

Owren, M. J., & Goldstein, M. H. (2008). Scaffolds for babbling: Innateness and learning in the emergence of contextually flexible vocal production in human infants. In D. K. Oller and U. Griebel (Eds.), *The Evolution of Communicative Creativity: From Fixed Signals to Contextual Flexibility*. Cambridge, MA: MIT Press.

SCAFFOLDS FOR BABBLING: INNATENESS AND LEARNING IN THE EMERGENCE  
OF CONTEXTUALLY FLEXIBLE VOCAL PRODUCTION IN HUMAN INFANTS

Michael J. Owren

Georgia State University

Michael H. Goldstein

Cornell University

INTRODUCTION

As one of the most distinctive features of human behavior, spoken language is widely believed to be critical to our species' notable flexibility in cognition, social interaction, and culture. Understanding the origins of this complex and apparently unique faculty poses significant challenges, and has engendered requisite controversy and discussion. Examples of critical issues include whether speech began relatively early or late in human evolution, whether any species other than humans exhibit or can acquire similar capabilities, and whether the evolutionary process involved was gradual and continuous or instead was so abrupt and discrete that it should be considered

an evolutionary discontinuity relative to our closest nonhuman primate relatives (hereafter *primates*).

The phenomenon of *babbling* in human infants is often considered particularly illuminating for these kinds of questions. By babbling, we mean prelinguistic production of meaningless, non-speech sounds that are nonetheless recognizable precursors of later phonemes. In this chapter, these *prelinguistic* vocalizations will be distinguished from *nonlinguistic* sounds such as laughter and crying, which are not considered direct precursors of phonemes or words. We will also distinguish between *early* and *canonical* forms of babbling, with the former referring to sounds with vowel- or consonant-like components that are not yet fully formed, and the latter applied to the speech-like and often reduplicated syllabic combinations of consonants and vowels that infants begin to produce prior to the emergence of words. While the terms babbling and canonical babbling are often used interchangeably, Oller (2000) argues that early babbling is characterized by *quasi-resonant vowels* and *primitive articulation*, the latter composed of slow, uncertain consonant-vowel transitions referred to as *going* or *marginal syllables*. In contrast, canonical babbling is marked by fully resonant vowels that are combined with rapid consonant-vowel alternation to form *canonical syllables*.

One common point of view is that babbling is a genetically based, evolutionary adaptation that arose in humans to facilitate production and perception of language-specific phonemes (e.g., Fry, 1966; Pulvermüller, 1999; Pulvermüller & Preissl, 1991; Pettito, 2004). This approach tends to emphasize the uniqueness of this behavior, noting that while babbling-like vocalizations occur in many songbirds (e.g., Doupe & Kuhl, 1999; Goldstein et al., 2003), it is rare or absent in nonhuman mammals (cf. Elowson et al., 1998; Knörnschild et al., 2006).

This chapter presents a different perspective, one that views the ontogeny of speech as an activity- and experience-dependent process from the very earliest stages. Specifically, we argue for the importance of learning both in the emergence of prelinguistic vocalization, and in the infant's subsequent progression to early babbling. In other words, while many researchers take the onset of canonical babbling as a starting point, we suggest that even nonlinguistic vocalizations play an important role in the vocal development that culminates in the emergence of speech. According to this *babbling-scaffold* view, each stage of vocal ontogeny acts as a bootstrap or scaffold for the next, beginning virtually from birth. Thus, nonlinguistic vocalization facilitates the emergence of early babbling, which in turn is the foundation of canonical babbling. Two particular factors we will point out in the progression of events preceding

canonical babbling are the infant's acquisition of volitional vocal control and its sensitivity to the instrumental value of vocalizing.

The hypothesis relies on a variety of kinds of evidence, beginning with the observation that the infant's nonlinguistic vocalizations are quite like primate calls. For example, they emerge in the absence of auditory experience, are largely unarticulated sounds, and are most likely mediated by subcortical, emotion-related brain areas. In both humans and primates, these kinds of sounds will be termed *innate*, in the limited sense that producing them requires little if any auditory experience or motor practice. However, vocal development in humans quickly diverges from this primate-like state, with infants soon showing levels of volitional vocal control over vocal production that are far beyond the capabilities of primates. Here, the evidence indicates that the infant first begins to control some aspects of their nonlinguistic vocalizations, with vowel-like sounds in particular providing the bridge from nonlinguistic to prelinguistic communication.

We argue that there are at least two different kinds of learning involved. The first derives from emotion-triggered production of nonlinguistic vocalizations, which exposes the infant's cortex to converging streams of proprioceptive and

auditory feedback. The proposal is that limbically mediated sounds act as a developmental scaffold that facilitates acquisition of volitional vocal control by providing sensorimotor experience to the infant's cortex. This process is interwoven with learning that nonlinguistic and early babbling vocalizations have important instrumental value, for instance in eliciting caregiving responses from others and promoting other kinds of social interactions as well. Overall, the argument is that even the earliest vocal development involves deeply intertwined learning about vocal production and the social roles of vocalization, and that both are critical to the infant's progress from nonlinguistic to prelinguistic communication.

To flesh out this argument, we begin by reviewing data from primate vocal behavior, emphasizing an evident dissociation between production and reception of vocalizations in these animals. The key point here is that primates show little flexibility in producing vocalizations, and are evidently not able to exert direct volitional control over calls acoustics or the circumstances of calling. In contrast, the same animals show great flexibility in responding to sounds. While humans, on the other hand, are very flexible in both, a difference we argue can be traced to more extensive connections between the human cortex and brainstem neurons that innervate peripheral vocal anatomy. However, studies of normally hearing and hearing-impaired

infants also show that fine-tuned volitional control over vocal production is an acquired ability in humans, which we propose develops gradually through stages of *laryngeal* and then *supralaryngeal* control. While neither laryngeal nor supralaryngeal production is likely under volitional control in primates, the hypothesis nonetheless includes that the heritage of nonlinguistic vocalization and flexible auditory learning capabilities that humans share with these animals are critical contributors to the developmental events that eventually produce our unique vocal communication abilities.

#### PRIMATE VOCAL BEHAVIOR

##### *Vervet monkey alarm calls*

Recent years have seen much interest in the question of whether primates and other nonhumans use vocalizations in ways that parallel word use in human language. In spite of these efforts, Seyfarth et al.'s (1980) pioneering demonstration of semantic-like properties in vervet monkey alarm calls arguably remains the most convincing example of language-like function in primate vocalizations. The work was particularly convincing because vervets produce a number of acoustically distinct alarms that the researchers could show are linked to specific, mutually exclusive predators and escape strategies. Vervet responses upon hearing *snake*, *eagle*, and *leopard* calls played from hidden

speakers provided requisitely compelling evidence that these sounds have specific representational value.

However, the development of vervet alarm calls is also strikingly different from the emergence of language in humans. Most importantly, Seyfarth et al. found that vervets produce recognizable alarm calls from a very young age, and in largely appropriate situations. In other words, vervet infants evidently do not need to learn how to produce alarm vocalizations with appropriate acoustic features, nor in which general circumstances to use the various calls. However, learning was found to be critical in responding to the calls. While able to call "correctly" from an early age, young vervets do not initially show differentiated escape reactions to alarm vocalizations from others. Instead, the youngsters tend to freeze, run to their mothers, or react in ways that can increase rather than decrease their exposure to danger. Infant vervets do acquire predator-appropriate responses relatively quickly, but in this case, experience with the calls and responses of other group members plays a critical role (Cheney & Seyfarth, 1990a).

#### *Vocal production*

*Little learning of vocal production to control acoustics.*  
The deep divide between producing and responding to calls shown by vervets is mirrored by evident innateness in the vocal production in other primates. For example, data from several

monkey species show that vocalizers produce recognizable species-appropriate calls from an early age, and that developmental effects on vocal acoustics are likely mainly maturational in nature. Evidence on this point is ably reviewed by Hammerschmidt and Fischer (this volume) and will not be repeated here except to note that the studies include some with monkeys that were deaf (Talmadge-Riggs et al., 1972), socially isolated (Winter, 1972; Hammerschmidt et al., 2000), or reared in an altered auditory environment (Owren et al., 1992a, 1992b, 1993).

