

Biological Specializations for Speech: What Can the Animals Tell Us?

J.D. Trout¹

Philosophy Department and the Parmly Hearing Institute, Loyola University Chicago, Illinois

Abstract

Language research over the past 40 years converges on the speech-is-special hypothesis (SiS), according to which speech perception and production are uniquely human adaptations. SiS is grounded in a variety of biological, developmental, and behavioral evidence. In some comparative studies, speechlike stimuli have seemed to cause nonhuman animals to exhibit humanlike performance functions. Auditorists—who believe that spoken language processing is executed by, and is explicable in terms of, general auditory mechanisms—have seized upon such studies as evidence that SiS is incorrect. However, it is difficult to identify biological and functional similarities across different species on the basis of behavior alone, and the elaborate training regimen that nonhuman animals require to achieve human performance levels undermines the significance of certain comparative studies. Both comparative and human behavioral research, including brain-imaging studies of functional localization, electrophysiological recordings of the neural basis of the perception-production link, and developmental studies of a time-locked schedule of language learning, favor SiS over auditorism.

Keywords

speech-is-special (SiS); auditorism; comparative research; speech perception and production

In humans, speech perception unleashes an intricate choreography of complex neural feats. Over the course of a sentence, the listener must spontaneously analyze the frequency spectrum, identify phonetic features, segment phonological units from the continuous stream of speech, apply rules of word formation, track intonational contours, access words from the mental lexicon, and enforce syntactic constraints on permissible grammatical formulation. The human brain executes this multilayered process with blinding speed. We perceive, and we produce, about three words per second, or one phone (the smallest perceptible unit of speech) every tenth of a second.

Two opposing theoretical explanations orient research programs on these marvels of spoken language processing. According to the view that speech is special (SiS), the mechanisms and processes of speech perception and production are uniquely human adaptations. Auditorism, on the other hand, holds that spoken language processing is executed by, and is explicable in terms of, general auditory mechanisms shared by many other species. Comparative research on animals (Hauser, 1996) has been

the main source of evidence in the SiS-versus-auditorism debate (Trout, 2001).

DIVERSE EVIDENCE FOR SiS

Many language scientists, including Eric Lenneberg, Noam Chomsky, and Steven Pinker, have promoted the view that there is biological specialization for language, and in particular for syntax. But there are several forms that human language takes: written, signed, and spoken. SiS draws its support mainly from the last of these—the capacity to perceive and produce spoken language. What, then, is the evidence for SiS?

Initially, casual observation would seem to conflict with claims of uniqueness. For example, many nonhuman animals have two eyes, two ears, one mouth, and one nose with two nasal passages. We share with fish a covering of our body that is vibrotactile (responsive to vibration via sensory contact), and our middle ear is traceable to the stapes and braincase in fish. There are similarities more specifically related to spoken language as well. Apes and birds communicate with sound, and some birds, such as Irene Pepperberg's remarkable African Grey parrot, Alex, can produce strings of words. Dogs, horses, pigs, apes, dolphins, and many birds can respond to language commands. Given these many shared traits and abilities, why would anyone think that the mechanisms used in speech perception and production are adaptations unique to humans?

In a certain respect, attributions of specialization are easy to trivialize: At some level of detail, any organism's psychological processes will appear unique. Is there an instructive notion of species specialization? Biological specialization is

evident in diverse brain-imaging and electrophysiological measures showing that words but not tones activate brain regions that do not exist in other species. Furthermore, although each ear is connected to both sides of the human brain, speech presented to the right ear is more quickly and accurately processed than speech presented to the left ear, indicating the enhanced efficiency typical of specialization. This *right-ear advantage* is achieved by a slightly stronger connection from the right ear to the left side than to the right side of the brain, and supports the idea of a brain that is specialized, because broadly lateralized, for language function.

Although not couched in biological terms, developmental evidence also supports species specificity, because developmental changes in production and perception of speech in humans follow a species-characteristic, time-locked pace, sequence, and scale. There is, for example, a critical window for spoken language acquisition in children that closes at about age 7, but no such critical period exists for audition.

