



WHAT JUMBO TELLS DUMBO

By Alex Antunes

WHAT DO ELEPHANTS AND HUMANS HAVE IN COMMON? ACCORDING TO MARQUETTE UNIVERSITY'S PATRICK CLEMINS, INDIVIDUALS IN BOTH SPECIES HAVE DISTINCT

voices, and we can train computers to recognize 85 percent of what they say. Running preidentified elephant sounds into a PC-based hidden Markov model (HMM) algorithm trains that program to then recognize and classify further elephant recordings. In short, combine a herd of six elephants, a digital tape deck, and a basic workstation running an HMM, and you get the start of a great conversation.

Clemins' bioacoustics team at Marquette hooked up with Disney's Animal Kingdom to test this acoustic analysis technology (see the "Dr. Dolittle, I Presume?" sidebar for why they chose elephants). Being able to identify which particular elephant is talking (the speaker model) and which sound or part of elephant speech an elephant is using (the vocalization type) are both important steps toward elephant speech translation. The team's HMM software can uniquely identify individual elephants in a herd by their distinct rumbles with more than 88 percent accuracy. "I don't think the individual speaker-identification task would be possible for humans to perform," says Kirsten Leong, a research associate at the Animal Kingdom.

To identify the vocalization type, the HMM software has a 94 percent success rate in immediately classifying new recordings into categories previously defined by the human researchers. In comparison, when researchers learned to identify vocalization types, "it usually took a couple weeks of intensive listening to tapes before we reached 85 percent," Leong estimates.

The software's accuracy not only bests human efforts for elephant speech, but it's also comparable with the same task—using the same approach and software—done on human speech.

Learning Elephantese

When communicating, elephants often group quietly for several minutes, then emit clusters of rumbles, trumpets, croaks, revs, snorts, and other sounds. In all, the elephants make from one or two to as many as 20 distinct vocalizations in a single minute. Then, they return to silence.

"Elephants almost always 'talk' at the same time," Leong notes. Although understanding "elephantese" will require that we understand social context, language, and grammar, the pure acoustic task of identifying vocalizations and individual speakers is a necessary first step.

Leong says that vocal, visual, tactile, and hormonal communication are all part of elephant "speech." Additionally, their vocal range is 7 to 200 Hz, compared to the human ear's range of 20 Hz to 20 KHz, so humans can't hear much of an elephant's speech.

"Over 85 percent of the elephant calls we recorded were rumbles, which ... have the added complication of containing infrasonic components [which humans can't hear],"



Figure 1. Elephants at Walt Disney World's Animal Kingdom. Using software running a hidden Markov model, researchers can identify which elephant is "speaking."

Computing

in SCIENCE & ENGINEERING

EDITOR IN CHIEF

Norman Chonacky
cise-editor@aip.org

ASSOCIATE EDITORS IN CHIEF

Denis Donnelly, Siena College
donnelly@siena.edu

Douglass E. Post, Los Alamos Nat'l Lab.
post@ieee.org

John Rundle, Univ. of California at Davis
rundle@physics.ucdavis.edu

Francis Sullivan, IDA Ctr. for Computing Sciences
fran@super.org

EDITORIAL BOARD MEMBERS

Klaus-Jürgen Bathe, Mass. Inst. of Technology, kjb@mit.edu

Antony Beris, Univ. of Delaware, beris@che.udel.edu

Michael W. Berry, Univ. of Tennessee, berry@cs.utk.edu

John Blondin, North Carolina State Univ.,
john_blondin@ncsu.edu

Michael J. Creutz, Brookhaven Nat'l Lab., creutz@bnl.gov

George Cybenko, Dartmouth College, gvc@dartmouth.edu

Jack Dongarra, Univ. of Tennessee, dongarra@cs.utk.edu

Rudolf Eigenmann, Purdue Univ., igenman@ecn.purdue.edu

David Eisenbud, Mathematical Sciences Research Inst.,
de@msri.org

William J. Feiereisen, Los Alamos Nat'l Lab, ill@feiereisen.net

Geoffrey Fox, Indiana State Univ., gcf@grids.ucs.indiana.edu

Sharon Glotzer, Univ. of Michigan, sglotzer@umich.edu

Anthony C. Hearn, RAND, hearn@rand.org

Charles J. Holland, Office of the Defense Dept.,
charles.holland@osd.mil

M.Y. Hussaini, Florida State Univ., myh@cse.fsu.edu

David P. Landau, Univ. of Georgia,
dlandau@hal.physast.uga.edu

B. Vincent McKoy, California Inst. of Technology,
mckoy@its.caltech.edu

Jill P. Mesirov, Whitehead/MIT Ctr. for Genome Research,
mesirov@genome.wi.mit.edu

