The Onset of Language

Nobuo Masataka

Primate Research Institute, Kyoto University, Japan

CAMBRIDGE
UNIVERSITY PRESS

published by the press syndicate of the university of cambridge
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

cambridge university press
The Edinburgh Building, Cambridge, CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 10/12 pt. *System* $\LaTeX\ 2_{\varepsilon}$ [tb]

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication data
Masataka, Nobuo, 1954–
The Onset of Language / Nobuo Masataka.
p. cm. – (Cambridge studies in cognitive and perceptual development)
Includes bibliographical references and index.
ISBN 0 521 59396 4
1. Language acquisition. I. Title. II. Cambridge studies in cognitive perceptual development.
P 118.M367 2003
401′.93 – dc21 2003043933

ISBN 0 521 59396 4 hardback

Contents

List of figures *page* vi
List of tables ix
Acknowledgements xi

1. Introduction 1
2. The development of the ability to take turns 27
3. Cooing in three-month-old infants 64
4. The development of vocal imitation 97
5. How infant-directed speech influences infant vocal development 126
6. From laughter to babbling 157
7. Earliest language development in sign language 188
8. From babbling to speaking 217
9. Summary and conclusion 241

References 250
Index 273

Figures

2.1 Distribution of intervals between sequential chucks by different callers *page* 30
2.2 Distribution of intervals between sequential chucks by same callers 31
2.3 Overall distributions of intervals between sequential coos by different callers (DC sequences) and of intervals between sequential coos by same callers (SC sequences) in the two Japanese macaque populations 37
2.4 A representative sonagram of a Japanese macaque coo call, in which the acoustic parameters measured for acoustic analyses are indicated 40
2.5 Distributions of intervals between two consecutive 3-month-old infant vocalizations in the contingent group (GpC-3) and the random group (GpR-3) 48
2.6 Distributions of intervals between two consecutive 4-month-old infant vocalizations in the contingent group (GpC-4) and the random group (GpR-4) 49
2.7 Distributions of intervals between the beginning of infant vocalization and the beginning of the mother's contingent response in the contingent group of 3-month-olds (GpC-3) and in the contingent group of 4-month-olds (GpC-4) 51
2.8 Comparison of distributions of intervals between two consecutive bursts in sucking by 2-week-old infants among three experimental groups 56
2.9 Comparison of distributions of intervals between two consecutive bursts in sucking by 8-week-old infants among three experimental groups 57
3.1 Comparison of preference scores in terms of the quality of vocalization and of manual action 84
3.2 Waddington's (1957) epigenetic landscape 95
4.1 Log frequency plots of the intervals between playback of

stimulus coo calls from an adult female Japanese macaque and the response coos of seven females 100
4.2 Regression of the maximum minus start frequency of the response coos of seven Japanese macaques on that of the playback stimuli (coo calls from an adult female) 102
4.3 Percentage of cross-validated vocalizations that were misclassified in each dyad in the discriminant analysis on the infant-initiated interaction as a function of the age of the infant in the dyad 118
4.4 Percentage of cross-validated vocalizations that were misclassified in each dyad in the discriminant analysis on the mother-initiated interaction as a function of the age of the infant in the dyad 119
5.1 Average pitch (a) and pitch range (b) for initial maternal utterances in A-I speech 145
6.1 Representative sonagrams of pant-hoot vocalizations of chimpanzees in two populations, Mahale and JMC 165
6.2 Overall developmental patterns in the occurrence of rhythmic leg and hand movements and laughter 171
6.3 Developmental changes in co-occurrences between laughter, and rhythmic leg and hand movement 172
6.4 The development of relative differences between the number of syllables in a laugh and the number of repetitions in co-occurring kicking and banging 173
6.5 Developmental changes in syllable length, laughter and babbling 175
6.6 Overall developmental patterns of marginal babbling and canonical babbling (upper) and of co-occurrences between babbling and banging (lower) 177
7.1 Schematic representation of the plane view (A) and the side view (B) of the mothers as recorded on film 194
7.2 Comparison of responsiveness for infant-directed (ID) signing versus adult-directed (AD) signing between hearing infants and deaf infants 199
7.3 Number of occurrences of signs and nonreferential gestures in two deaf infants, S1 (top) and S2 (bottom), as a function of age in months 203
7.4 Density of space usage in manual babbling of deaf infants and hearing infants at the ages of 6 months and 9 months 207
7.5 Comparison of trajectories of the right hand in the frontal plane in manual babbling between two deaf infants at the age of 9 months 209

7.6 Comparison of five types of manual babbling and the first recognized signs that appear to have developed from each, as produced by deaf infants 214
8.1 Mean number (per 30 min) of infants' manual activities as a function of their ages 233
8.2 Percentage of occurrences of infants' manual activities when familiar and unfamiliar toys were presented, within reach and beyond reach 235
8.3 Relation between the onset of pointing and the proportion of common nouns and frozen phrases in the first 50 words in 10 children 238

Tables

1.1 Comparison of three-stage-evolution of vocal communication in animals and three-stage-development of language in human children *page* 24
2.1 Distribution of two consecutive chucks separated by intervals longer than 0.5 sec by caller 32
2.2 Distribtuion of two consecutive chucks separated by intervals of 0.5 sec or less by caller 33
4.1 Mean values of the three key acoustic variables of infant vocalizations across five pitch contours of following maternal utterance 109
4.2 Mean values of the two key acoustic variables of infant vocalizations across three vowels in preceding maternal utterance 111
4.3 Mean values of the three key acoustic variables of infant vocalizations across five pitch contours of following maternal utterance 114
4.4 Mean values of the two key acoustic variables of infant vocalizations across three vowels in preceding maternal utterance 115
4.5 Abrupt change points determined in the discriminant analyses 120
4.6 Likelihood of infant vocalization in response to echoic and non-echoic maternal responsiveness 123
5.1(a) Distribution of maternal speech/infant's response sequences when mother made only one or two utterances before infant's response 146
5.1(b) Distribution of maternal speech/infant's response sequences when mother made more than two utterances before infant's response 147
5.2 Distribution of three terminal contour variants by ten communicative function categories 150