In addition, the same conclusion can be drawn from recent reports of babbling-like vocalizations in two nonhuman mammals, namely pygmy marmosets (*Cebuella pygmaea*; Elowson et al., 1998) and sac-winged bats (*Saccopteryx bilineata*; Knörnschild et al., 2006). Early vocal production in these species resembles human canonical babbling in that infants produce a mix of adult-like sounds that are divorced from their normal calling contexts. In contrast to human babbling, however, the sounds have no syllabic character, and there appears to be little or no acoustic learning involved. Instead, the "babbled" vocalizations are acoustically recognizable as adult-like calls from the beginning, do not arise from more approximate, intermediate forms, and do not seem to require practice.



*Lack of articulation.* Primate vocalizations are also different from human speech in that key acoustic features typically reflect *laryngeal* rather than *supralaryngeal* aspects of vocal production. By *laryngeal*, we mean acoustic characteristics due to vocal-fold vibration patterns, whereas *supralaryngeal* refers to filtering effects traceable to cavities and tissues above the larynx. Linguistic contrasts in human speech typically rely on flexible positioning of the tongue, mandible, and lips, referred to as the *articulators*. In primates, however, major sound-classes within a given repertoire primarily reflect differentiated vocal-fold action (Brown et al., 2003; Owren, 2003). Compared to humans, primates have thinner tongues, larynges positioned higher in the neck, and a relative lack of flexible soft tissues in the *supralaryngeal* vocal tract. While *supralaryngeal* filtering effects are likely important at least in some sounds (Owren et al., 1997), primates rarely seem to use active articulation to create functionally significant sound contrasts (Lieberman, 1975; cf. Hauser et al., 1993; Riede & Zuberbühler, 2003).

*Limbic neural mechanisms.* Evidence about the neural circuitry underlying vocal production in primates is also extensively reviewed by Hammerschmidt and Fischer (this volume). A major point is that subcortical structures play a central role, for instance with electrical stimulation of midbrain

(periaqueductal gray, or PAG) eliciting fully formed, natural-sounding vocalizations in several species, including squirrel monkeys, rhesus monkeys, gibbons, and chimpanzees (Jürgens 2002). Natural-sounding vocalizations can also be produced by stimulation in the hindbrain, as well as in limbic-system structures such as hypothalamus and cingulate cortex. However, stimulating these latter regions is less reliable, slower, and elicits less natural-sounding calls. Stimulating cerebral cortex does not elicit vocalizations, nor does cortical lesioning interfere with vocal production. In addition to limbic pathways, motor neurons innervating the lips, mandible, tongue, and vocal folds have also been found to have connections to the facial area of motor cortex. But these *corticobulbar* pathways have not been shown to have a role in vocal production.

*Vocal conditioning effects.* Some evidence of plasticity in vocal production has been provided by operant conditioning studies (Pierce, 1985), but may also mainly underscore the primacy of limbic, rather than cortical mechanisms. For example, several laboratory studies have shown that rhesus monkeys working for food-reward can show contingent changes in the rate, duration, and intensity of species-typical *coo* calls (Sutton et al., 1973). Masataka (1992) has also reported anecdotal evidence that provisioned, free-ranging Japanese macaques can learn to produce coos for food reward (see also Sugiura, 1998).

However, macaque coos are feeding-related vocalizations that both rhesus and closely related Japanese monkeys produce spontaneously in the context of foraging and the arrival of caretakers with food. Motivational state plays a strong role both in triggering these calls, and in shaping call acoustics (e.g., Green, 1975; Hauser & Marler, 1993; Owren et al., 1992b). It therefore seems likely that experiments demonstrating conditioned modification of these vocalizations have succeeded in large part because the calls are ones that are triggered by food-related cues under natural circumstances. In addition, acoustic dimensions such as rate, intensity, and duration are characteristics that are highly likely to be affected by vocalizer arousal and emotion in both primates (Rendall, 2003) and humans alike (Bachorowski & Owren, in press). Volitional control would be much more convincingly demonstrated if primate subjects could be trained to produce affectively unrelated sounds such as threat, alarm, or copulation calls for food reward, but that has not been done and is likely very difficult or impossible.

Brain-lesioning studies with rhesus monkeys have also shown that conditioned effects on coo acoustics critically involve the cingulate cortex, a limbic-system component (Sutton et al. 1974; Sutton et al., 1981; Trachy et al., 1981). This structure has reciprocal connections both to other limbic areas and to

neocortex, monitors overall internal state, and triggers emotion-related vocalization (Vogt & Barbas, 1986). While lesioning the cingulate abolishes conditioned vocalization, spontaneous calling and PAG-elicited sounds are not affected. In contrast, lesions in PAG abolish all vocalization (Jürgens, 2002). In other words, although cortically mediated learning is probably involved in conditioned vocalization, effects on call acoustics are nonetheless still mediated by emotion-related, subcortical regions. Taken together, then, both behavioral and neural evidence indicates that primate cortex cannot exert direct control over vocal output or trigger vocalizations independently of the caller's affective state.

#### *Responding to vocalizations*

*Flexibility in perception, cognition, and behavior.* While the sophisticated cognitive capabilities of monkeys and apes seem to take a back seat to relatively involuntary affective processes in vocal production, a very different picture emerges when these animals are responding to vocalizations. As in the case of vervets learning to execute particular escape strategies when hearing various alarm calls, primate responses to vocalizations are highly labile and subject to both short- and long-term modification. A listener can respond quite differently to a sound, depending on the call-type involved, the context in

which it hears the call, and its relationship to, as well as knowledge of the vocalizer.

Flexibility in responding to sound has in fact been amply demonstrated in decades of studies of captive primates. For example, behavioral tests of auditory processing have been heavily reliant on the ability of monkeys and apes to learn to perform arbitrary motor actions in response to particular acoustic features of interest (Moody, 1995; Niemiec & Moody, 1995; Sinnott, 1995). Captive primates tested in learning tasks can also readily associate sounds with other stimuli or events, a capability that has been shown based on both affective and instrumental responses. The same conclusion can be drawn from field studies, where primatologists playing back species-typical calls and other auditory stimuli to wild monkeys and apes necessarily rely on that flexibility in conducting their experiments (Cheney & Seyfarth, 1990a; Hauser, 1996; Rendall, 2004). Overall, results have been impressive in that regard. For example, researchers in laboratory and field have shown that listening animals can attend and respond to very subtle aspects of vocalizations, for example related to group membership, biological kinship, and dominance rank, as well as more individualized factors such as signaling reliability, recent calling behavior, or the emotional tone of recent interactions between vocalizer and listener.

*Widespread neural involvement.* The dramatic discrepancy in flexibility that primates show when producing versus responding to calls is consistent with differences in underlying neural connectivity. As noted, vocal production is primarily mediated by subcortical, limbic structures. In contrast, processing of sounds such as vocalizations includes both subcortical and cortical regions (Kaas et al., 1999). Lesioning and neurorecording studies have thus confirmed that the cortex is intimately involved when animals respond to any of a variety of natural or artificial sounds (Brosch & Scheich, 2003; Newman, 2003; Cheung et al., 2005). While vocal production is thus importantly "limbic" in nature, responding to vocalizations can be said to bring the entire primate brain into play.

#### HUMAN VOCAL BEHAVIOR

Based on the material reviewed so far, it seems difficult to make substantive connections between development of speech in human infants and vocal ontogeny in primates. Most importantly, learning and environment clearly play a critical role as the infant moves from prelinguistic sounds to language (e.g., Oller, 2000), while little auditory or motor learning is required for primates to produce acoustically appropriate calls. We suggest that vocal behavior in humans and primates may nonetheless be importantly related through the innate, nonlinguistic vocalizations and sophisticated auditory learning capabilities

present in both. The critical difference becomes that significant expansions of direct corticobulbar pathways in humans brings these characteristics into play as scaffolds that facilitate development of volitional vocal control and the emergence of prelinguistic sound production.