Charles Peskin, Courant Inst. of Mathematical Sciences,
peskin@cims.nyu.edu

Constantine Polychronopoulos, Univ. of Illinois,
cdp@csrd.uiuc.edu

William H. Press, Los Alamos Nat'l Lab., wpress@lanl.gov

John Rice, Purdue Univ., jrr@cs.purdue.edu

Ahmed Sameh, Purdue Univ., sameh@cs.purdue.edu

Henrik Schmidt, MIT, henrik@keel.mit.edu

Leong says. “Blue whales are an example of another species with similar vocal challenges for researchers.”

Name that Sound

The team’s automated approach takes individual extracted vocalizations and applies a trained HMM to identify the vocalization in nearly real time. Given a sequence of distilled data reductions—in this case, coefficients derived from logarithmically rescaled fast Fourier transforms of the elephant vocalizations—the researchers built and trained a Markov chain to analyze new sound sequences.

This chain is a series of steps in which the probability for each next step is based only on the current state of that step’s portion of the data, with no dependence on what the previous data was or which chain steps processed it so far. Do enough “walks” to train the chain, and it settles into an equilibrium that you can use to identify new, unknown samples. Unlike rules-based systems, in which humans have to discover and then codify what makes each vocalization unique, an HMM simply requires the researchers to label different vocalizations according to type and then let the software decide how to uniquely differentiate and distinguish the types. Once Clemins’ team trained the chain, they ran new vocalizations through a Viterbi pattern-matching algorithm to rapidly classify them into previously identified speaker identifications and vocalization categories, producing a log of “which elephant said what.”

The HMM identification software runs on an ordinary (2-GHz) workstation using Cambridge University’s freely downloadable HTK 3.1.1 software and Clemins’ own Java additions and scripting. Among the additions was a *silence model*, which defined “not speaking” as a valid part of speech so that the HMM didn’t miscategorize initial pauses or lulls in the vocalization data.

It took the software just five minutes to train the speaker ID experiment 150 times with 150 known vocalizations. To provide the training data, the team gave each elephant its own microphone, which not only greatly reduces noise but also ensures that the training data for identifying individual elephants is accurate. Even so, numerous false positives for the shorter vocalizations and missed rumbles coincident with environmental noise occurred in the automated tests. The need for initial human processing might suggest further room for error. Fortunately, Leong says that in her experience, human error in classifying vocalization types—unlike that in identifying speakers—is nearly zero. Still, the researchers needed several human-tagged samples of the herd.

“The more training data, the better generalized the mod-

Dr. Dolittle, I Presume?

Patrick Clemins blames his decision to use elephants for his dissertation work on a vacation taken by his advisor, Mike Johnson. "Mike got interested in bioacoustics during an anniversary trip to Disney World. At [Disney's] Animal Kingdom, he noticed that they were recording elephants and labeling the data with behavior. His wife, Patricia, talked him into asking more about it and he eventually met [conservation biologist] Anne [Savage]," Clemins says. "Since Anne was our first collaborator, and Disney was recording elephants, elephants it was."

Johnson and Savage began the Dr. Dolittle Project (<http://speechlab.eece.mu.edu/dolittle/>) a bioacoustic effort to correlate behaviors and vocalizations not just for identification purposes, but also to assess habitat and species survival issues for African elephants, Beluga whales, Sumatran rhinos, tigers, and even chickens.