6.1 Variability of the degree of co-occurrence of laughter with rhythmic leg movement according to variability of syllables included in a laugh 186
6.2 Variability of the degree of co-occurrence of laughter with rhythmic hand movement according to variability of syllable length of a laugh 186
7.1 Average duration, size (by angle degree) and repetition rates for infant-directed and adult-directed signs 196

1 Introduction

This book outlines an approach to the development of expressive and communicative behavior in early infancy until the onset of a single word which is rooted in ethology and dynamic action theory. Here the process of expressive and communicative actions, organized as a complex and cooperative system with other elements of the infant's physiology, behavior and social environments, is elucidated. Overall, humans are provided with a finite set of specific behavior patterns, each of which is probably phylogenetically inherited as a primate species. However, the patterns are uniquely organized during ontogeny and a coordinated structure emerges, which eventually leads us to acquire spoken language. A dynamic model is presented where elements can be assembled for the onset of language in the infant in a more fluid, task-specific manner determined equally by the maturational status and experiences of the infant and by the current context of the action.

No doubt, communication is a social phenomenon and the most prominent feature of human speech and language. The complex organization of human societies is mediated by the ability of members to inform one another and is dependent on the exchange of information. Therefore, not surprisingly, many scientists have focused attention on how children acquire language ability.

Although children do not produce linguistically meaningful sounds or signs until they are approximately one year old, the ability to produce them begins to develop in early infancy, and important developments in the production of language occur throughout the first year of life. Unless they are hearing-impaired, infants acquire phonology during their first year. In spoken language, the acquisition of phonology consists of learning to distinguish and produce the sound patterns of the adult language. At birth, the newborn has the ability to distinguish virtually all sounds used in all languages, at least when the sounds are presented in isolation. The newborn produces no speech sounds, however. During the first year of life, speech-like sounds gradually emerge, beginning with vowel-like coos at six to eight weeks of age, followed by some consonant sounds, then followed by true babbling. By the end of the first year, children are typically babbling sequences of syllables that have the intonation contour of their target languages. Finally, meaningful words are produced; that is, the onset of speech occurs.

The factors that underlie these developments include: physical growth of the vocal apparatus, neurological development, and language experience. Language experience exerts its influence on both the perception and the production of speech sounds. Characteristics of the vocal apparatus that enable us to acquire language, features of neurological development, and features of the manner in which the experience of ambient language influences children's linguistic behavior are all uniquely human, and this uniqueness can only be adequately comprehended when we view the process of early language development from a comparative perspective. Moreover, the predisposition of humans to acquire language is not restricted to a specific modality but rather is somewhat amodal. When humans have difficulty acquiring spoken language, other possibilities can be explored – a further biological predisposition that has phylogenetically evolved exclusively in humans.

A primate behaviorist's view of language acquisition

By comparing human language with the communicative behavior of nonhuman primates, this book will take an ethological perspective in exploring the changes that occur during this earliest stage of language development. Animal societies are equally dependent on the exchange of information. Any organism that lives in complex social groupings must rely on communicating some aspects of its status to others. Such an exchange of information, the process that defines a communication system, implies the existence of a common language or a common set of rules that govern the encoding and decoding of signals in the communication system.

It is tempting to think of animal communication systems as being composed of simple invariant designators or external manifestations of some basic internal states such as hunger, pain or reproductive readiness. For nonhuman primates, however, it is known that, in addition to these states, many other individual and societal factors such as individual identities, kinship, roles, dominance relations and coalitions play an important part in social organization and social behavior. The complexity of many primate societies kindled interest in the communication systems mediating social behavior. For this reason, the objective and quantitative description of vocal communication began earlier in nonhuman primate studies than in studies of human infants.

Carpenter (1934), a pioneering researcher, introduced in his observations of howler monkeys the basic method that is still used – describing vocalizations and the situations in which they were used. Rowell and Hinde (1962) were the first to characterize the vocal repertoire of a monkey, the rhesus macaque, by publishing sound spectrograms. Winter, Ploog and Latta (1966) added a quantitative dimension to the analysis by measuring acoustic features of the sounds recorded in their colony of squirrel monkeys. Struhasaker (1967) statistically analyzed the vocalizations recorded in his field study of vervet monkeys.

As a primate behaviorist, these early pioneering works influenced my initial interest in language. Consequently, my first exposure to the study of language did not involve human infants, children or even adults. In 1979, I was living in the upper Amazonian basin in Bolivia observing groups of a free-ranging New World primate, Goeldi's monkey. While there, I recorded their vocalizations. During my observations, I found that the animals exhibited two different types of responses when group members encountered a predator and emitted an alarm call. One was to climb down to the ground and to freeze there. The other was to climb up to the highest strata in the canopy and to mob. Different types of alarm calls appeared to be associated with different types of predators and the behavioral responses were assumed to vary with call type. However, the sound spectrographic analyses that I conducted upon returning to Japan showed that the entire sample of alarm calls fell along a graded continuum. Therefore, I chose to focus my doctoral thesis on how Goeldi's monkeys perceive conspecific alarm calls. Using captive animals, I investigated their responses to experimentally produced conspecific natural calls as well as to synthesized versions of them that varied in the acoustic parameters that defined the calls under study. Although natural alarm calls showed considerable individual heterogeneity, playbacks of synthesized versions of these calls that varied in a single acoustic parameter produced gross differences in behavioral responding across a narrow acoustic boundary.

With respect to speech perception in humans, if one creates synthetic speech stimuli representing equal steps along the continuum of a single acoustic parameter (for example, voice-onset-time ranging from simultaneous voicing to increasingly delayed voicing) and plays these stimuli to subjects, subjects report the experience of hearing either of two different sounds (for example, /ba/ or /pa/) rather than a graded series of sounds. That is, they perceptually group several different stimuli as /ba/ and certain other stimuli as /pa/. There is no apparent ambiguity between /ba/ and /pa/. A given stimulus from any point on the continuum is labeled as one or the other phoneme, and the two phonemes are strictly categorized; this phenomenon is known as categorical perception. The findings I obtained on vocal perception in Goeldi's monkeys appear analogous to this categorical perception that humans demonstrate with speech sounds, though at present such a perception is thought to be restricted to speech sounds.