#### *Early vocal ontogeny*

Early vocal behavior in infants has long been a topic of interest to those studying the ontogeny of speech and language, for instance including nonlinguistic crying and prelinguistic sounds that can arguably be linked to later canonical babbling (Oller, 2000). Recent studies by Scheiner et al. (2002, 2004) examined both nonlinguistic and prelinguistic vocalizations produced during the first year of life, in particular comparing development in normally hearing versus hearing-impaired infants. Scheiner and colleagues documented the occurrence of at least 12 sound-types, including 10 that were compared between the two groups. Nine of these are in our view nonlinguistic in nature, including *cry*, *short cry*, *coo/wail*, *moan*, *whoop/squeal*, *groan*, *croak*, *hic*, and *laugh*, whereas the last category comprised prelinguistic *babble* sounds.

Little difference was found in either the acoustics or time of emergence for any of the nonlinguistic vocal-types, indicating that they are produced independently of auditory experience. These sounds included both noisy and harmonically

structured vocalizations, but as in primates, the acoustic distinctions involved were primarily laryngeal rather than supralaryngeal in origin. Scheiner et al. were also able to sort the sounds by overall type based on whether the vocalizing infant was most likely in a positive versus negative emotional state. Harmonically structured coo/wail sounds could, for instance, be further broken down into coos, occurring in positive contexts, and wails, occurring in negative contexts. The former were quieter and more vowel-like, while the latter were higher-pitched and less vowel-like, although remaining tonal in quality.

The development of babble sounds was markedly different, however. This category included a variety of more clearly prelinguistic vocalizations, and these sounds were either delayed or failed to appear in the repertoires of the hearing-impaired infants over the course of the study period. This difference between nonlinguistic and prelinguistic sounds was striking, but also consistent with earlier work that has documented lengthy delays in the onset of canonical babbling in hearing-impaired infants (Oller & Eilers, 1988; Eilers & Oller, 1994; Koopmans-van Beinum et al., 2001).

#### *Emergence of canonical babbling*

Although canonical babbling has often been considered a starting point for speech development, Oller (1980, 1986, 2000)



has presented a compelling argument for the occurrence of earlier prelinguistic stages. In his view, canonical babbling is importantly preceded by *quasi-resonant* vowels in a *simple phonation* stage (0-2 months), a *primitive articulation* stage (1-4 months) associated with gooing and *marginal syllables* (1-4 months), and an *expansion* stage (3-8 months) in which a variety of new sounds are produced. During the latter period, infants begin to position their vocal-tract articulators in new ways, produce sequences of articulated sounds, and exhibit slow, but recognizable consonant-vowel transitions (*marginal babbling*).

While the role of learning in these early stages is not well documented, the importance of auditory input for subsequent vocal ontogeny is clear from studying hearing-impaired infants. Reviewing the available evidence, Eilers and Oller (1994) suggested that canonical babbling begins only after the infant has accumulated some threshold level of auditory experience. This argument was based on finding a rather remarkable correlation of  $+0.69$  between the ages at which impaired infants received hearing aids and the age of canonical babbling onset. While a corresponding regression value was not reported, fitting a regression line to graphical data shown reveals an approximate slope of  $+0.75$  and a canonical babbling delay of at least 6 months. In contrast to nonlinguistic vocalization, then, emergence of canonical babbling requires a substantial amount of

previous auditory experience (see also Lynch et al., 1989, for evidence that the necessary threshold may also be reached through other sensory modalities) These findings thus belie any argument that canonical babbling should be considered primarily maturational or "genetic." Oller (2000) instead suggests that sounds occurring during all three prelinguistic stages are building blocks for the later speech-like, consonant-vowel transitions of canonical babbling.

*Neural mechanisms of volitional production*

While spontaneous, nonlinguistic vocalizations in both humans and primates are associated with specific limbic and brainstem structures, human speech involves integrated activity across a variety of structures (Lieberman, 2002). It is therefore common to view linguistic versus nonlinguistic vocalization in humans as involving distinct neural structures (Meyers, 1976; Ploog, 1988; Deacon, 1997), with cortical areas playing critical roles in both production and perception of speech (Jürgens, 2002; Lieberman, 2002). It is also likely not coincidental that cortex has a central role in speech and language, given the flexible and markedly volitional nature of this form of communication.

It is also not surprising that anatomical studies have revealed important differences between humans and primates in corticobulbar connectivity. As mentioned earlier, while primates

do have some direct cortical connections to the brainstem neurons innervating peripheral vocal-production anatomy it is not clear that these pathways are involved in vocalization, as opposed to other functions (Jürgens, 2002). Connections are less extensive than in humans, giving primates less opportunity for using the larynx, jaw, tongue, and face in an integrated, volitional fashion for the purpose of making sounds (Deacon, 1997). In addition, humans exhibit at least one corticobulbar pathway that is lacking in primates, namely connections from motor cortex to the *nucleus ambiguus* of the medulla. This pathway in particular is therefore believed to be central to volitional vocal production (Deacon, 1997; Jürgens, 2002).

#### THE BABBLING-SCAFFOLD HYPOTHESIS

The observations and empirical evidence outlined to this point present an intriguing pattern of similarities and differences between vocal behavior in humans and primates. A human infant's earliest vocalizations, for instance, are clearly primate-like, being innately grounded, nonlinguistic, and largely unarticulated. Conversely, primates resemble humans in showing pronounced auditory learning capabilities and flexible responding to vocalizations they hear from others. Nonetheless, whereas humans exhibit routine volitional control over vocal production, primates do not. At the neural level, this

difference likely has much to do with the more extensive corticobulbar connections present in humans.

We interpret these similarities and differences as suggesting that neither canonical nor even early babbling arises de novo during vocal ontogeny. Instead, we propose that the combination of innate, nonlinguistic vocalizations, typically primate auditory-learning capabilities, and uniquely human corticobulbar connections is key. While the first two are shared with other primates, the addition of direct corticobulbar pathways completes an otherwise open "circuit" between the production and reception sides of primate vocalization. Closing this loop creates the possibility of cortically controlled vocal communication, which is arguably realized through a self-organizing cascade of developmental events in which each stage becomes a foundation for the next level of communicative achievement. However, none of these events necessarily requires that humans have novel neural substrates or new learning capabilities relative to primates.

The next sections outline an argument for two developmental stages that we propose bring the infant from nonlinguistic to prelinguistic vocal behavior. The first stage involves acquisition of volitional control of laryngeally based vocal production, with nonlinguistic vocalizations suggested to be playing an important role. The second stage concerns development

of the rudimentary supralaryngeal control that is the hallmark of early prelinguistic behavior, and that is a prerequisite for the subsequent emergence of more speech-like sounds in canonical babbling.

*Nonlinguistic vocalizations, cortical maps, and laryngeal control*

Infants prototypically produce their earliest vocalizations shortly after birth in the form of crying. Other innately grounded, nonlinguistic sounds also emerge early, with Scheiner et al. (2004) reporting a total of 6 nonlinguistic sound-types appearing within the first 2 months of life. As in primates, crying and other similarly innate vocalizations are likely limbically controlled signals triggered by particular needs or emotional responses. Caregivers are very sensitive to these non-volitional sounds, as illustrated by the responsiveness of both parents and non-parents alike to crying (Green & Gustafson, 1997; Wood & Gustafson, 2001).

However, infants also rather quickly begin to exhibit some control over their nonlinguistic production, for example pausing during crying in an apparent effort to gauge caregiver reactions (Bell & Ainsworth, 1972). Over time, infants begin to use their sounds instrumentally as an attention-getting device and to coordinate vocalizations with gestures and other skeletal motor actions (Wolff, 1969; Papousek & Papousek, 1984; Gustafson &

Green, 1991; Lester & Boukydis, 1992). In other words, vocal behavior in human infants shows subtle but unmistakable elements of volitional control from an early age, thereby diverging from its initial, more purely primate-like form. This discrepancy is not likely to be traceable to major, between-species differences in caregiver reactions, as primate mothers are also responsive to infant distress vocalizations (Patel & Owren, 2007). Nor is the difference likely to reflect that primates cannot learn contingent relationships between vocalizing and caregiver response. As discussed earlier, primates show sophisticated learning about the social significance of calls heard from others, and laboratory studies have confirmed that conditioning can lead to affectively mediated changes in vocal production.

