C E N T E R   F O R   R E S E A R C H   I N   L A N G U A G E

•   •   •

FEATURE ARTICLE

In Search of the Statistical Brain

Javier Movellan
Department of Cognitive Science, UCSD

ANNOUNCEMENTS

Postdoctoral Fellowship: The Center for Research in Language announces postdoctoral fellowships in Language, Communication, and Brain. The fellowships are supported by the National Institutes of Health and provide an annual stipend ranging from $19,608 to $32,300. Deadline for submission of applications is May 15, 1995. Contact Jan Corte at CRL for more information (jan@crl.ucsd.edu).

WWW: The CRL Newsletter is now on World Wide Web at URL: http://crl.ucsd.edu/newsletter.html

EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition (1-10 pages, sent via e-mail) and welcome response from friends and colleagues at UCSD as well as other institutions.

SUBSCRIPTION INFORMATION

If you are currently receiving a hardcopy of the newsletter and have access to e-mail, please help us save printing and mailing costs by forwarding your e-mail address to CRL. If you require a hardcopy in addition, please request it and we will be happy to send you one.

If you know of others who would be interested in receiving the newsletter, please forward the e-mail or postal mailing address. Please forward correspondence to:

Jay Moody, Editor
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (619) 534-2536 • E-mail: crl@crl.ucsd.edu

Back issues of this newsletter are available from CRL in hard copy as well as soft copy form. Papers featured in previous issues include the following:

In Search of the Statistical Brain

Javier Movellan
Cognitive Science, UCSD

The brain is a fantastic random number generator and its operating principles are inherently statistical. I refer to this conceptual framework as "The Statistical Brain". This approach to understanding the brain and human information processing is not new. In my case, it was inspired by the work of signal-detection theorists in psychology, the PDP work on harmony theory, and the ideas of the late von Neumann, the designer of the digital computer.

According to von Neummann, a crucial difference between the brain and the digital computer is that the brain is statistical in nature whereas digital computers are deterministic. Digital computers were designed with an eye on reliability: At a hardware level, natural noise is cancelled by the use of digital technology, and at a functional level, computation is grounded on deterministic Boolean logic. Contrary to the digital computer, brains are designed to operate in natural environments, where uncertainty rules. They are made of massive numbers of simple stochastic processors and their representations are probabilistic and flexible, as required by the enoromous variability present in real-life situations.

But what are the designing principles of these natural stochastic computers? Can we envision new computers based on the same statistical principles as the brain? I am trying to address these questions in two ways: 1) analyzing the formal properties of stochastic networks, 2) studying how humans combine different sources of uncertain information in perceptual tasks.

With respect to the study of stochastic networks, the goal here is to find a formal framework to better understand how stochastic dynamical systems work. At this level, analysis is grounded on continuous stochastic calculus, a generalization of ordinary calculus. One of the things we learn from these network models is the need to think in terms of probability distributions evolving through time. When we initialize these networks, probability concentrates in particular states and, as time progresses, probability, and thus information, evolves according to well defined diffusion principles. From this point of view, one can think of the statistical brain as a continuous representational web and of probability as a substance diffusing through this web in response to internal and external forces. Understanding network dynamics in probability space may give us important insights about the way the statistical brain operates.

I am also trying to understand the designing principles of the statistical brain by studying how humans combine different sources of information (stimulus, context and prior knowledge) in perceptual tasks. The task I am working on now is audio-visual (AV) speech perception.

We know that the brain uses both visual and acoustic information to recognize speech. For example, when acoustic information for the syllable /ba/ is synchronized with video images of lip movements for /ga/, subjects report hearing /da/ or /ta/. This phenomenon, known as the McGurk-McDonald effect, raises crucial questions about the way the brain combines information from different sources: What kind of representations facilitate this intermodal integration? Is bimodal speech perception based on relatively unmodified and independent elements from each modality? Is it based on non-independent amodal representations? What are the temporal dynamics of information integration?

I like to approach these problems first from an engineering point of view. If I had to develop an optimal system to recognize speech, how would I go about it? Here is where I find probability theory in general and pattern recognition theory in particular so useful. In pattern recognition ideal optimal systems are called "Bayesian classifiers", or maximum posterior classifiers (MAP). A system that follows the MAP principles is guaranteed to achieve minimum error rates. MAP is a very useful framework to understand the type of problems that the brain needs to solve in real life. MAP tells us in a general way how information needs to be combined to achieve optimal performance. However, MAP itself does not tell us about specifics, unless we are willing to make assumptions. Here is where modeling results from human experiments may help: Is the human data consistent with the assumtions we are making?

Experimental psychologists have studied the problem of AV speech recognition with some success but their approach has been too limiting. In my opinion current psychological models have two major shortcomings:

1) Lack of temporal dynamics: Current models of speech perception typically do not pay attention to the temporal dynamics of the visual and acoustic signals. Figure 1, for example, shows preliminary results from an ongoing experiment in my laboratory. The figure shows the percentage of fused AV responses in a McGurck-like experiment as a function of the temporal delay between the visual and acoustic signals. The curve marked with circles is for a high-volume acoustic signal, and the curve marked with squares for a low-volume signal. As the figure

File Name : jett.ps
Title : Fontographer 3.5
Creator : DeltaGraph 2.0
CreationDate : 8/12/92 2:49:04 PM

FIGURE 1

indicates, synchrony between the visual and acoustic signals plays a well defined role in the percentage of combined responses. This type of temporal dynamics is ignored in present psychological models but needs to be addressed if we want to develop realistic models of information integration.

