VERTEBRATE VOCAL PRODUCTION *IN VIVO* AND *IN SILICO*[1]

TECUMSEH W. FITCH

In recent years, despite great gains in our understanding of animal sound perception, there has been a growing realization that an important component of animal vocal communication has been neglected: vocal production (the mechanisms by which animal sounds are made). Indeed, although acoustic communication logically begins with sound production, a cursory glance in any of the leading journals will show that research labeled "bioacoustics" typically concerns only the sound-transmission and perception side of the vocal communication equation. This lack is made more salient by the fact that many if not most of the sounds that the auditory system is "designed" to hear are created by other animals, often of the same species (as opposed to vision, where nonbiological stimuli play a role equally as important as visual stimuli generated by conspecifics). Part of the reason for overlooking vocal production has been historical: only recently have the tools (both intellectual and technological) necessary to study vocal production become available. But part of the reason is systemic: vocal production is a highly interdisciplinary topic, encompassing branches of physics, physiology, neuroscience, ethology, and evolutionary biology, and it is not always easy to bridge the terminological and conceptual gaps between experts in these different disciplines.

In April 2003, the Wissenschaftskolleg sponsored a highly successful interdisciplinary workshop on vertebrate vocal production, organized by Fellow Tecumseh Fitch with the support of Britta Cusack and Reinhart Meyer-Kalkus. This workshop brought together two leading physicists working on human vocal production (speech and singing) with some of the top biologists currently studying sound production in animals (mammals, birds, and frogs). The importance of interdisciplinary collaboration is perhaps best illustrated by an example. Vocal anatomists deal with extremely complex three-dimensional systems, and it is quite difficult, based on anatomy alone, to derive clear conclusions about vocal produc-

1 Seminar, sponsored by and held at the Wissenschaftskolleg on April 9–12, 2003.

tion. Similarly, bioacousticians can analyze recordings of sounds, but it is often not even clear where a particular sound is being produced anatomically, much less how this production system is controlled. For instance, birds possess a complex specialized organ, termed the syrinx, which lies deep in the chest and generates bird song. Because of its small size and inaccessible location, birdsong researchers have had to make educated guesses, based on the anatomy of the syrinx, about which of the many membranes in the syrinx actually generate sound. Recently, endoscopic technology had advanced to the level where it was finally possible to directly view the syrinx of a singing bird *in vivo*. Using a medical imaging system designed to snake through human arteries into the heart, conference participants Ole Larsen and Franz Goller passed an endoscope down the trachea of an anesthetized bird and became the first humans to actually witness the vibrating membranes involved in birdsong. The sound source turned out to be a completely different structure than those previously supposed responsible!

The first and most basic goal of the conference was to share information across disciplines, so each participant gave a detailed introduction to his subject area. The second and most exciting goal was to use computer modeling as a basis for forging a common language that all of these disciplines can use to communicate. Computer simulation has become one of the most important tools available to natural scientists seeking to understand complex systems. Whether in physics, geology, meteorology, or biology, many questions at the current limits of our scientific understanding are far too complex to be modeled by equations simple enough to be solved by inspection. Instead, scientists use these equations to make models, at varying levels of complexity, which can then be "run" on a computer. By varying the many parameters of such an *in silico* model and then observing the outcomes, scientists using simulations can thus both increase their understanding of the system and generate predictions about what is and is not crucial. These in turn serve as indicators for biologists about what to look at and measure more carefully *in vivo*. Just as in theoretical and experimental physics, these two enterprises should form complementary, interdependent components of a successful research program. Unfortunately, though, the computational complexities of computer modeling have rendered this the domain of physicists and engineers, while biologists typically have their hands full with the anatomy, physiology, and neural control side of the problem. Thus an important desideratum for this conference was to begin to create an integrated international community of scientists from both sides, with the goal of making adequate computer models of the vocal production system as the "glue" to bind them together. While the immediate goal of computer models is increased understanding of vocal production, ultimately, models of sound production in other animals allow a scientist to flexibly create realistic animal sounds. Such synthetic calls can then be used in playback experiments in the field to improve understanding of the meanings of animal sounds (as I have done in my own field work), and thus greatly deepen our understanding of animal communication.