After earning my doctorate, I briefly conducted research in Texas, USA. There, I investigated the perception of conspecific alarm calls in a group of Japanese macaques that had been translocated from the Kyoto area of Japan ten years prior. In my work with Japanese macaques, I employed the same experimental paradigm as in my previous work with Goeldi's monkeys. I found that Japanese macaques also perceive their conspecific alarm calls categorically, as demonstrated in human speech perception. From my studies, I learned that what is perceived as a single unit of behavior by human observers (i.e., what is heard as a single class of vocalization, in this case) may not actually be perceived as such by members of other species. These findings, together with similar results with other nonhuman primate species (see Snowdon 1982, for review), were rather astonishing because previous researchers attempting to construct vocal repertoires for nonhuman primate species (e.g. Rowell and Hinde, 1962) have noted the complex call structure of animals that was highly variable both between individuals and within the repertoire of a single individual. That is, many calls could not be easily categorized into discrete classes but rather call structures seemed to intergrade with one another. Researchers have assumed that in many cases these intergradations corresponded to hypothetical underlying motivational continua, thus the intergrading call structure was said to map a continuous motivational system. Despite this sort of variability and complexity, findings like my own suggest that we must be very cautious about how we define units of behavior in nonhuman primates. Based on such reflection, thereafter, primatologists working with vocal communication started to seek new methodologies that could reconcile the continuous variability in calls with the discrete messages they appear to carry. In addition, they successfully expanded the notion of vocal communication in traditional ethology. In so doing, they sought to elucidate the evolutionary continuity between nonhuman primate vocalization and human language.

Implications and limits of the traditional ethological approach to communication

The term "ethology" refers to the biological study of behavior (Tinbergen, 1951). It has been claimed that the discipline of ethology offers a unique integration of a unifying theory, evolutionary biology, with a methodological heritage, naturalistic observation (Blurton-Jones, 1972; Charlesworth, 1980). The operational translation of the evolutionary perspective on to behavior was provided by an early pioneer of ethology, Nicholas Tinbergen. Tinbergen (1951) defined ethology as follows:

> the science [of ethology] is characterized by an observable phenomenon (behavior, or movement), and by a type of approach, a method of study (the biological method). The first means that the starting point of our work has been and remains inductive, for which description of observable phenomena is required. The biological method is characterized by the general scientific method, and in addition by the kind of questions we ask, which are the same throughout Biology and some of which are peculiar to it. (1951: 411)

The modern synthetic theory of evolution provides an integrative framework for many disciplines and content areas. Naturalistic observation provides not only essential descriptive data but it also serves as an invaluable source of ecologically valid hypotheses. Current ethology does not stress biological determinism but rather a multilevel perspective that can expand and enrich our understanding of development. Tinbergen argued that the question, “Why does this animal behave in this way?” included four different questions in the “why.” The first question asks why the animal performed a particular behavior now, the question of immediate causal control of the behavior. The second question asks how the animal grew to respond in that particular way, the question of individual development. The third question asks why this kind of animal does this particular behavior, the question of survival value or function of the behavior. Finally, there is the question of why this group of animals came to solve this problem of survival in this way, the question of evolutionary origins of the behavior.

Until the mid-1980s, virtually all investigators interested in the vocal communication systems of nonhuman primates were concerned with the problem of human language in terms of these four questions. Those engaging in research with nonhuman primates looked for clues to illuminate the evolutionary background and biological heritage of human language. These kinds of clues, hints of the rules by which socially important information is encoded into and decoded from speech sounds, are especially relevant to hypotheses on the origins of human language since there are no fossil records available and one has to rely on comparative studies alone. The uses of vocalizations and their relationship to social behavior may be investigated when both the auditory and social parameters of behavior are available. In fact, in many nonhuman primates, certain features of the social situations in which the sounds are emitted are accessible to the investigator.

The approach to language that I adopt in this book might surprise those who have little knowledge about recent advances in primatology with respect to vocal communication. For example, linguists and developmental psychologists who regard language as a capability beyond the reach of animal research subjects might conclude that primate vocal communication falls outside their own purview as investigators and scholars. Such reactions would not be unexpected given that mainstream modern linguistics has been more concerned with theories of grammar than social communication and ecologically valid models of language use. Further, language has also been defined in very abstract terms and treated by many linguists as though it were synonymous with generative morphology and syntax.

By considering the general characteristics of vocal systems and how they are used, a number of primatologists interested in communicative behavior have recently revived the traditional ethological paradigm in order to place the interspecies comparison of vocal sounds in perspective for nonhuman primates. The conceptual framework for this book is inspired by the theories and methods of this recently expanded ethology as well as by current knowledge about vocal communication in nonhuman primates. The arguments raised and the paradigms developed in recent research also contribute to our understanding of the nature of linguistic capacity and are particularly indispensable to understanding how preverbal human infants acquire language. However, before I explore arguments surrounding language development in human infants, I will outline recent advancements in research on nonhuman primate vocal communication. A focus on such research will help show why evolutionary and comparative perspectives as formulated in the discipline of ethology are crucial to guide a program of developmental research on humans in general. Indeed, this is particularly important in that recent trends in developmental psycholinguistics research cast nonhuman primates in a more interesting light than ever before.

It is now recognized that language, whether spoken or signed, rests on several different types of motor and phonetic learning systems and a range of potentially contributory precursive behaviors (Bullowa, 1979; Papoušek, Jürgens and Papoušek, 1992; Oller, 2000; Speidel and Nelson, 1989). Hence, it is now deemed legitimate to investigate infants' cognitive and neural development as well as their social perceptual experiences in the quest for understanding how and why they begin to speak. Such an approach is also a theoretical necessity. That is, if infants engage in behaviors that facilitate language before they possess the cognitive capability to fully appreciate its existence, then their behaviors must be motivated by one or more non-linguistic factors (Locke and Snow, 1997). Merely owning the genes of a species known to possess the capacity for language would be insufficient. Linguists have argued that language requires specialized mental mechanisms that are encapsulated or dissociated from other, more generalized processing systems. However, linguists have not yet presented actual evidence for this. I propose that an ethological approach to language development provides one possibility for a breakthrough on this issue.