A host of nondevelopmental findings about specialization bolsters the case for SiS as well, by signaling the independence of speech from auditory function. Humans can exhibit speech and language disorders without auditory disorders. Moreover, there is a characteristic symmetry between speech perception and speech production. This symmetry is easily explained by SiS: Neural representations encoding linguistic (rather than merely auditory) information provide the common currency to both guide our spoken language perception and regulate the motor functions of speech production. This symmetry is more difficult to explain if our speech perception and production are fundamentally regulated by auditory information, as

auditorism holds, for certain perceptual judgments of speech are prompted by the linguistic information available in the visible movement of the lips and jaw. This is the simplest illustration of the additional, nonauditory bases of symmetry accommodated by SiS but not auditorism. But there are many others. For example, brain-imaging measures show that when people make perceptual judgments about speech, areas of the brain that are involved in speech production are activated. Also, when native speakers of Japanese are trained to perceive the contrast between /r/ and /l/, their pronunciation of this contrast is improved.

Finally, as noted, the visible movement of the articulators influences how speech is perceived. This tendency is illustrated by the McGurk effect, in which one dubs the sound of the syllable [ba] onto the visible mouth movements for [ga]. Normal adults report hearing [da], a sound articulated between the other two. Not all McGurk-type impressions represent blends that express the compromise that is most likely given the language experience of the listener. An audio [ga] and a video [ba] produce the impression of [bga], an impossible speech object in English. These irregular audio-visual impressions are difficult to explain using an auditorist approach because they violate linguistically predictable relations between an auditory impression and its visual correlate. Auditorism has little to say about the influence of visual information on the perception of speech.

DEBUNKING SiS

Auditorists have focused their critical efforts on behavioral evidence, which represents a slender sample of all of the evidence for SiS. Even nonauditorists who value

comparative research criticize SiS for the “persistent tendency to assume uniqueness even in the absence of relevant animal data” (Hauser, Chomsky, & Fitch, 2002, p. 1574). However, this is no stubborn assumption of SiS. A substantial portion of the evidence for SiS requires no experimental comparison; it falls out of textbook discussions of the functional anatomy of each species’ brain and the schedules of different species’ development. For example, the lateralization effects found in human language (such as the right-ear advantage) are not present in quail because the quail brain is not lateralized; it does not have separate hemispheres, each devoted to distinct functions. Birds and chinchillas do not have language disorders—reading or otherwise—and so of course there is no reason to hold out for a relation between a bird audiogram (a graph displaying hearing sensitivity at different frequencies) and bird aphasia (a language disorder that affects the ability to produce or comprehend grammatical and semantic structure). Unlike humans, birds and chinchillas have no spontaneous developmental schedule for speech. Finally, they have neither an ability to produce language nor homologues to Broca’s area and Wernicke’s area (brain areas involved in producing and perceiving speech), and so there is no possibility of reproducing brain-imaging investigations to compare results with those obtained in studies of human perception of words versus tones. In short, these species are so different from humans biologically that comparative efforts on this wide range of speech abilities have little rationale.

If comparisons of brain activation and developmental progression are typically inappropriate, what remains to compare across species? Surely the most revealing comparative research will be on apes and monkeys, because they

are closest to humans on the evolutionary tree. Historically, however, such systematic comparisons have not been made.

In the classic behavioral experiments used to undermine SiS, chinchillas (Kuhl & Miller, 1975) and quail (Kluender, 1991) were trained on a categorical perception task. Chinchillas would display a barrier-crossing behavior, and quail would peck, to identify /da/ or /ta/. After a period of training of no less than 4,000 trials, the quail and the chinchilla were prompted to generalize their responses to digitized "graded" versions of /da/ and /ta/ (synthetic versions that were "between" /da/ and /ta/ along a selected acoustic dimension). To everyone's surprise, chinchillas and quail, like humans, starkly categorized these graded versions as either /da/ or /ta/ and yielded a performance function² arguably similar to that of humans for a similar task. Thus, although categorical perception was once thought to be a hallmark of evidence for speech specialization in humans, an auditory basis for categorical perception is present in animals whose lineage long predates the evolution of humans.

REFUTING AUDITORISM: FLEAS, ELEPHANTS, AND THE CIRCUS PROBLEM

Auditorists argue that similarity of response, by itself, implies a similar functional or biological mechanism. Parsimony, it is said, demands that we infer sameness of mechanism from similar behavior. However, there is an obvious problem with this inferential pattern: Mere behavioral similarities come cheap. For example, under a strict regimen of circus training, horses can count, and fleas can pull carriages. Similarly, elephants can be trained to walk bipedally. How-

ever, we infer neither that the elephant has the same mechanisms as humans for bipedal gait nor that humans are not equipped with special mechanisms for bipedal gait.