Instead, the discrepancy appears to be rooted in the elaboration of corticobulbar connections in humans, which allows infants to gain more direct, cortical control over vocalizations than is possible for primates. It is certainly the case that both primates and human infants receive proprioceptive and auditory feedback to cortical regions when producing affectively triggered, nonlinguistic vocalizations. However, only the human cortex in turn has direct contact with premotor neurons innervating the peripheral vocal anatomy that is producing these sounds.

This is a crucial difference in our view, with the convergence of motor and sound feedback being precisely what is needed to bootstrap the cortical sensorimotor mapping that is critical to language. That mapping has been explored by Westermann (this volume; Westermann & Miranda, 2004) and others, whose models have demonstrated that separate, but interconnected networks can produce highly coordinated, reciprocal mapping of perceptual and motor information. Kuhl and Meltzoff (1992) have made a similar point in the context of later speech production, arguing that a cortically mediated link between perceptual and motor experience is critical for mastering the sounds of a given language. They suggest that sensory experience with particular phonemes establishes stored auditory patterns that guide the infant's motor behavior as it works through successive approximations to the target sounds. Our proposal is that this kind of learning begins even earlier, with cortical maps first beginning to form based on the infant's own nonlinguistic production.

*Laryngeal control and pitch modulation.* The suggestion also includes that this first stage of vocal development most importantly concerns laryngeally based production. Volitional control first appears particularly evident in the infant's use of pitch, which reveals growing mastery of both vocal-fold tension and air pressure. Although vocal pitch is important in

primate vocalizations, the psychophysical evidence indicates that humans are especially sensitive to this acoustic dimension (Owren, 2003). The functional importance and salience of pitch variation to human infants has been amply demonstrated, for instance through their attentiveness to and preference for highly pitch-modulated "infant-directed speech" (Fernald, 1992). Caregivers are also very attentive to pitch in the infant's vocalizations, consistently using pitch patterning as a basis for interpreting their emotional significance for infants as young as 2 months of age (Papousek, 1989). Even for nonlinguistic sounds, adult listeners interpret vocalizations that end with rising pitch as expressing requesting or wanting (D'Odorico, 1984; Flax et al., 1991; Furrow, 1984; Masataka, 1993).

Analyzing the acoustic variation and social significance involved in pitch-modulated vocalization is furthermore heavily reliant on cortical processing – both in humans and primates alike. Once again, however, it is only in humans that the cortex in turn has extensive, direct contact with the brainstem neurons used to control vocal output. In the infant, these connections can create reciprocal contact between its nascent cortically based sensorimotor maps, social learning, and peripheral vocal-production anatomy. Early emergence of volitional control of the larynx in particular figures into this argument because pitch is



laryngeally based in vocal production. It primarily reflects the vocal-fold vibration rate, which is in turn a function of laryngeal muscle tensions and air pressure from the lungs.

Pitch also likely represents the avenue of greatest opportunity for infants that are beginning to gain some control over laryngeal production. Due to the perceptual prominence of pitch modulation, any control the infant can gain over pitch acoustics will be very salient both to itself and to caregivers. As discussed earlier, nonlinguistic vocalizations such as crying are also very salient, likely to both parties. However, many are grounded in less controllable, or even chaotic forms of vocal-fold vibration (Owren, 2003; Robb, 2003). Both for very rudimentary control that is first exerted over limbically mediated sounds and for more fully volitional sounds that the cortex is coming to directly control, the infant can be expected to work on readily achievable modulations that are nonetheless the most salient to itself and to others. We suspect that vocal pitch meets both these criteria.

The final point to make here is that the success the infant begins to enjoy in gaining control over its sounds is likely to be highly reinforcing. In other words, any initial achievement in volitional control can only foster greater motivation to explore the available "vocal space." As noted in the next section, such exploration is likely encouraged both by the self-

stimulation afforded by sound production, and through reciprocal effects on and from the environment. Here again, initial efforts are expected to focus on laryngeally produced modulations, and pitch modulation in particular being a prime candidate for early, purposeful exploration. Indeed, by 2-3 months of age, infants are routinely producing sounds with highly variable and dramatically exaggerated pitch changes, such as squeals (Oller & Griebel, this volume).

*From nonlinguistic to prelinguistic: Emerging articulation*

While the distinction drawn between nonlinguistic and prelinguistic vocalizations is conceptually useful, we are also proposing that the line is soon blurred by the infant's earliest, rudimentary attempts at volitional vocal control. Achieving even a modest degree over nonlinguistic sounds such as crying and squealing means that these vocalizations could also be considered prelinguistic in a sense. However, the distinction remains useful in that nonlinguistic vocalizations emerge at a very early age, including in hearing-impaired infants for whom the absence of auditory input means that fully volitional control and canonical babbling may never develop. Furthermore, it is clear from vocalizations such as adult laughter and crying that cortical and limbic vocal mechanisms never become fully, or

perhaps even substantially integrated. Human adults can inhibit or simulate these kinds of nonlinguistic sounds to some degree, but volitional suppression nonetheless fails in the face of strong emotion -- most people are likely unable to produce accurate and convincing, non-spontaneous versions of either laughter or crying.

Instead, it is volitional production of vowel-like sounds that prelinguistic infants first master, and that can be considered the primary vehicle of the nonlinguistic-to-prelinguistic transition we are proposing. In their nonlinguistic instantiation, Scheiner et al. (2002) and others refer to these sounds as coos, while for Oller (2000) the quasi-vowels in this early stage exemplify simple phonation. Oller further notes that these sounds are subject to an early, primitive form of articulation, and in this version might better be called goos. With increasing laryngeal control, we suggest that the initially emotion-triggered and limbically controlled vowel-like coos become the substrate of the next important steps in infant vocal development, namely the emergence of rudimentary articulation. This is another significant step, as supralaryngeal modulation has been found to play a prominent role in the more sophisticated social interactions that somewhat older infants enjoy with their caregivers.

The physiology of the vocal tract above the larynx is in fact changing significantly after the first 2 to 3 months of life. Prior to that point, the larynx is positioned high in the throat and overlaps the nasopharynx, which predisposes the infant to breathe nasally (Kent & Vorperian, 2006).

In addition, the small size of the oral cavity in very young infants restricts the range of possible tongue movement. By about 3 months of age, laryngeal descent has opened up the supralaryngeal vocal tract, reduced the previously nasalized character of vowel-like sounds, and significantly increased the diversity of acoustic effects achievable by repositioning the tongue within the oral and pharyngeal cavities.

We suggest that social learning is also playing a critical role at this early babbling stage. Infant vocal behavior is, for example, known to be responsive to contingent, trial-and-error learning long before the onset of traditional cognitive milestones of communication (Locke, 2001). Even simple caregiver responses such as touching and shaking a rattle, have been shown to be effective reinforcers of infant vocalizations that lead to increased rates of production (Rheingold et al., 1959; Weisberg, 1963; Routh, 1969; Poulson, 1983).

Associative learning is thus proposed to be at work even in very young infants, likely by first affecting incidental articulation effects occurring in nonlinguistic vocal

production. In one prominent example, Goldman (2001) found that caregivers report hearing infants as young as 1 month old produce their first "mama"-like sound in the context of crying. Emergence of "mama" sounds (or something similar) was found to peak at 2 to 3 months of age in this study, with the critical observation being that infants are not producing a fully formed word at this point. Rather, caregivers are interpreting incidental labial contact as an articulatory gesture, typically attributing communicative intent related to wanting or requesting.

While caregivers continue to be highly responsive to all manner of infant vocalizations at this age, coos and "a" sounds figure prominently among the sounds that elicit increasing proportions of verbal responses over the first 3 to 4 months (Keller & Schölmerich, 1987). In addition, caregivers have been found to be responsive specifically to vocal acoustics associated with speech-like supralaryngeal effect. Infants that produce more fully resonant sounds, for example, are rated as being more attractive or appealing by adults (Papousek, 1989; Bloom & Lo, 1990; Beaumont et al., 1993; Bloom et al., 1993).