Discrepancy between ethologists' traditional view and linguists' view of human speech

In his formulation, Tinbergen aptly recognized that a full understanding of behavior includes both proximate and distal "causes" and that one must always view individual animals within the ecological context of the species. In sharing this view, my purpose in this book is in part to illustrate how Tinbergen's formulation can be used to direct research on a class of common, but puzzling infant behavior: language acquisition. That a combination of evolutionary biology and naturalistic observation potentially has much to offer our understanding of human behavior has been pointed out a number of times over the past few decades. However, Tinbergen's formulation has only been successfully extended to human behavior, more specifically, human language, in just a few investigations. As partial explanation for this, Tinbergen also cautioned that one should not confuse questions asked at one level with those asked at another. For example, Blurton-Jones (1972) argued that the persistence of unproductive nature–nurture arguments in behavioral research is a consequence of the confusion between issues of development and those of adaptation and evolution. More importantly with respect to communicative behavior, it must be acknowledged that ethologists have not understood how linguists distinguish human language from nonhuman communicative behavior on the one hand, and that linguists have not understood the significance of the ethologists' view of language on the other.

Traditional ethology conceived of animal communication as genetically fixed, developmentally immutable, stereotyped activity. Within the communicative repertoire of a species there were thought to be only a relatively small number of invariant signals (Moynihan, 1970) that were used in an equally small number of motivational or contextual situations (Smith, 1977). Although the critical importance of context in the interpretation of signals has been recognized for many years, the prevailing view that has been provided of communication in nonhuman animals has been of a restricted signal repertoire and a restricted set of communicative referents.

According to the traditional ethological view, which assumes discontinuity between human and animal communication, human communication is not stereotyped and is considerably modifiable during development. Human communication employs a signal repertoire of enormous size compared with the repertoires of nonhuman species. Human communication has signal invariants that are easily perceived by human recipients even though it is often difficult for humans to discern the physical structure of signals. If one ascribes to this view, one cannot analyze human communication from an ethological perspective. Earlier studies of sounds produced by nonhuman animals (other than primates) also confirmed that these sounds could be regarded as a sort of fixed action pattern. Before sound spectrum analysis became possible in the 1950s, all sounds were identified by labels that were often idiosyncratic to the person who used them. With the new method, different individuals were now able to agree on the pattern of a signal based on its objective and permanent representation. Pioneering sound spectrographic analyses revealed that many of the vocalizations recorded from a number of bird species could be easily discriminated from one another. However, as noted by Rowell and Hinde (1962), nonhuman primate vocalizations frequently appeared to intergrade with one another and hence were not clearly classifiable into discrete categories.

Therefore, in the ethological view, nonhuman primate vocalizations should be classified into the category of human communication because of their signal feature of forming a graded continuum. However, the ethologists were so naive regarding linguistics in general that they failed to appreciate that the human system does not necessarily use continuous units exclusively. On the contrary, although language employs continuous parameters whereby small changes in acoustic value result in corresponding changes in transmission value (e.g., as one raises one's voice gradually, one may sound increasingly angry or upset), such continuous variations merely correspond to "paralinguistic" signaling. They are not regarded by linguists as playing a role in differentiating lexical items. Linguists concluded that while nonhuman primate vocal communication systems appear in some cases to rely heavily on signal dimensions that vary continuously for communicative value, human vocal communication systems maintain a fundamental distinction between dimensions that are manipulated continuously for paralinguistic effect and segmental features. Moreover, in linguistics, the latter are treated as phonetic units and are interpreted categorically in terms of their lexical effect.

A typical expression of this sort of linguistic view of nonhuman primate vocalizations and human language is Hockett's (1960) characterization of human language, as a communication system, in terms of "design features" (e.g., "discreteness" and "duality of patterning"). According to Hockett, the human system possesses discreteness in that the alphabet-level (segmental phonetic) units have categorical values. That is, a change in the acoustic characteristics of one sound segment (say the b in "bay") is regarded as irrelevant from the standpoint of transmission value (meaning) unless it precedes a shift to a new meaning category (say "pay"). Human language usually includes lexicons of thousands of words constructed from such discrete alphabetic/phonetic units. Nonhuman vocal communication systems often include an inventory of discrete calls or call types (e.g., one for threat, one for affinity, one for alarm). However, their categorical lexicon is usually small in number of meaningful units by comparison with human languages, and importantly, as already noted, it is usually characterized by stereotypy.

The power of the human system to create an extensive lexicon lies in its dependency on the duality of patterning referred to by Hockett. According to Hockett, duality of patterning concerns individual alphabetic units of the human phonetic/phonemic system that are independent of meaning; duality of patterning refers to the fact that these units can be recombined and reordered to construct different units of meaning. Thus the words act, cat and tac(k) all share the same phonemic units while lexically they are entirely distinct.

It is important to emphasize the "recombinability/reorderbility" characteristic implied by this duality because recombinability enables a small number of phonemic units to be utilized to create an enormous lexicon, by merely stringing the phonemic units in unique patterns. With respect to potential recombinability, studies of nonhuman primate vocal systems appear to show either that no restructuring is possible or that changes are far more limited than those that can occur in human speech. A system that has no recombinability is restricted to a lexical inventory size, which can be no greater than the number of discrete units in the system.

Thus, the use of continuous variations of sounds for communicative purposes that has been recognized in nonhuman species is indeed shared by humans, but in humans continuous variation is only used as a paralinguistic component of vocal communication and not as a component of language itself. Humans also apparently differ from nonhuman primates in making greater use of the categorical features of sound in their vocal communication. Linguists have assumed that through the acquisition of such distinct means, humans exclusively are equipped to produce and use language. The evolution of language is thought to have occurred some time after the emergence of vocal communication like that found in living nonhuman primates, for instance after the acquisition of a unique vocal apparatus as bipedal walkers. In order to produce sounds with the features needed for language, sounds generated by the air stream must be morphologically chopped by vibrating vocal folds.