At the heart of our interdisciplinary endeavor was physicist Hanspeter Herzel (Institute of Theoretical Biology, Humboldt University, Berlin), a leading light in contemporary speech acoustics and a pioneer in understanding the role of nonlinear dynamics in vocal production. Herzel and his colleagues (including Patrick Mergell and Jürgen Neubauer) have worked for several years with the author to develop models of the mammalian larynx, and several papers have already come out of this collaboration (which has been supported by the Wissenschaftskolleg over several years). Herzel's approach, typical for a physicist, is to build gradually from what is understood and thus to start with simple models that can be clearly and intuitively understood (indeed the vocal fold model Herzel developed with Ina Steinecke has become one of the standard models in the field of speech science). Taking such simple human models and adding components that are seen in animals has been a productive way to better understand animal vocal production. The second major contributor to the *in silico* component of the conference was Ingo Titze, Director of the National Center for Voice and Speech in Denver. Titze is the world's leading expert on the physiology of singing and an accomplished singer himself, as well as one of the first scientists to develop complex computer models of the human vocal folds. Trained in engineering, Titze's traditional approach has been the complement of Herzel's: to make as few simplifying assumptions as possible and to model the vocal production system in great detail. Together, these two scientists thus represented the two poles of approaches to computer simulation of the voice.

The workshop officially started with an evening colloquium, open to the public, by Tecumseh Fitch on the evolution of human language. Most of the conference participants arrived for this and joined the Fellows for cocktails afterwards. The *in vivo* component of the conference was launched the next day with an introduction to the evolution of the vertebrate vocal system by Tecumseh Fitch. This was followed by a comprehensive introduction to amphibian vocalization and vocal production by frog acoustics expert Andreas Elepfandt (Institute of Biology, Humboldt University, Berlin). An introduction to the anatomy underlying vocal production in mammals, along with a survey of the many fascinating modifications of mammalian vocal anatomy, was given by anatomist Roland Frey

(Institute for Zoo Biology and Wildlife Research, Berlin). Finally, a detailed overview of vocal production systems in nonhuman primates was given by Uwe Jürgens (German Primate Center, Göttingen), who provided a masterful introduction to the physiology of primate vocalization, from respiratory muscles all the way up to cortical control. This was followed by Thursday night dinner with the other Wiko Fellows.

The second day started with a detailed and illuminating introduction to voice modeling and voice models by Hanspeter Herzel. Thereafter the session focused on birdsong and included two of the leading biologists currently studying sound production in birds: Franz Goller (Institute of Biology, University of Utah) and Ole Larsen (Institute of Biology, Odense). (A German expert on avian vocal and respiratory anatomy, Hans-Reiner Duncker, was unable to participate due to illness.) These experts gave a comprehensive overview of what is known about bird vocal production, much of it thanks to their own efforts, and showed stunning videos of the vibrating syrinx in action. They were joined by the youngest participant, Ph.D. student Coen Elemans (Department of Animal Sciences, University of Wageningen) who described new results derived from a physical (rather than computer-simulated) model of the avian vocal tract.

The conference ended with a wrap-up discussion highlighting outstanding problems and sketching out a road map for future modeling efforts (work already underway in Herzel's lab). The most exciting immediate development was the proposal of a hypothesis on the function of a membrane (the medial tympaniform membrane) long suspected to play a role in birdsong, based on an analogy with the human falsetto singing voice discussed by Ingo Titze, and a planning of two distinct modeling approaches to test this hypothesis. More generally, this intense three-day workshop helped to resolve outstanding issues concerning the usefulness of simple vs. complex models and the kinds of data that biologists need to collect to support future computational efforts. It is now quite clear that the study of animal sounds has reached a watershed moment and that, with concerted interdisciplinary efforts, our understanding of how animals make sounds and what they mean will be considerably deepened in the coming years. This is valuable not just due to the intrinsic interest of the subject, but also because it provides crucial background for understanding the evolution of speech and language in humans. It is a testament to the Wiko's understanding of the importance of interdisciplinary collaboration that they made this conference possible and created the conditions under which it could be such a success.

Participants

Coen Elemans, Department of Animal Sciences, University of Wageningen
Andreas Elepfandt, Institute of Biology, Humboldt University, Berlin
Tecumseh W. Fitch, Wissenschaftskolleg zu Berlin and School of Psychology, University of St. Andrews
Franz Goller, Department of Biology, University of Utah
Roland Frey, Institute for Zoo Biology and Wildlife Research, Berlin
Hanspeter Herzel, Institute for Theoretical Biology, Humboldt University, Berlin
Uwe Jürgens, German Primate Center, Göttingen
Ole Larsen, Institute of Biology, University of Odense
Ingo Titze, National Center for Voice and Speech, Denver