Young infants are in turn very sensitive to sounds and responses from caregivers. For instance, experimental work has shown that 3-6 month-old infants readily and successfully imitate the absolute pitch of vocalizations (Kessen et al.,

1979), and produce more fully resonant sounds in vocal turn-taking with caregivers when these adults are responding contingently to this feature (Bloom et al., 1987). When exposed to combination of faces and voices with either matched or mismatched articulation, preverbal infants have also been shown to preferentially imitate vowels associated with the congruent pairings (Legerstee, 1990; see also Kuhl & Meltzoff, 1996).

Successful learning continues to keep the infant engaged with its environment throughout the early babbling stage, as demonstrated in experiments involving naturalistic play sessions. In these studies, caregivers are found to be more responsive to speech-like vocalizations that include fully resonant vowels and consonant-vowel-like transition than they are to sounds that lack these features (Keller & Schölmerich, 1987; Hsu & Fogel, 2003; Gros-Louis et al., 2006). When engaged in face-to-face interactions, mothers and infants take turns vocalizing (Anderson et al., 1977; Papousek et al., 1985), with both parties playing active roles in coordinating this joint behavior (Jaffe et al., 2001). As part of that process, mothers playing with their 2- to 5-month-old infants often match the infants' vocalizations, and do so accurately (Papousek & Papousek, 1989; Papousek, 1991). However, outcomes are very different when caregivers do not routinely produce speech or show well-coordinated responses to infant vocalizations. Hearing

infants of non-speaking deaf parents, for example, produce vocalizations that are acoustically disorganized and are delayed in using spoken language relative to infants with hearing parents (Petitto et al., 2004).

*The origins of canonical babbling*

Taken together, the evidence indicates that infants have a rich history of speech-relevant experience well before they begin to produce speech-like, prelinguistic sounds. As noted earlier, however, work on early language development often takes canonical babbling as the starting point (e.g., MacNeilage, 1998). Further, some approaches characterize canonical babbling as being preprogrammed, preordained, or a maturational outcome. Deacon (1997), for example, points to myelination of cortical neurons as the critical factor in the emergence of canonical babbling. While these sorts of maturational events are likely both necessary and important, we suggest that the evident impact of auditory input and social engagement on vocalization indicates that they cannot in and of themselves be sufficient explanations for the emergence of canonical babbling.

REFLECTIONS ON THE HYPOTHESIS

The babbling-scaffold hypothesis differs most fundamentally from traditional views of infant vocal development in imputing a central role for learning virtually from the moment of birth. On the production side, the infant is argued to first acquire

volitional control of the larynx, while also more gradually gaining proficiency in supralaryngeal articulation. Self-stimulation through vocalization is proposed to be important throughout this development, beginning with cortical mapping initially produced by nonlinguistic vocalization, but then continuing as early babbling gives rise to canonical babbling and then full-blown speech. Finally, social learning is seen as being critical at every stage of the process. In this view, even the earliest vocal behavior is importantly subject to contingency learning, with nonlinguistic and prelinguistic vocal behavior both promoting and being influenced by social interaction with caregivers. Although discussion is necessarily brief, we now examine some of the assumptions and implications of our approach in the larger context of development, the effects of hearing impairment, and vocal ontogeny in humans and primates.

#### *Self-stimulation in development*

One premise of the babbling-scaffold hypothesis is that the infant's own behavior plays an important role throughout its vocal development, specifically including a facilitating effect of innately grounded, nonlinguistic sounds on cortical mapping and volitional control of the larynx. Although this argument may seem unusual, it is in fact common for behavior or activity generated by one part of the immature brain to be critical for



some other behavior or brain area. In comparative psychology, such occurrences are well-known and have played a central role in nature-nurture debates (Ho, 1998).

In a classic study of chick development, for instance, Kuo (1966) showed that the embryonic heartbeat plays an important role in stimulating and entraining raising and lowering of the head, opening and closing of the beak, and later swallowing of amniotic fluid. While still in the egg, in other words, activity of the chick's heart stimulates development of coordinated movements in quite different systems. At the purely neural level, many instances of spontaneous correlated activity in one part of the nervous system providing critical experience to neural circuits in other regions in advance of external stimulation have recently been uncovered (Feller 1999; Wong 1999).

There are fewer demonstrations of this kind specifically in the vocal domain, but some telling examples are available. For instance, work by Gottlieb (1963) has demonstrated that a newly hatched but socially isolated wood duckling can recognize conspecific vocalizations expressly because it has heard its own calls while in the egg. The influence that vocal self-stimulation can have across brain systems has been shown even more dramatically by Cheng (1992). In this work, the "nest" coos of female ring doves were found to stimulate hormone release in

the vocalizer herself, thereby playing a functional role in advancing the courtship and mating process with a male. While the brain systems are different in the two cases, this last example in particular illustrates the kind of self-stimulation we are proposing to be important in the early stages of human vocal development.

#### *Hearing impairment and canalization*

In proposing that experience plays a critical role throughout infant vocal development, the babbling-scaffold hypothesis must also account for the seeming imperturbability of this process up to and including the onset of canonical babbling. Here, the important observation is that canonical babbling emerges predictably in the face of a variety of possible risk factors, including deprivation due to socio-economic circumstances, hearing multiple languages in early infancy, and mental retardation (Oller, 2000). Although the reliable emergence of canonical babbling in spite of such handicaps has been taken as evidence of biological depth and canalization, it need not be. Finding that a trait emerges in both normative and non-normative circumstances does not in and of itself imply that it is guided by biology, as it could be that critical experiential factors remain operative in each instance.

In the case of vocal ontogeny, we suggest that none of the important factors – the presence of a primate-like, nonlinguistic repertoire, development of nascent cortical mapping through proprioceptive and auditory feedback, and contingent caregiver responses to vocalization – are likely to be importantly affected by socio-economic circumstances or simultaneous exposure to multiple languages. The learning processes proposed to be at work during the nonlinguistic and early prelinguistic stages are also expected to be robust to many mental retardation effects.

However, attaining full-fledged canonical babbling and later speech is a long and challenging process. The infant must go far beyond the rudimentary volitional control of laryngeal and supralaryngeal production that we have focused on, including integration of auditory, motor, and social input that is both qualitatively and quantitatively much more complex. While we thus expect mentally retarded children to routinely acquire volitional vocal control and rudimentary speech, mature language depends both on being able to process speech effectively and to be sensitive to a variety of kind of feedback to their sounds. If either aspect is impaired, language development is likely to be requisitely incomplete.

The babbling-scaffold approach is arguably also consistent with finding a strong correlation between the age at which

hearing-impaired infants receive effective hearing aids and observed delays in the onset of canonical babbling. From this perspective, significant hearing impairment must have immediate and detrimental effects on vocal development, for instance in hindering both self-stimulation by the infant's nonlinguistic vocalizations and early learning about caregiver responses. Consistent with Eilers and Oller's (1994) report, the hypothesis therefore predicts that even hearing impairment that is corrected very early will delay vocal development at later stages, particularly the acquisition of volitional control required for canonical babbling. While this prediction may seem obvious, it does not follow from the alternative perspective of viewing nonlinguistic vocalization, early babbling, and canonical babbling as being separable developmental events. The same rationale applies to finding that the amount of delay in canonical babbling is proportional to the age at which a hearing-impaired infant receives aid, although that expectation is likely compatible with a variety of theories of babbling. A more specific prediction is that at the time of receiving a hearing aid, the infant's progress in volitional control of laryngeal production, instrumental use of nonlinguistic vocalization, and its transition from nonlinguistic to prelinguistic sounds should be correlated with the amount of onset delay observed later for canonical babbling.