Methodological characteristics of ethology in investigating nonhuman primate vocalizations

Hockett initially proposed his model in order to criticize naive comparisons between nonhuman sounds and human language. However, having rejected the position of traditional ethologists, one might revisit the original question: how are nonhuman sounds similar to or different from the sounds of human language? Hockett's model provides a framework for discussing only how the sounds "function" (similarly or differently in humans and nonhuman species) but it does not really address the issue of the relationship between human and nonhuman sounds per se. In order to investigate how preverbal infants come to produce sounds that characterize human language, a purely acoustic description of preverbal infant vocalizations could still be meaningful. In this regard, findings obtained from comparisons between the vocal sounds of humans and nonhuman primates could offer an important perspective.

Further, mostly owing to our ever-developing knowledge of human speech perception, the distinction between discrete and continuous vocalizations has blurred recently. Knowledge concerning human speech perception came first from findings on categorical perception, a topic in which I was interested in my doctoral work. Namely, several of our speech sounds appear to form a continuous distribution when examined spectrographically and yet we rarely have difficulty distinguishing the category into which a particular sound falls. Findings such as these make it difficult to apply the graded-discrete distinction between the signals of primates (humans included) versus the signals of other animals as was done in the earliest nonhuman primate vocalization studies. Whether a repertoire appears large or small depends on how one characterizes signals and how one deals with graded signals. Along with improvements in the detection of signals, early estimates of repertoire size have been altered; while obviously valuable in itself, this has made it even more difficult to draw any conclusions about repertoire size.

In response to the oversimplified dichotomy between animal and human communication, primate behaviorists have sought methods to identify more precisely each call type within a vocal repertoire. As a result, advancements have been made in the techniques used to analyze vocalizations. These advancements fall primarily into three domains that I will discuss presently: (1) contextual analysis, (2) sorting techniques, and (3) playback techniques.

Contextual analysis

First, there came to be much more detailed analysis of the contexts in which calls occurred than in previous investigations. For example, in his study of Japanese macaques, Green (1975) found that one call type, the coo call, actually consisted of several variants, each of which was associated with a different behavioral situation. In classical studies of primate vocalizations (e.g., Rowell and Hinde, 1962) data comprised a few representative sound spectrograms on the graded nature of calls. Actual isolation of discrete vocalizations based on physical characteristics was difficult because of this variability and because this variability was interpreted as representative of a behavioral continuum of arousal or motivation. In his study, Green therefore isolated additional sources of variability in the vocalizations of Japanese macaques. He sorted spectrograms into categories of similar appearing acoustic patterns and found that these categories represented vocalizations uttered in similar social contexts. Social contexts were differentiated by various factors such as age, biological state (e.g., “estrous female”) and dominance relationships. His success in grouping calls according to their acoustic characteristics, which could then be correlated with social context, provided further support for the argument that vocalization variability is a function of behavioral categories.

Subsequently, for a number of vocalizations in other primate species that had been classified as single types, other researchers have found that an apparently unitary call type can further be divided into several variants (e.g., pygmy marmoset trills, Snowdon and Pola, 1978; cotton-top tamarin chirps and long calls, Cleveland and Snowdon, 1982). My own findings with Goeldi’s monkey alarm calls provide another example. Examining the correlation between different structural variants and different behaviors is one way of discriminating call types.

Sorting techniques

The second important area of methodological advancement concerns the development of sorting techniques. Indeed, in order to find variant and functionally meaningful forms within a category formerly classified as unitary, it is necessary to establish reliable sorting techniques for the sounds.

In general, analyses of animal vocalizations have largely been dependent upon sound spectrographs that provide a visual analog of vocalizations. At the onset of the analysis, an investigator would typically attend to the set of characteristics that appear most salient, at least to him or to her. Then, the investigator would proceed to sort the spectrograms into categories based on those particular structural differences. Thus, by using different criteria for sorting, different investigators could conceivably generate different numbers of vocal signals. For example, it is possible that one investigator might sort only according to obvious differences in call type whereas another may sort according to minor or subtle differences in call structure. The repercussions of this difference between investigators can be quite striking.

As such, I will discuss two important points concerning sorting techniques. First, while most of the early studies of primate vocalizations employed only quantitative sorting techniques, qualitative techniques (e.g., visually inspecting sound spectrographs and anecdotally classifying them into several types) have been used to supplement the quantitative techniques (Smith, Newman and Symmes, 1982). When examining subcategories or variants within a larger category, quantitative statistical techniques can be useful. Quantitative techniques can help identify the acoustic parameters that differentiate call variants and address questions concerning whether or not the differences the investigator has perceived in sorting are indeed valid. If no quantitative basis is available to support a finer division of call types, then over-classification is likely to occur. Overlooking true call variants that distinguish different populations is also likely to occur. Second, the use of quantitative techniques enables one to distinguish the parameters that determine differences in the content of the call from the parameters that identify the individuals or population making the call.

Early research pioneers used a calibrated graticule to obtain measures of the temporal and frequency parameters for each spectrogram (Snowdon, 1982). Later, owing to technological advances, it became possible to enter natural vocalizations directly through its analogue-to-digital converter and sample them. The sequentially digitized samples are stored in sequential order there. Discrete Fourier Transforms are obtained by applying a Fast Fourier Transform to the digitized representations of vocalizations. Then, using call context as one variable and individual animals or other factors as other variables, resulting digitized values can be subjected to univariate or multivariate analysis of variance or discriminant analysis. Through this method it is possible to determine which parameters, if any, vary depending on the behavioral situation and which, if any, vary between individual animals. However, the fact that statistical techniques may occasionally reveal a degree of complexity in vocal structure which qualitative sorting ignored does not automatically mean that this greater acoustic complexity is functionally significant in the communication system of the animals – that must be determined empirically.