2) Insufficient specification of computational mechanisms: Current psychological models are typically built from the top-down. Based on the response confusions made by humans, simple representational models are developed that generate the same type of confusions that humans do. This top-down approach typically used in psychology (from responses to representations) is insufficient; it needs to be complemented with a bottom-up approach (from physical stimuli to internal representations). The bottom-up approach to modeling emphasizes the importance of models capable of processing physical signals through time. For example, in the AV speech recognition case, we may start with images and acoustic signals, process them with biologically inspired models of the acoustic and visual system and train models of AV speech integration that would actually work in real-life situations. Once we have a model built from the bottom-up we can test whether the responses generated by the system match the data obtained from humans. This strategy has the advantage of forcing us to be very specific about hidden

assumptions in our models. Moreover, it allows us to visualize the kind of representations that may be sufficient to solve the task under study.

Figure 2 shows an example of this bottom-up approach. The figure shows typical representations learned by a purely visual synthetic speech recognizer developed in my laboratory. The system is based on a simple stochastic network trained to recognize the first four digits in English. Each column is a different digit, starting with "one." Each row represents different time steps. The two pictures within each cell are related to intensity and to intensity derivatives, a crude measure of flow. The network uses dynamic probability distributions to represent possible ways in which people say the digits in English. Since we cannot visualize entire probability distributions evolving through time, the figure just shows the most-likely paths. The fact that the network representations are entire probability distributions, not just fixed patterns, allows it to be robust to variations in the way people look and act when they say things.

This particular system achieved a 89.5% correct generalization, which compares well with the 89.9% correct obtained by untrained humans. However, trained lip-readers achieved a 95% correct rate,

start

finish

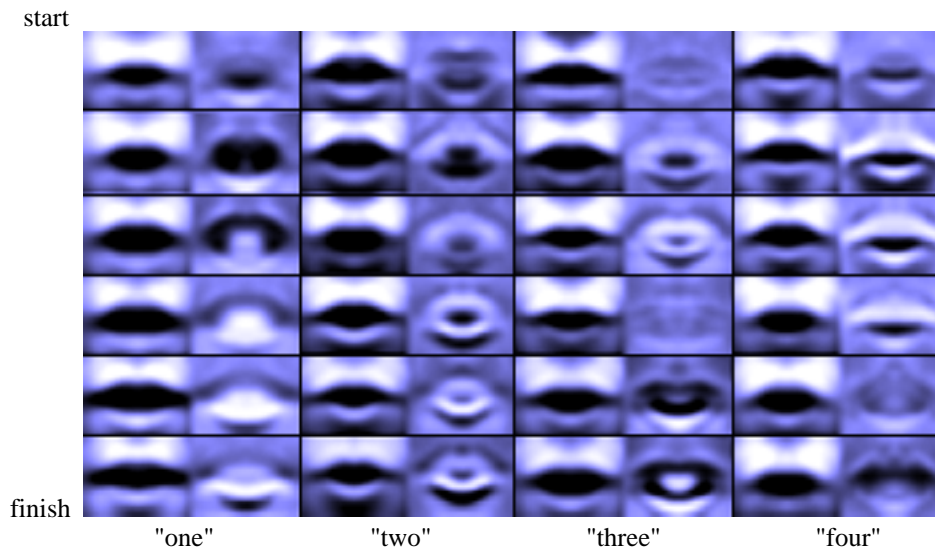"one"   "two"   "three"   "four"

FIGURE 2

indicating that there is still room for improvement. Interestingly the type of mistakes made by humans and by the synthetic system had a 0.99 correlation (98% of the variance in human confusions can be accounted for by the artificial model). This suggests that the probability distribution of representational states learned by the artificial system is a reasonable model of the stochastic representational space used by humans.

Presently we are developing a combined audio-visual system. The acoustic signal will be handled by a biologically inspired model of the auditory system, that converts the incoming waveform into a statistical representation of the pattern of activity in the cochlea. The visual input will be handled by a model of the MST, a center in the brain related to optical flow computation. This model, which was developed by Sereno and Zhang in our Department, computes optic flow in a robust and inexpensive way. Learning and information integration will be handled by a stochastic neural network. One of the most exciting aspects of this project is that it will help us find optimal ways to combine visual and acoustic representations. Is it a good idea to do low-level integration of the representations and base perceptual decisions on these multimodal representations? Is it better to keep the two channels separate and base the perceptual decisions on independent modal representations? This project will provide answers to these questions.

This is just an example of the possibilities opened by integrating the study of the brain, human information processing and computational analysis. In my case, probability theory and statistics are invaluable tools to guide my research and to bridge the gaps between

these three fields. Hopefully our quest to understand the designing principles of the stochastic brain will take us to new, unexplored territories.

POST SCRIPT:

If you are interested on the specifics of the AV speech recognition project at my lab, you may contact me at `movellan@cogsci.ucsd.edu`. Or you can get more information and copies of related papers from my World Wide Web pages at: `http://cogsci.ucsd.edu/~movellan/`.

Note especially these links:

    AV Speech Recognition
    Models of Optic Flow
    Speech Recognition Information
    Pattern Recognition Information