*Vocal development in humans and primates*

A central theme of the chapter has been that vocal development and vocal behavior are fundamentally different in humans and primates, yet also importantly similar. For primates, we have emphasized the separation between vocal production and reception, and traced it to a lack of direct cortical connections to brainstem neurons innervating peripheral vocal anatomy. Consistent with this observation, projects attempting to teach apes to produce spoken language have met with little success despite heroic training efforts (Hayes, 1951; Kellogg & Kellogg, 1933). In contrast, bonobos, chimpanzees, orangutans, and gorillas have all been able to acquire significant language-related skills when the medium has involved manually controlled actions such as gesturing or selecting among visual symbols. The key difference is most likely that these animals have full volitional control over these kinds of motor actions, but not over vocal production.

It follows that the great divide between vocal communication in humans and primates does not merely reflect differences in relative cognitive abilities. In fact, primates have been found to best demonstrate language-relevant skill when the tasks involve the reception rather than the production side of the equation. Monkeys, for instance, have been found to be sensitive to grammar-like statistical regularities in auditory

input (Hauser et al., 2001; Fitch & Hauser, 2004), and apes exposed to spoken language have acquired significant comprehensive vocabularies, even in the absence of explicit training (Brakke & Savage-Rumbaugh, 1995; Lyn & Savage-Rumbaugh, 2000; Williams et al., 1997).

We interpret such evidence to suggest that at least some primate species are probably capable of forming the sensorimotor maps, semantic representations, and illocutionary motivations involved in human language. In fact, linked cortical mapping of proprioceptive and auditory information from hearing their own calls may be routine in primates, at least at a rudimentary level. If so, however, neither that mapping nor their demonstrated processing capabilities can take them into the realm of speech-like communication in the absence of more direct cortical control of vocal production. We have argued that these corticobulbar connections are critical at every stage of vocal-production learning in human infants, and that without them, primates have no possibility of even beginning such a process. Instead, animals whose communicative production is mediated solely or primarily through affective and limbic mechanisms will forever remain inflexible and context-bound. While we do not claim that the simple appearance of expanded corticobulbar connections in a primate would inexorably lead to language-like vocal communication, we do argue that it would be a critical

step. In humans, we see these pathways as the linchpin of a developmental process that may not have required any other special abilities or evolutionary adaptations. Rather, we suggest that the developmental process contributes much of the uniqueness of spoken language, proceeding as a cascade in which each new capability and its associated behaviors opens the door to additional and qualitatively more sophisticated communicative skills and interactions.

CONCLUSIONS: DEVELOPMENT OF CONTEXTUAL FLEXIBILITY  
IN VOCAL COMMUNICATION

The ubiquity of spoken language in humans can sometimes lessen the appreciation of its sophistication and complexity, and the extended period of development that is involved in learning to use it. Thus, while the rapid pace of speech ontogeny can suggest innateness, the learning required for becoming a fully competent speaker of a language extends well into middle childhood or beyond (Ferguson et al., 1992; Kuhl, 2007). In this chapter, we have argued for a critical role of learning from the very earliest stages of prelinguistic production as well, by linking the emergence of volitionally controlled prelinguistic sounds to the innate and primate-like nonlinguistic vocalizations that precede them. While the

connections are indirect, we are thereby arguing that continuity does exist between vocal communication in primates and spoken language in humans. However, the two are nonetheless qualitatively discontinuous, with the expanded corticobulbar connections present in humans opening the door for speech-related sensorimotor and instrumental learning processes that primates might be capable of, but have no opportunity to realize.

The view that human vocal development is a process that begins virtually at birth has guided proposals about self-stimulation through nonlinguistic vocalization, as well as gradual acquisition of volitional control over first laryngeal and then supralaryngeal production. Some predictions follow, including that even very early correction of hearing impairment should speed vocal development, and that the child's progress in instrumental and volitional control of vocalization should be correlated with the amount of delay observed in canonical babbling. Another expectation that flows from the proposed connection between nonlinguistic and prelinguistic vocalizations in early infancy is that greater variety in the earliest nonlinguistic vocalizations in a hearing infant should produce earlier emergence of volitional vocal abilities as a result of forming better sensorimotor cortical mapping. Finally, although hearing-impaired infants produce the same kinds of nonlinguistic



vocalizations as their normally hearing peers, we expect that more detailed comparisons will reveal that they are nonetheless being hindered in the instrumental use of sounds and in the transition from nonlinguistic to volitional control of vocal production. If so, these are impediments occurring long before the well-documented canonical-babbling delays occurring in impaired infants, and are likely to be contributing to those effects.

#### *Contextual flexibility*

In its simplest form, the babbling-scaffold hypothesis argues that the unique flexibility of spoken language in humans may ultimately be traceable to combining an evolutionary innovation (increased corticobulbar connections) with an evolutionary legacy (innate vocal production but more sophisticated auditory learning). We have throughout put cortical control over vocal production at the center, more or less equating it with volitional control. That characterization is of course over-simplified, particularly in light of evidence that many brain areas are involved in language-related processing, both in production and in comprehension. We can nonetheless assert that the absence of direct cortical control over peripheral vocal anatomy that characterizes many mammals

other than humans powerfully restricts the flexibility that can be achieved in their vocal signaling. There is perhaps some irony in noting that these are asymmetrical constraints, with the potential for flexibility and sophistication in signal reception and processing far outstripping possible vocal expressiveness. Bringing the cortex into play in vocal production may thus have been quite a small evolutionary change occurring in some early human ancestor, while the symmetry that was thereby created between producing and responding to vocal signals was likely the key unlocking the vast potential of contextually flexible communication.

## REFERENCES

- Anderson BJ, Vietze P, Dokecki PR (1977) Reciprocity in vocal interactions of mothers and infants. *Child Devel.* 48: 1676-1681.
- Bachorowski J-A, Owren M J (in press) Vocal expressions of emotion. In: *The Handbook of Emotion*, 3rd Ed. (Lewis M, Haviland-Jones JM, Barrett LF, eds), New York: Guilford Press.
- Beaumont SL, Bloom K (1993) Adults' attributions of intentionality to vocalizing infants. *First Lang.* 13: 235-247.
- Bell SM, Ainsworth MDS (1972) Infant crying and maternal responsiveness. *Child Devel.* 43: 1171-1190.
- Bloom K, D'Odorico L, Beaumont S (1993) Adult preferences for syllabic vocalizations: Generalizations to parity and native language. *Inf. Beh. Devel.* 16: 109-120.
- Bloom K, Lo E (1990) Adult perceptions of vocalizing infants. *Inf. Behav. Devel.* 13: 209-219.
- Bloom K, Russell A, Wassenberg K (1987) Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14: 211-227.
- Brakke K, Savage-Rumbaugh ES (1995) The development of language skills in bonobo and chimpanzee—I. Comprehension. *Lang. Comm.* 15: 121-148.

- Brosch M, Scheich H (2003) Neural representation of sound patterns in the auditory cortex of monkeys. In: Primate Audition: Ethology and Neurobiology (Ghazanfar A, ed), pp 151-175. Boca Raton, Fl: CRC Press.
- Brown CH, Alipour F, Berry DA, Montequin D (2003) Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*). J. Acoust. Soc. Am. 113: 2114-2126.
- Chapman CA, Lefebvre L (1990) Manipulating foraging group size: Spider monkey food calls at fruiting trees. Anim. Behav. 39: 891-896
- Cheney DL, Seyfarth RM (1990a) How Monkeys See the World. Chicago: University of Chicago Press.
- Cheng M-F (1992) For whom does the female dove coo? A case for the role of self-stimulation. Anim. Behav. 43: 1035-1044.
- Cheung SW, Nagarajan SS, Schreiner CE, Bedenbaugh PH, Wong A (2005) Plasticity in primary auditory cortex of monkeys with altered vocal production. J Neurosci. 25:2490-2503.
- Clark AP, Wrangham RW (1993) Acoustic analysis of chimpanzee pant hoots: do chimpanzees have an acoustically distinct food arrival pant hoot? Am. J. Primatol. 31: 99-109.
- Deacon TW (1989) The neural circuitry underlying primate calls and human language. Hum. Evol. 4: 367-401.
- Deacon TW (1997) The Symbolic Species. New York: Norton.