The determination requires three steps. First, one must determine that the vocal structure shows a significant association with the behavioral and social context in which a call is produced (the process described above as “contextual analysis”). Second, one must form hypotheses about call function. Finally, most importantly, one must test whether or not the statistically significant associations between variants of calls that are classified into a single category and the contexts in which they are recorded really imply that the associations are biologically significant. In order to verify the associations, it must be demonstrated that conspecific animals really perceive the variants differently. That said, I have now touched on an important issue that research primatologists have fervently sought to address and I will elaborate on it further in the section to follow.

Playback technique

Primatologists have attempted to use playbacks of calls in appropriate and inappropriate contexts to solve the problem of contextual relevance. Ideally, the identification of a certain number (e.g., five) of vocal structure categories or subcategories will be accompanied by the identification of the same number (e.g., five) of types of social-behavioral situations in which calls are uttered, with a perfect association of one given call structure with one social-behavioral situation. But the normal state of affairs is far from ideal; one call type may appear in more than one situation and one situation may be associated with more than one call type (Gouzoules, Gouzoules and Marler, 1984). In order to verify that conspecifics really perceive the variants differently, it is crucial to demonstrate empirically that they respond differently to the variants. For this purpose, auditory playback experiments have been developed. Snowdon (1982) summarizes the theoretical implications of this experimental paradigm as follows:

> One hypothesizes which behavior should occur following a call in situation A and which behavior should occur following a call in situation B. One can establish that call X is given functionally in situation A by playing back call X in situations A and B. Animals should give their normal responses to call X only in situation A and not in situation B if call X is most closely associated with situation A. Playback of call X in situation A should also be more effective at eliciting appropriate behavior than playback of call Y. (p. 215)

Snowdon further pointed out several advantages of the playback technique. The playback technique offers experimental data to resolve questions that could be answered at best roughly by the use of correlational techniques. One can conduct experiments both in captivity (i.e., in a closed captive population that preserves the normal social structure found in nature) and in free-ranging situations (i.e., in a population of animals in their natural habitats). In captivity, behavioral responses occurring with regularity to natural emissions of the stimulus are defined as criterion responses to the experimental playback of the signals. The advantages of this sort of experiment stem from the fact that natural social groups are investigated, thus typical responses to signals are likely to occur within normal social contexts. Moreover, one can identify individual animals and record their individual responses with high precision.

However, the playback technique does have some limitations. First, it is often difficult to get animals to respond repeatedly to a playback stimulus. For example, for effective simulation, the speakers must be well concealed in parts of the environment where other members of the social group are likely to be. Further, stimulus presentation must be kept within relatively low frequencies to avoid habituation to the stimuli. Second, it is possible that the behavioral response one observes might not be an accurate reflection of what the animals can discriminate. There might be signals that the animal can differentiate but which do not always lead to differential spontaneous overt behavioral responses. Therefore, the method might produce a bias toward finding categorization of stimuli rather than differentiation of stimuli. One might discover that sounds, which are easily discriminated in an operant conditioning situation, may not normally lead to different “natural” behavioral responses by the animals and so would not appear to be discriminable using the playback technique.

Nonetheless, for determining how animals naturally respond to signals in their normal environments, it is an extremely powerful method. Indeed, the best evidence for discrimination provided by the methods of studying animal perception that were traditionally used in experimental psychology (i.e., discriminative conditioning and the habituation-dishabituation paradigm), may not be the most appropriate data for understanding natural processes in animals because the traditional methods present the sounds to animals outside of the normal context in which these sounds would normally be produced. Therefore, researchers have employed playbacks in free-ranging situations in order to compensate for the disadvantages of the playback experiment in captivity. Obviously, when executed, this paradigm is more difficult than other methods employed in perception studies. Experimenters must make certain that all animals are within audible range of the playback stimulus. As with playbacks used with captive populations, it is necessary to camouflage playback equipment so that animals will respond to playbacks normally. Despite the technical difficulties, this paradigm has the greatest ecological validity in that everything is natural except for the stimuli being played back. For this reason, several primate investigators have made tremendous efforts to conduct field playback experiments and with a great deal of success.

Methodological advantages of ethology in investigating human language development

Only by use of the playback method is it possible to determine the biologically relevant acoustical components comprising a vocalization in nonhuman animals. As argued by Snowdon (1982), in other paradigms adopted in traditional experimental psychology, nonhuman animals are more likely to base discrimination simply on "acoustic" features which are ecologically irrelevant for themselves rather than on so-called "phonetic-like" features of the stimuli (described in detail below), because of the lack of a normal context and because of the parameters typically used for stimulus presentation. He proposed that "most operant studies with animals will produce evidence for a failure to categorize stimuli. That is, most stimuli will be more or less equally discriminable as the discrimination is likely to have been based on 'acoustic' features" (p. 412).

Just as Snowdon referred to biologically relevant features of nonhuman vocalizations as phonetic-like, an analogous phenomenon has been reported in the speech perception of humans. For example, researchers have reported several experiments that challenge the notion that categorical perception is something special and speech-specific. Pisoni (1977) has shown that categorization by voice-onset-time occurs with pure tones; thus phonemes of speech are not necessary for categorical perception. It has also been shown that categorical discrimination can be made continuous if subjects are given the appropriate set (e.g., by sequencing the presentation order of stimuli as they would occur on the relevant acoustic continuum). Moreover, continuous perception appears by minimizing the memory load during the discrimination task; that is, when subjects are simply asked to detect a change in stimuli rather than whether the last sound heard is more similar to a first or a second comparison sound (Pisoni and Lazarus, 1974). These findings imply that speech perception is similar to perceptual categorization in other modalities. Namely, there is a labeling function that is categorical and a discrimination function that may be categorical or continuous depending on the method of stimulus presentation and the demands of the experimentation task.

In contrast to other areas of developmental psychology, the human language research literature pays little attention to the preverbal period. This is partly because conventional units of linguistic analysis are not useful for such study. Indeed, the conventional units may not always be defensible or optimal, even though much of the current understanding of speech and language development is based on conventional linguistic units such as words, syllables, phonemes and phonetic features. Consequently, the terminology used to describe preverbal vocalizations has varied between researchers. Van der Stelt and Koopmans-van Beinum (1986) called it the "descriptive chaos in studies on infant sound production" (p. 140). Although various problems have created this confusion, it must be true that linguists confronting the subject of infant vocalizations are daunted by its inherently chaotic characteristics: babies produce many different sounds. As long as they depend upon the conventional units of behavioral analysis, scientists may resist examining these chaotic sounds. Such resistance may have contributed to the long-held but false view that babbling is a phenomenon entirely unrelated to speech, the latter being dominated by an innate linguistic capacity that only comes into play at the point when real words are acquired (Jakobson, 1941).