Most comparative research records behaviors of short duration (e.g., pressing or pecking at a bar, pressing a computer key). If instead we take a developmental approach and sample behavior over several periods, two striking differences emerge between nonhuman and human patterns of learning. The first concern the scale of the information acquired, and the second concerns its mode of its acquisition. For example, all of the quail required at least 4,000 trials to identify a consonant (or an instance of its multiple realizations) 90% of the time (Kluender, 1991). Several of the quail required more than twice that number of training trials. Chinchillas required 10 to 15 min of training every day for 7 months, or more than 5,000 trials (Kuhl & Miller, 1975) to reach the same criterion. Given these facts, what reason do we have for thinking that the same mechanisms are being used by quail and chinchillas, on the one hand, and human children, on the other? Quail and chinchillas require thousands of trials in an intensive training regime to perform at a 90% identification rate on a small handful of consonants or vowels. In contrast, human children spontaneously learn rules and an arbitrarily large store of words from casual, unregulated exposure and (by comparison) uncontrolled feedback. Moreover, their accuracy is generally better than 90%.

Hauser et al. (2002) acknowledged that a difference in scale between rates of spontaneous learning versus elaborate training threatens the validity of cross-species comparisons: "The rate at which children build the lexicon is so massively different from nonhu-

man primates that one must entertain the possibility of an independent mechanism" (p. 1576). Despite this concession, they also said that the animal studies support the "continuity thesis," the "null hypothesis of no truly novel traits in the speech domain" (p. 1574). However, these two assertions apparently conflict, because all of the nonhuman animal studies using human language stimuli involved elaborate, validity-threatening training. If one applies the criteria that Hauser et al. defended, the comparative studies provide counterevidence for the claim of a shared, cross-species, sensorimotor system that can be recruited for spoken language perception.

FUTURE DIRECTIONS AND CONCLUDING REMARKS

One important area for future research will examine the way perception and production (of language and other behaviors) are linked in the brain. Until very recently, the perception-production link in humans was explained by mechanisms that remained largely theoretical. However, the existence of mirror neurons, discovered in primates in the past several years, holds particular promise for elucidating the special connection between speech perception and production (Arbib, 2003; Fowler, Galantucci, & Saltzman, 2003). Researchers have found neurons in the premotor cortex³ of monkeys that react when the monkey performs an act as well as when it perceives a similar action performed by another monkey (or even by a human). These neurons have been called "mirror" neurons because they seem to be involved in a neural mechanism that allows monkeys, and, it turns out, humans, to copy, imitate, or otherwise produce actions by perceiving their produc-

tion. The existence of mirror neurons promises to explain a range of brain-imaging findings establishing that, in intelligent pursuits from speech to social evaluation, passive perceptual judgment alone is sufficient to activate motor representations used for production. Mirror neurons may supply the common code that underwrites the perception-production link, as the pioneering researchers in this field have indicated: "In primates, the aim of the motor system is to create internal copies of actions and to use these internal copies for generating actions as well as for understanding motor events" (Rizzolatti, Fogassi, & Gallese, 2000, p. 550).

Once more fully understood, the function of mirror neurons could give systematic voice to the view that "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands" (Lieberman & Mattingly, 1985, p. 2). A suite of specialized perceptual mechanisms "processes the acoustic signal so as to recover the coarticulated gestures that produced it. These gestures are the primitives that the mechanisms of speech production translate into actual articulator movements, and they are also the primitives that the specialized mechanisms of speech perception recover from the signal" (Lieberman & Mattingly, 1989, p. 491).

Another challenging area of research concerns the evolutionary basis of speech. How did the capacity for language evolve? It is difficult to find principled answers to this question. One recent and influential proposal (Hauser et al., 2002) suggests that the cognitive capacities required for language processing derived from independent cognitive abilities to count, navigate, and coordinate socially. Admittedly, the evidence here is scant. Like Kipling's *Just So Stories*, fanciful evolutionary stories can be

told about any trait or ability. How, precisely, does an ability to count, to navigate, or to scheme socially contribute to spoken language perception and production, including sentence processing, planning, and execution? After all, counting, navigating, and scheming are deliberate and effortful cognitive activities, whereas syntactic functions such as planning and parsing are implicit processes and so relatively effortless. Even if humans recognize addition by the age of 3 or so, but chimpanzees do not, how would that human ability provide a basis for parsing? The problems that face this evolutionary proposal are both exciting and speculative.