- Dennett D (1983) Intentional systems in cognitive ethology: The "Panglossian" paradigm defended. *Behav. Brain Sci.* 6: 343-355.
- D'Odorico L (1984) Non-segmental features in prelinguistic communications: An analysis of some types of infant cry and non-cry vocalizations. *J. Child Lang.* 11: 17-27.
- Doupe AJ, Kuhl PK (1999). Birdsong and human speech: Common themes and mechanisms. *Ann. Rev. Neurosci.* 22, 567-631.
- Eilers RE, Oller DK (1994) Infant vocalizations and the early diagnosis of severe hearing impairment. *J Pediat.* 124: 199-203.
- Elowson, AM, Snowdon CT, Lazaro-Perea C (1998) "Babbling" and social context in infant monkeys: Parallel to human infants. *TICS* 2: 35-43.
- Feller MB (1999) Spontaneous correlated activity in developing neural circuits. *Neuron* 22: 653-656.
- Fernald A (1992) Maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In: *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Barkow JH, Cosmides L, Tooby J, eds.) pp 391-428. Oxford: Oxford University.
- Ferguson CA, Menn L, Stoel-Gammon C (eds.) (1992). *Phonological Development: Models, Research, Implications*. Timonium MD: York Press.

- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Sci.* 303: 377-380
- Flax J, Lahey M, Harris K, Boothroyd, A (1991) Relations between prosodic variables and communicative functions. *J. Child Lang.* 18: 3-20.
- Furrow D (1984) Young children's use of prosody. *J. Child Lang.* 11: 201-213.
- Fry DB (1966) The development of the phonological system in the normal and deaf child. In: *The Genesis of Language* (Smith F, Miller GA, eds) pp 187-206. Cambridge, MA: MIT Press.
- Goldman HI (2001) Parental reports of 'MAMA' sounds in infants: An exploratory study. *J. Child Lang.* 28: 497-506.
- Goldstein MH, King AP, West MJ (1998) Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *J. Comp. Psychol.* 113: 52-58.
- Goldstein MH, King AP, West MJ (2003) Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Nat. Acad. Sci.* 100: 8030-8035.
- Gottlieb G (1963) A naturalistic study of imprinting in wood ducklings (*Aix sponsa*). *J. Comp. Phys. Psych.* 56: 86-91.
- Green S (1975) Variation of vocal pattern with social situation in the Japanese macaque (*Macaca fuscata*): A field study.

In: Primate Behavior Vol. 4 (Rosenblum LA, ed). New York: Academic Press.

Green JA, Gustafson GE (1997) Perspectives on an ecological approach to social communicative development in infancy. In: Evolving explanations of development: Ecological explanations to organism-environment systems (Dent-Read C, Zukow-Goldring P, eds), pp 515-546. Arlington, VA: American Psychological Association.

Gros-Louis J, West MJ, Goldstein MH, King AP (2006) Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Devel.* 30: 509-516.

Hammerschmidt K, Freudenstein T, Jürgens U (2001) Vocal development in squirrel monkeys. *Behav.* 138: 1179-1204.

Hauser MD (1996) *The Evolution of Communication*. Cambridge MA: MIT Press.

Hauser MD, Evans CS, Marler P (1993) The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations. *Anim. Behav.* 45: 423-433.

Hauser MD, Marler P (1993) Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behav. Ecol.* 4: 194-205.

Hauser MD, Newport EL, & Aslin RN (2001) Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cog.* 78: B53-B64.

- Hayes C (1951) *The Ape in Our House*. New York: Harper.
- Ho, M-W (1998). Evolution. In: *Comparative Psychology, a Handbook* (Greenberg G, Haraway MM, eds.), pp 107-119. New York: Garland.
- Hsu H, Fogel A (2003) Stability and transitions in mother-infant face-to-face communication during the first 6 months: A microhistorical approach. *Devel. Psychol.* 39: 1060-1082.
- Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD (2001) Rhythms of dialogue in infancy. *Monographs Soc. Res. Child Devel.* 66: Serial No. 265.
- Jürgens U (2002) Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26: 235-238.
- Kaas JH, Hackett TA, Tramo MJ (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobio.* 9: 164-170.
- Keller H, Schölmerich A (1987) Infant vocalizations and parental reactions during the first 4 months of life. *Devel. Psychol.* 23: 62-67.
- Kellogg WN, Kellogg LA (1933) *The ape and the child: A study of environmental influence upon early behavior*. New York: McGraw-Hill.
- Kent RD, Vorperian HK (2006) In the mouths of babes: anatomic, motor, and sensory foundations of speech development in children. In: *Language Disorders From a Developmental Perspective* (Paul R, ed). Mahwah, NJ: Lawrence Erlbaum.



- Kessen W, Levine J, Wendrich KA (1979). The imitation of pitch in infants. *Inf. Behav. Devel.* 2: 93-99.
- Knörnschild M, Behr O, von Helversen O (2006) Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissen.* 93: 451-454.
- Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science*, 4, 61-70.
- Kuhl PK (2007) Is speech 'gated' by the social brain? *Devel. Sci.* 10: 110-120.
- Kuhl PK, Meltzoff AN (1992) The bimodal perception of speech in infancy. *Sci.* 218: 1138-1141.
- Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100, 2425-2438.
- Kuo ZY (1966) *The Dynamics of Behavior Development: An Epigenetic View*. New York: Random House.
- Larson CR, Kistler MK (1987) Brainstem neuronal activity associated with vocalization in the monkey. In: *Vocal Fold Physiology: Laryngeal Function in Phonation and Respiration* (Baer T, Harris KS, Sasaki C), pp 154-167. Boston: Little, Brown.

- Legerstee M (1990) Infants use multimodal information to imitate speech sounds. *Inf. Behav. Devel.* 13: 343-354.
- Lester BM, Boukydis CFZ (1992) Infantile colic: Acoustic cry characteristics, maternal perception of cry and temperament. *Inf. Behav. Devel.* 15: 15-26.
- Lieberman P (1975) *On the Origins of Language*. New York: Macmillan.
- Lieberman P (2002) The neural bases of language. *Yearbook Phys. Anthro.* 45: 36-62.
- Locke J (2001) First communion: The emergence of vocal relationships. *Soc. Devel.* 10: 294-308.
- Lyn H, Savage-Rumbaugh, ES (2000). Observational word learning in two bonobos (*Pan paniscus*): Ostensive and non-ostensive contexts. *Lang. Comm.* 20: 255-273.
- Lynch, M. P., Oller, D. K., & Steffens, M. (1989). Development of speech-like vocalizations in a child with congenital absence of cochleas: The case of total deafness. *Applied Psycholinguistics*, 10, 315-333.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences*, 21, 499-546.
- Masataka N (1992) Attempts by animal caretakers to condition Japanese macaque vocalizations result inadvertently in individual-specific calls (Nishida T, McGrew WC, Marler P,

- Pickford M, de Waal FBM, eds), pp 271-278. Tokyo: University of Tokyo.
- Masataka N (1993) Relation between pitch contour of prelinguistic vocalizations and communicative functions in Japanese infants. *Inf. Behav. Devel.* 16: 397-401.
- Masataka, N., & Bloom, K. (1994). Acoustic properties that determine adults' preferences for 3-month-old infant vocalizations. *Inf. Behav. Devel.* 17: 461-464.
- Moody DB (1995) Classification and categorization procedures. In: *BioMethods, Vol. 6, Methods in comparative acoustics* (Klump GM, Dooling, RJ, Fay, RR, Stebbins WC, eds), pp 293-305. Basel, Switzerland: Birkhauser Verlag.
- Myers RE (1976) Comparative neurology of vocalization and speech: Proof of a dichotomy. *Ann. NY Acad. Sci.* 280: 745-757.
- Newman JD (2003) Auditory communication and cortical auditory mechanisms in the squirrel monkey: Past and present. In: *Primate Audition: Ethology and Neurobiology* (Ghazanfar A, ed), pp 227-246. Boca Raton, Fl: CRC Press.
- Niemiec AJ, Moody DB (1995) Constant stimulus and tracking procedures for measuring sensitivity. In: *BioMethods, Vol. 6, Methods in comparative acoustics*, (Klump GM, Dooling RJ, Fay RR, Stebbins WC, eds), pp 65-78. Basel, Switzerland: Birkhauser Verlag.