However, the descriptive chaos in studies on infant sound production cannot be overcome by simply adopting systematic linguistic or prelinguistic categories. For instance, as pointed out by Delack (1976), a universally accepted definition of babbling was obvious but descriptions changed with the discipline of the researcher because the processes underlying the development of sound production are not simple. Nevertheless, the adult listener is able to recognize haphazard linguistic categories. Delack (1976) suggests "an early link between perception and sound production which implies a continuity of sound-meaning correlation has been overlooked by many investigators" (p. 494). This is precisely the view adopted by primatologists struggling with the problem of distinguishing "phonetic-like" features from "acoustic" features of vocalizations.

A primary advantage of the primatological approach to early language development is that it can help clarify the appropriate units of behavioral analysis. At least, it makes us keenly attentive to the assumptions that underlie the choice of such units. This is not surprising, not only for primate behaviorists but also for ethologists as a whole. The perspectives presented by primate behaviorists are shared with ethologists in general because both usually undertake their investigations of behaviors in nonhuman species by conducting longitudinal observations of the target behaviors under naturalistic circumstances. It is a commitment to natural history as a starting point for behavioral studies that is a hallmark of the ethological approach. The assembly of the "ethogram" (or catalogue of behavior in its natural context) is essential for any comparative study and also provides the basis for further ecological or causal analyses (Eibl-Eibesfeldt, 1970).

Advantages of an evolutionary view on language

Another advantage of the ethological approach is that it offers a useful evolutionary perspective on emerging behavior. As will be argued later in more detail, with respect to the vocal tract, it is inappropriate to think of the infant mechanism as being simply a scaled-down version of the adult structure. Although data on vocal tract development remain sparse, existing data suggest that the infant's vocal tract differs substantially from that of the adult. The infant's vocal tract is not only shorter (thereby accounting for absolute differences in dimensions and configurations) but it also differs in the relative or proportionate size of its subdivisions. These anatomical differences may impose restrictions on the degree to which the units and dimensions of adult language can be used to characterize infant vocalizations.

As the universalist theories of Jakobson (1941), Smith (1977) and Stampe (1973) fall under criticism, the redefinition of data, concepts and issues in infant language acquisition is well underway. The major criticisms of these theories (Locke, 1993; Oller, 2000) are as follows: empirical evidence violates the predicted orders of acquisition; language acquisition is not adequately explained by a process of successive acquisition of phonetic oppositions; and cognitive development is ignored. Further, these theories generally predict a gradual convergence on, or progression toward, an adult system, when in fact phonology acquisition is characterized both by regression and by overgeneralization. These theorists apparently view the development of an organism as a continuous process of adaptation to the environment in which it lives.

Charles Darwin was perhaps the first and most noteworthy scientist to adopt this adaptation-to-the-environment view of the processes underlying language acquisition in human infants. Of course, Darwin is often credited with establishing the scientific approach not only to language acquisition but also to the entire discipline of developmental psychology. Although his major interests were in the theory of evolution, he could also be considered the first developmental psychologist. Indeed, in 1877 he published a short paper, "A Biographical Sketch of an Infant", describing the development of his infant son, Doddy. In the paper, he reported that his son "understood intonation and gestures" before he was a year old, whereas his linguistic competence was still very limited. In addition, he was impressed by the playfulness of his son and by his capacity for emotional expression.

In his studies of his own infant son, Darwin particularly sought to understand the evolution of innate forms of human communication. Underlying his studies is the notion that one can best understand development as the progressive adaptation of the child to the environment. Thus, this very commonly held notion could now be traced directly to Darwin and the influence of evolutionary theory. The introduction of systematic and objective methods to the study of development, another of Darwin's contributions, also must not be overlooked. His studies of development were always undertaken on the basis of actual observation of developing children, and the major biological foundations of behavioral development were virtually laid after the publication of Darwin's study. Nevertheless, many studies conducted in the field of psychology after Darwin have remained philosophical or anecdotal despite his work. Objective and quantitative investigations are crucial to understanding behavioral development in nonhuman animals because they never talk. This perspective has spawned the development of a scientifically driven discipline to study the vocalizations of nonhuman animals on a "purely" acoustical level. Some scientists working in the field of nonhuman animal vocalizations, myself included, have therefore come to develop interests in the vocalizations of preverbal infants.

In a larger sense, pragmatic and interactive data combined with a comparative, evolutionary perspective on language in relation to human biological adaptation complete an ethological paradigm that emphasizes natural selection. The evolutionist paradigm, such as the one initiated by Darwin, regards infant language acquisition as a particularly human progressive adaptation of brain and behavior to varied and distinct environments. An infant is thought to develop from exhibiting a particular range of expressive resources in spontaneous exploratory productions to exploiting those expressive resources on the basis of experience in gaining control over expressive skills in language. Such a view is also in harmony with the paradigmatic notion that human language evolved as a tool that helped generalist hominids to survive and exploit diversity in the physical and social environment. Diversity is potentially life threatening to species that are too behaviorally and ecologically specialized to learn and to adapt to new environmental habits and niches. Developmentally, the cognitively based theory of language acquisition characterizes the infant as an active seeker and user of information. The infant actively solicits linguistic information and tests and revises hunches.

Interestingly, recent comparative studies (e.g., Hauser, 1996) suggest the existence of cognitive parallels in the development of human and nonhuman primates. The capabilities we share with nonhuman primates are the first capabilities to develop in human infants. Here the significance of ethological approaches to human language acquisition becomes clear. For example, much of nonhuman primate socialization revolves around affective signaling by voice and by face and this also applies to the interactions of human infants. Both groups of primates (i.e., human and nonhuman) retain their capacity to communicate on that level. Humans additionally take on more arbitrary and codified means for communicating. Therefore, investigations of interactions between infants and people who are talking should be important for understanding language development because language develops in a social context. Physical cues, particularly the vocal variations of caregivers, help define infant social development. As infants orient to the cues, these cues may start them down a developmental growth path that leads to language acquisition. If this scenario is true, the question of how infants acquire language becomes a relevant ethological issue and this book presents evidence for this very proposal.