Johnson adds, "Our current research focuses on four primary areas: development of telemetry devices to track movement of species in the wild, reproductive biology of captive and wild animals, animal communication, and our field biology program, [which] encompasses both local and international conservation efforts."

els are and the more accurate the classification," Clemins says. "To a bioacoustics person, large amounts of data means hundreds of individual calls. To a speech person, this means hours of continuous speech consisting of hundreds of thousands of words. So, there is a definite gap here. The main disadvantage for us is [the] lack of high-quality, noise-free training data."

A big hurdle for real-time vocalization translation remains the initial extraction of each vocalization. The software expects discrete elephant vocalizations and can't handle a raw stream of sound. Thus, separating out the vocalizations from the raw sound tapes also requires a human operator—in this case, Leong. Disney's Wildlife Tracking Center has real-time spectrogram software, which scrolled the audible and infrasound spectrograms and let Leong identify, tag, and dump individual vocalizations into digital files. This step, which takes about twice as long as simply listening, is therefore the slowest part of the recording process. Despite these hurdles, Clemins is looking to the technology's future. "I don't think it would be a stretch to say that the algorithm could be implemented in real time on a Palm Pilot or Pocket PC, although we haven't tried this." Clemins says they've used the same algorithms as the speech-recognition programs available for such platforms.

Even real-time work would require a training period in which users label each discrete vocalization, but as the library



The Argonne National Laboratory Named Postdoctoral Fellowship Program

The Director's Office initiated these special postdoctoral fellowships at Argonne National Laboratory, to be awarded internationally on an annual basis to outstanding doctoral scientists and engineers who are at early points in promising careers. The fellowships are named after scientific and technical luminaries who have been associated with the Laboratory and its predecessors, and the University of Chicago, since the 1940's.

Candidates for these fellowships must display superb ability in scientific or engineering research, and must show definite promise of becoming outstanding leaders in the research they pursue; the Laboratory intends to award four such fellowships this coming year. Fellowships are awarded for a two-year term, with a possible renewal for a third year, and carry a stipend of \$71,000 per annum with an additional allocation of up to \$20,000 per annum for research support and travel.

Requirements for Applying for an Argonne Named Postdoctoral Fellowship:

The following documents must be sent **via e-mail to: fellowships@anl.gov by October 14, 2005.**

- Letter of Nomination (Recommendation from individual who supports your candidacy for the fellowship.)
- Curriculum Vitae (Include the names of the Nominator and two additional references.)
- Two letters of reference (It is the candidate's responsibility to arrange that the two reference letters be sent to the Laboratory **via e-mail** prior to the **October 14, 2005 deadline.**)
- Bibliography of publications
- Bibliography of preprints
- Description of research interests to be pursued at the Laboratory (We encourage applicants to contact Argonne staff in their areas of interest in order to explore possible areas of research.)
- Name of Argonne Division(s) in which you would like to work

All correspondence should be addressed to ANL Named Postdoctoral Fellowship Program. One application is sufficient to be considered for all named fellowships. For additional details, visit the Argonne web site at <http://www.anl.gov>. Argonne is an equal opportunity employer.

Argonne is operated by The University of Chicago for the U.S. Department of Energy Office of Science.

EDITORIAL OFFICE

COMPUTING in SCIENCE & ENGINEERING
 10662 Los Vaqueros Circle, PO Box 3014
 Los Alamitos, CA 90720
 phone +1 714 821 8380; fax +1 714 821 4010;
 www.computer.org/cise/

DEPARTMENT EDITORS

Book Reviews: Bruce Boghossian, Tufts Univ., bruce.boghossian@tufts.edu
Computing Prescriptions: Isabel Beichl, Nat'l Inst. of Standards and Tech., isabel.beichl@nist.gov, and Julian Noble, Univ. of Virginia, jvn@virginia.edu
Computer Simulations: Muhammad Sahimi, Univ. of Southern Calif., moe@usc.edu, and Dietrich Stauffer, Univ. of Köln, stauffer@thp.uni-koeln.de
Education: Denis Donnelly, Siena College, donnelly@siena.edu
Scientific Programming: Paul Dubois, paul@pfdubois.com, Konstantin Läufer, Loyola Univ., klauser@luc.edu, and George K. Thiruvathukal, Loyola Univ., gthiruv@luc.edu
Technology Reviews: Norman Chonacky, Columbia Univ., cise-editor@aip.org
Visualization Corner: Jim X. Chen, George Mason Univ., jchen@cs.gmu.edu, and R. Bowen Loftin, Old Dominion Univ., bloftin@odu.edu
Your Homework Assignment: Dianne P. O'Leary, Univ. of Maryland, oleary@cs.umd.edu