- Oller DK (1980) The emergence of speech sounds in infancy. In: Child Phonology: Vol. 1. Production (Yeni-Komshian G, Kavanagh J, Ferguson C, eds), pp 93-112. New York: Academic Press.
- Oller DK (1986) Metaphonology and infant vocalizations. In: Precursors of Early Speech (Lindblom B, Zetterstrom Z, eds), pp 21-35. New York: Stockton Press.
- Oller DK (2000) The Emergence of the Speech Capacity. Mahwah NJ: Erlbaum.
- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59, 441-449.
- Owren MJ (2003) Vocal production and perception in nonhuman primates provide clues about early hominids and speech evolution. ATR Symp. HIS Ser. 1: 1-19.
- Owren MJ, Dieter JA, Seyfarth RM, Cheney DL (1992a). Evidence of limited modification in the vocalizations of cross-fostered rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques. In: T. Nishida T, McGrew WC, Marler P, Pickford M, de Waal F, eds), Topics in Primatology, Vol. I: Human Origins, pp 257-270. Tokyo: University of Tokyo.
- Owren MJ, Dieter JA, Seyfarth RM, Cheney DL (1992b) 'Food' calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species.

- Behav. 120: 218-231.
- Owren MJ, Dieter JA, Seyfarth RM, Cheney DL (1993) Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Devel. Psychobio.* 26: 389-406.
- Owren MJ, Rendall D (1997) An affect-conditioning model of nonhuman primate vocal signaling. In: *Perspectives in Ethology: Vol. 12. Communication* (Owings DH, Beecher MD, Thompson NS, eds), pp 299-346. New York: Plenum Press.
- Patel ER, Owren MJ (2007) Acoustic and behavioral analyses of gecker distress vocalizations in young rhesus macaques (*Macaca mulatta*). *J. Acoust. Soc. Am.* 121: 575-585.
- Papousek H, Papousek M (1984) Qualitative transitions during the first trimester of human postpartum life. In: *Continuity of Neural Function Prenatal to Postnatal Life* (Prechtl HFR, ed), pp 220-244. London: Spastics International Medica.
- Papousek, M. (1989) Determinants of responsiveness to infant vocal expression of emotional state. *Inf. Behav. Devel.* 12: 507-524.
- Papousek, M. (1991) Early ontogeny of vocal communication in parent-infant interactions. In: *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (Papousek H, Jürgens U, Papousek M, eds.), pp 230-261. Cambridge, England: Cambridge University.

- Papousek M, Papousek H (1989) Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Lang.* 9: 137-158.
- Papousek M, Papousek H, Bornstein MH (1985) The naturalistic vocal environment of young infants. In: *Social Perception in Infants* (Field TM, Fox N, eds) pp 269-298. Norwood, NJ: Ablex.
- Petitito LA, Holowka S, Sergio LE, Levy B, Ostry DJ (2004) Baby hands that move to the rhythm of language: Hearing babies acquiring sign language babble silently on the hands. *Cog.* 93: 43-73.
- Pierce JD (1985) A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26: 202-213.
- Ploog D (1988) Neurobiology and pathology of subhuman vocal communication and human speech. In: *Primate Vocal Communication* (Todt D, Goedecking P, Symmes D, eds) pp 195-212. Berlin: Springer-Verlag.
- Poulson CL (1983) Differential reinforcement of other-than-vocalization as a control procedure in the conditioning of infant vocalization rate. *J Exp Child Psychol.* 36: 471-89.
- Pulvermüller F (1999) Words in the brain's language. *Behav. Brain Sci.* 22: 253-336.
- Pulvermüller F, Preissl H (1991) A cell assembly model of language. *Comput. Neural Sys.* 2: 455-468.

- Rendall D (2003) Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J. Acoust. Soc. Am.* 113: 3390-3402.
- Rendall D (2004) 'Recognizing' kin: Mechanisms, media, minds, modules and muddles. In: *Kinship and Behaviour in Primates* (Chapais B, Berman C, eds), pp 295-316. Oxford: Oxford University.
- Rendall D, Cheney DL, Seyfarth RM (2000) Proximate factors mediating 'contact' calls in adult female baboons and their infants. *J. Comp. Psychol.* 114: 36-46.
- Rheingold HL, Gerwitz JL, Ross, HW (1959) Social conditioning of vocalizations in the infant. *J. Comp. Physiol. Psychol.* 52: 68-73.
- Riede TR, Zuberbühler K (2003) Pulse register phonation in Diana monkeys alarm calls. *J. Acoust. Soc. Am.* 113: 2919-2926,
- Robb MP (2003) Bifurcations and chaos in the cries of full-term and preterm infants. *Folia. Phoniatr. Logop.* 55: 233-240
- Routh DK (1969) Conditioning of vocal response differentiation in infants. *Devel. Psychol.* 1: 219-226.
- Scheiner E, Hammerschmidt K, Jürgens U, Zwirner P (2002) Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *J. Voice* 16: 509-529,

- Scheiner E, Hammerschmidt K, Jürgens U, Zwirner P (2004) The influence of hearing impairment on preverbal emotional vocalizations of infants. *Folia Phoniat. Logopaed.* 56: 27-40.
- Seyfarth RM, Cheney DL (2003) Signalers and Receivers in Animal Communication. *Ann. Rev. Psychol.* 54: 145-173
- Seyfarth RM, Cheney DL, Marler P (1980) Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav.* 28: 1070-1094.
- Sinnott J (1995) Methods to assess the processing of speech sounds by animals. In: *BioMethods, Vol. 6, Methods in comparative acoustics*, (Klump GM, Dooling RJ, Fay RR, Stebbins WC, eds), pp 281-292. Basel, Switzerland: Birkhauser Verlag.
- Sugiura H (1998) Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Anim. Behav.* 55: 673-687.
- Sutton D, Larson C, Lindeman RC (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Res.* 71: 61-75.
- Talmadge-Riggs G, Winter P, Ploog D, Mayer W (1972) Effects of deafening on the vocal behavior of the squirrel monkey. *Folia Primatol.* 17: 404-420.



- Trachy RE, Sutton D, Lindeman RC (1981). Primate phonation: Anterior cingulated lesion effects on response rate and acoustical structure. *Am. J. Primat.* 1: 43-55.
- Vogt BA, Barbas H (1986) Structure and connections of the cingulated vocalization region in the rhesus monkey. In: *The Physiological Control of Mammalian Vocalization* (Newman JD, ed), pp 203-225. New York: Plenum.
- Weisberg P (1963) Social and nonsocial conditioning of infant vocalizations. *Child Devel.* 34: 377-388.
- Westermann G, Miranda E (2004) A new model of sensorimotor coupling in the development of speech. *Brain Lang.* 89: 393-400.
- Williams SL, Brakke KE, Savage-Rumbaugh ES (1997) Comprehension skills of language-competent and nonlanguage-competent apes. *Lang. Comm.* 17: 304-317.
- Winter P, Handley P, Ploog D, Schott D (1973) Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behav.* 47: 230-239.
- Wolff PH (1969) The natural history of crying and other vocalizations in early infancy. In: *Determinants of Infant Behavior Vol. 4* (Foss BM, ed), pp 113-138. London: Methuen.
- Wong RO (1999) Retinal waves and visual system development. *Ann. Rev. Neurosci.* 22: 29-47.
- Wood RM, Gustafson GE (2001) Infant crying and adults'

anticipated caregiving responses: Acoustic and contextual influences. *Child Devel.* 72: 1287-1300.

### Acknowledgments

This work was supported in part by NIMH Prime Award 1 R01 MH65317-01A2, Subaward 8402-15235-X, and in part by the Center for Behavioral Neuroscience, STC Program of the National Science Foundation under Agreement No. IBN-9876754. Many thanks to Kurt Hammerschmidt, Chuck Snowdon, and Kim Oller for their helpful comments on earlier versions of this chapter.