions.

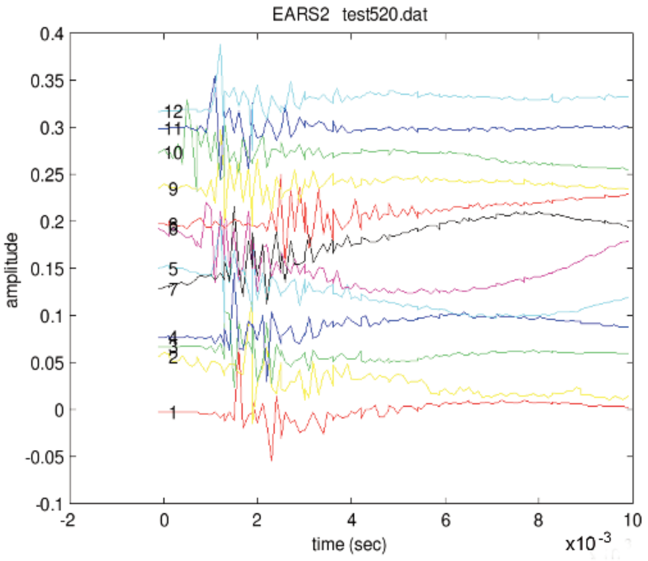


Fig. 8. Amplitude versus time for all the clicks in one coda from Fig. 3 plotted. Amplitude in each click is offset for clarity.

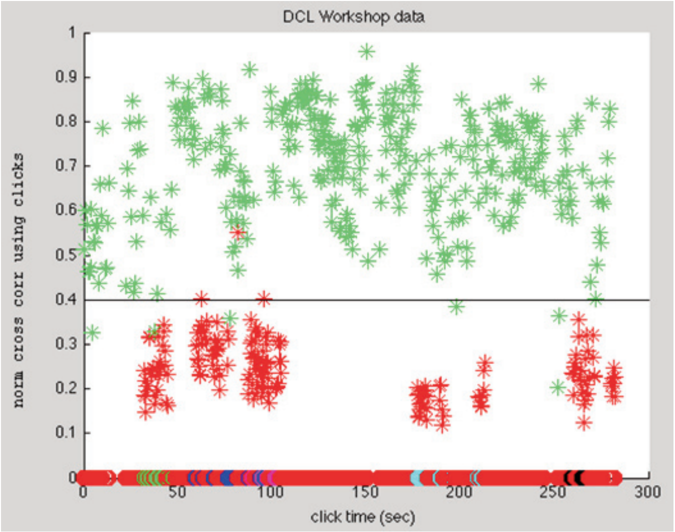


Fig. 9. Normalized cross correlation value plotted versus click time. Christopher Tiemann click train analysis results indicated by red and green star colors. Click of speaking whale unchanged from previous click colored green, and click of speaking whale changed from previous click colored red, summarized as green: CT class(i)=class(i-1); red: CT class(i) not equal to CT class(i-1). For more than 98.5% of the clicks, green stars are above the threshold and red stars are below.

Business Directory

Creating a quieter environment since 1972

DESIGN & SURVEY FIELD TESTING

- Industrial Noise Control
- Auditoriums & Music Halls
- Classroom & Education Facilities
- HVAC Mechanical Noise
- Multifamily Structures
- Transportation Noise
- Seismic Vibration Surveys
- Building Acoustics
- RT60, C80, D50. G
- ANSI 12.60
- AMCA, AHRAE, ISO
- ASTM ASTC, AIIC
- E966, HUD, FAA
- Scientific, Residential

Sound Power: OEM Acculab Reference Sound Source //////

Angelo Campanella P.E., Ph,D., FASA
3201 Ridgewood Drive Columbus(Hilliard), OH 43026-2453
614-876-5108 // cell = 614-560-0519
a.campanella@att.net//www.CampanellaAcoustics.com

Campanella Associates
CONSULTANTS IN ACOUSTICS

HGC ENGINEERING
noise · vibration · acoustics

Consulting Engineers specializing in Acoustics, Noise and Vibration

HOWE GASTMEIER CHAPNIK LIMITED

Enhancing where people live, work and play through the application of the principles of acoustical engineering.

Mississauga, Ontario
P: 905-826-4044
F: 905-826-4940
www.hgcengineering.com

ACOUSTICS TODAY BUSINESS DIRECTORY

**Cost-effective advertising for consultants and service providers.
Each space offers 1/8 page business card size ad.**

FOR ADVERTISING RESERVATIONS AND INFORMATION

Please contact:

DEBORAH BOTT
Advertising Sales Manager

American Institute of Physics
Two Huntington Quadrangle
Suite 1NO1, Melville, NY 11747
Tel: **(800) 247-2242** or **(516) 576-2430**
Fax: **(516) 576-2481** Email: **dbott@aip.org**
To view a pdf of the media kit, visit : **http://scitation.aip.org/AT**

Classified Advertisements

ACOUSTICS TODAY RECRUITING AND INFORMATION CLASSIFIED ADVERTISEMENTS:

Positions Available/Desired and Informational Advertisements may be placed in ACOUSTICS TODAY in two ways: — display advertisements and classified line advertisements.

Recruitment Display Advertisements:

Available in the same formats and rates as Product Display advertisements. In addition, recruitment display advertisers using 1/4 page or larger for their recruitment display ads may request that the text-only portion of their ads (submitted as an MS Word file) may be placed on ASA ONLINE — JOB OPENINGS for 2 months at no additional charge. All rates are commissionable to advertising agencies at 15%.

Classified Line Advertisements:

2012 RATES: Positions Available/Desired and informational advertisements

One column ad, $35.00 per line or fractions thereof (44 characters per line), $185 minimum, maximum length 40 lines; two-column ad, $46 per line or fractions thereof (88 characters per line), $320 minimum, maximum length 60 lines. (**Positions** desired ONLY – ads from individual members of the Acoustical Society of America receive a 50% discount)

Submission Deadline: 1st of month preceding cover date.

Ad Submission: E-mail: Christine DiPasca at cdipasca@aip.org; tel. : (516) 576-2434. You will be invoiced upon publication of issue. Checks should be made payable to the Acoustical Society of America. Mail to: Acoustics Today, Acoustical Society of America, 2 Huntington Quadrangle, Suite 1NO1, Melville, NY 11747. If anonymity is requested, ASA will assign box numbers. Replies to box numbers will be forwarded twice a week. Acoustics Today reserves the right to accept or reject ads at our discretion. Cancellations cannot be honored after deadline date.

It is presumed that the following advertisers are in full compliance with applicable equal opportunity laws and, wish to receive applications from qualified persons regardless of race, age, national origin, religion, physical handicap, or sexual orientation.

Index to Advertisers

Advertising Sales & Production

For advertisement reservations or further information please contact:

Deborah Bott
Advertising Sales Manager
Acoustics Today
c/o AIP Advertising Department
2 Huntington Quadrangle
Suite 1NO1
Melville, NY 11747
Ph: (800) 247-2242
Ph: 516-576-2430
Fax: 516-576-2481
Email: dbott@aip.org

For advertisement production inquiries, please contact and send files to:

Christine DiPasca
Sr. Advertising Production Manager
Acoustics Today
c/o AIP Advertising - Production Operations
2 Huntington Quadrangle
Suite 1NO1
Melville, NY 11747
Ph: 516-576-2434
Fax: 516-576-2481
Email: cdipasca@aip.org

To view a pdf file of the Media Kit visit:
http://scitation.aip.org/journals/doc/ASALIB-home/corp/pdf/ATmediakit.pdf