STAFF

Senior Editor: Jenny Ferrero, jferrero@computer.org
Group Managing Editor: Steve Woods
Staff Editors: Kathy Clark-Fisher and Rebecca L. Deuel
Contributing Editors: Cheryl Baltes, Keri Schreiner, and Joan Taylor
Production Editor: Monette Velasco
Magazine Assistant: Hazel Kosky, cise@computer.org
Technical Illustrations: Alex Torres

Publisher: Angela Burgess, aburgess@computer.org
Assistant Publisher: Dick Price
Advertising Coordinator: Marian Anderson
Marketing Manager: Georgann Carter
Business Development Manager: Sandra Brown

AIP STAFF

Jeff Bebee, Circulation Director, jbebee@aip.org
 Charles Day, Editorial Liaison, cday@aip.org

IEEE ANTENNAS AND PROPAGATION SOCIETY LIAISON

Don Wilton, Univ. of Houston, wilton@uh.edu

IEEE SIGNAL PROCESSING SOCIETY LIAISON

Elias S. Manolakos, Northeastern Univ., elias@neu.edu

CS MAGAZINE OPERATIONS COMMITTEE

Bill Schilit (chair), Jean Bacon, Pradip Bose, Doris L. Carver, Norman Chonacky, George Cybenko, John C. Dill, Frank E. Ferrante, Robert E. Filman, Forouzan Golshani, David Alan Grier, Rajesh Gupta, Warren Harrison, James Hendler, M. Satyanarayanan

CS PUBLICATIONS BOARD

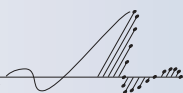
Michael R. Williams (chair), Michael R. Blaha, Roger U. Fujii, Sorel Reisman, Jon Rokne, Bill N. Schilit, Nigel Shadbolt, Linda Shafer, Steven L. Tanimoto, Anand Tripathi



IEEE Antennas &
 Propagation Society

IEEE

Signal Processing Society



of processed vocalizations grows, the software would encounter fewer unknown or uncertain matches. “Eventually, the software could be set to just run and pick the best choice and analyze a whole tape in near real time,” Clemens says. As users fix the mistakes, the software learns and improves.

Supporting the Science

Leong says that although all the rumble vocalizations they used in the experiment sound essentially the same to the human ear, their speaker ID work supports the claim of distinct voices. Many researchers believe elephants do have distinct voices, but there is thus far little strong scientific evidence to directly support this stance.

Animal Kingdom’s Joseph Solti believes the animals have unique voices, and provides physical evidence using principal components analysis (PCA) of vocalization formant frequencies. For example, the dominant frequency of an elephant’s voice is influenced by the shape of its vocal tract. Clemens notes that Solti’s use of a more traditional bioacoustics method has a lower accuracy, but other inherent advantages.

“The main drawback to our method is that it’s hard to visualize in a spectrogram what cepstral coefficients are and how the vocalizations are different,” Clemens says. “With Joseph’s paper, he was able to point out that the location of the formant frequencies varied the most between the speakers using PCA. This kind of result gives bioacoustic researchers something to visualize—cepstral coefficients do not.”

Given the real-time extraction and computer identification of elephant vocalizations, a “dictionary” of elephant sounds and the ability to play back samples and communicate with these animals might be a possibility—at least Clemens would like to think so. However, he points out, significant hurdles remain. “I’d be surprised if it doesn’t take years for us to unlock the various meanings of each vocalization. Our software only matches labels [elephant names and vocalization types] to vocalization data,” he says. “We need humans to try and understand what these labels might be.”

However, the computerized system does provide an unexpected benefit. “In the course of our research, we discovered two new types of vocalizations that had never been described before,” Leong says. “If we were relying on an automatic classification scheme, we would have missed those entirely.”

Alex “Sandy” Antunes is a freelance science writer based in Laurel, Maryland.