A scientific research program that recruits biological evidence must distinguish between traits that are homologies (traits based on common descent—e.g., eyes in vertebrates) and those that are homoplasies (or analogies—similar structures that evolved independently of each other, and so are based on different mechanisms). This is an especially formidable task, because some traits, such as teeth in jawed fish and the capacity for live birth in reptiles, have been lost and regained at least once in a species' history. The mechanisms underlying similar cross-species behavior are different if the biological basis was once lost in one species and not in the other. For the purpose of answering the homology-or-analogy question, behavioral similarities are not especially probative. Only biological research will establish homologies, and only functional and biological work will characterize species specializations.

There has been an explosion of comparative research on the mentality of nonhuman animals in the past two decades, and this line of comparative research on audition has been especially distinguished and fruitful for hearing science. However, it is unlikely that success in hearing science will render trac-

table many of the durable theoretical problems of speech science. There is a hopeful (and so, vulnerable) population with language disorders that is desperate for solutions. Thus, it can be tempting to embrace alternative research programs whose hypotheses appear more clinically motivated than those of SiS. SiS promises a multi-layered model integrating higher-level processes of linguistic representation and lower-level neural mechanisms, but many people, understandably, are hoping for successful remediation in the short term and are pinning their hopes on the tractability of speech disorders via simple auditory processing. SiS is chiefly concerned with durable theoretical questions about the nature of spoken language processing, and, unlike auditorism, SiS is not prominently associated with clinical applications in the short run. Nor is SiS focused on generating consulting fees, patents for dramatic technological advances, or opportunities for third-party payment in clinical settings. Indeed, the enthusiasm for clinical applications can pose a special threat to the integrity of a scientific research program when recognized standards of good science conflict with accepted standards of clinical practice. This threat is amplified when practitioners at clinical speech centers belong to the very professional organization that regulates their clinical standards. In such cases, the standard for change, and so for improvement, tends to be conservative; it may be difficult for people to impose standards that would, if implemented, threaten their livelihood. Consequently, some accreditation standards and clinical remedies may not receive the critical scrutiny characteristic of good science (Dawes, 1994).

We have come a long way since Descartes's simple a priori pronouncement that dogs, horses, and monkeys could not express thought

(for more recent developments, see Allen, 2003). SiS is an overarching scientific hypothesis that unifies diverse findings about spoken language perception and production. The evidence for it is derived both from comparative research that taps the natural abilities of nonhuman animals and from behavioral, developmental, and biological studies of humans. Much basic research remains to be done. Knowledge gained from this research will yield both a model of language function and an account of spoken language as part of the human genetic endowment.

Recommended Reading

Aitchison, J. (2000). *The seeds of speech: Language origin and evolution*. New York: Cambridge University Press.

Liberman, A.M. (1996). Introduction: Some assumptions about speech and how they changed. In A.M. Liberman (Ed.), *Speech: A special code* (pp. 1–44). Cambridge, MA: MIT Press.

Pinker, S. (1994). *The language instinct*. New York: HarperCollins.

Trout, J.D. (2001). (See References)

Notes

1. Address correspondence to J.D. Trout, Philosophy Department and the Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Rd., Chicago, IL 60626; e-mail: jtrout@luc.edu.

2. A performance function is a graph that depicts how behavior on a task—such as percentage of responses that are correct—changes with carefully controlled changes along a dimension of the stimulus.

3. The premotor cortex is an area near the front of the brain that appears to store much gestural information about motor activities such as moving the hands or the mouth.

References

Allen, C. (2003, Summer). Animal consciousness. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved July 23, 2003, from <http://plato.stanford.edu/archives/sum2003/entries/consciousness-animal>

Arbib, M. (2003). The mirror system, imitation, and the evolution of language. In C. Nehaniv & K. Dautenhahn (Eds.), *Imitation in animals and artifacts* (pp. 229–280). Cambridge, MA: MIT Press.

Dawes, R. (1994). *House of cards*. New York: Free Press.

Fowler, C.A., Galantucci, B., & Saltzman, E. (2003). Motor theories of perception. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 705–707). Cambridge MA: MIT Press.

Hauser, M. (1996). *The evolution of communication*. Cambridge, MA: MIT Press/Bradford Books.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Kluender, K. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America*, 90, 83–96.

Kuhl, P., & Miller, J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69–72.

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.

Liberman, A.M., & Mattingly, I.G. (1989). A specialization for speech perception. *Science*, 243, 489–494.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 539–552). Cambridge, MA: MIT Press.

Trout, J.D. (2001). The biological basis of speech: What to infer from talking to the animals. *Psychological Review*, 108, 523–549.