Structure and function of nonhuman primate vocalizations

As previously argued, findings emanating from new analytic techniques for nonhuman primate research have led us to re-evaluate the traditional view of human/nonhuman differences in vocal ability. Snowdon (1982) summarizes four major functional levels of variability in vocalizations throughout the primate order: (1) individual variability, (2) population variability, (3) localization variability, and (3) “phonetic-like” variability.

Individual variability

It has been known for some time that substantial individual variability exists in the form of nonhuman primate calls in several species (Rowell and Hinde, 1962; Marler and Hobbette, 1975). However, until recently it has not been clear that conspecifics really perceive and make use of the variation. Playback studies reveal this to be the case. Japanese macaque mothers show selective responses to playbacks of recorded vocalizations of their offspring and differences in the calls of mothers are responded to selectively by their infants, even those younger than one month old (Perreira, 1986; Masataka, 1985). The same phenomena were confirmed in rhesus macaques (Hansen, 1976) and in squirrel monkeys (Kaplan, Winship-Ball and Sim, 1978). In a field experiment, Cheney and Seyfarth (1980) played “lost” calls of infants to groups of mothers, all of whose infants were out of sight, and found that mothers responded selectively to calls of their own infants. Interestingly, other mothers appeared quite aware to whom the infants were related, as evidenced by the observation that recorded calls of other infants caused them to look at the infant’s mother. A playback experiment using chimpanzee pant-hoot calls and control sounds showed that chimpanzees discriminate between calls of familiar and strange animals as well as between male and female pant-hoots (Bauer and Philip, 1983). Finally, in their study of pygmy marmoset contact calls, Snowdon and Cleveland (1980) found individually distinctive acoustic features in the calls that elicited differential individual responses upon playback.

Population variability

It is common to find dialects or geographical variations in acoustic patterns of vocalizations between different populations, in birds and humans. Also, for nonhuman primates there is a growing literature documentating population differences in vocal structure. Maeda and Masataka (1987) undertook a quantitative acoustic analysis of long calls of red-chested moustached tamarins from the primary forest of northwestern Bolivia and found that their acoustic structure varied between populations. Because there was no evidence to suggest underlying genetic differences between the populations, the authors concluded that vocal variability was comparable to dialects. Subsequently, a playback experiment revealed that animals really perceive the difference between the acoustic quality of long calls recorded from their natal populations and those recorded from alien populations (Masataka, 1988). In terms of long calls, researchers have confirmed similar findings in chimpanzees and Japanese macaques (Masataka and Fujita, 1989; Kajikawa and Hasegawa, 2000).

Essentially, animals perform these vocalizations most frequently when they are separated during travel from conspecifics living in the same groups. Antiphonal calling takes place mostly among affiliative individuals. The vocalizations of individuals living in alien groups, in general, provoke vigorous avoidance responses in hearing animals. Such behavior would serve functionally to encourage intragroup cohesion and intergroup spacing.

Localization variability

This form of variability concerns changes in call structure that occur in association with the distance of callers from other animals in their groups. Several studies have examined the design features of vocalizations, namely, those acoustic features of sounds that maximize or minimize detectability in a given environment (Waser and Waser, 1977; Wiley and Richards, 1978). Frequency modulation is a very important acoustic cue for sound localization (Brown, Beecher, Moody and Stebbins, 1979). For instance, Pola and Snowdon (1975) found that pygmy marmosets used three trill variants that were physically different from one another and yet appeared to convey identical behavioral messages. These trill variants could be ordered according to their cues for sound localization. In a subsequent field study in Peru, Snowdon and Hodun (1981) reported that the most localizable trill variant was heard most frequently when calling and responding animals were far apart, whereas the least localizable variant was used by animals in close proximity. A similar variation in call structure depending on distance between animals has been observed in the calls of captive cotton-top tamarins (Cleveland and Snowdon, 1982).

Masataka and Symmes (1986) recorded isolation calls of captive squirrel monkeys by separating infants from their natal group members and then permitting vocal contact between the “lost” baby and the group at systematically varied distances. Separated infants gave longer calls at greater separation distances from their natal group members and responding adults and juveniles similarly extended the length of their vocalizations. In the longer variants, a high-frequency element was prolonged. At first this appears rather disadvantageous for long-distance sound transmission. However, in the habitat occupied by squirrel monkeys, insect noise in the range 5–8 kHz operates to mask some portions of vocalizations and to produce the unlikely result that higher frequencies are better distance signals. The longer variants enjoy the advantage of relatively clear acoustic channels in noisy environments – an example of the net advantage of a “frequency window” in ambient environmental noise.

“Phonetic-like” variability

With respect to the three types of variability described thus far, the results described are arguably not very surprising. Intelligent, long-lived primates, who spend most of their lives in close proximity to relatives and fellow group members, would readily learn to associate individual vocal characteristics with other attributes of social and environmental relevance. This is similar to our attention to the paralinguistic elements of human speech. More interesting, then, would be the discovery of “phonetic-like” elements in nonhuman primate vocalizations. A series of playback experiments seriously challenged the long-held view that nonhuman primates communicate primarily about internal states and that they communicate relatively little, if anything, about external objects or events.

Since Struhsaker’s (1967) early fieldwork in the mid-1960s, primatologists concerned with the origin of language have had an interest in the calls of vervet monkeys. This is because vervet monkeys give acoustically different alarm calls to at least three types of predators (to large mammalian carnivores like leopards; to eagles; and to snakes such as pythons). Further, each call type is associated with an adaptively appropriate escape response; for example, when on the ground, leopard calls lead the monkeys to seek refuge in the trees whereas snake calls lead them to search the ground. In their playback experiment with free-ranging animals, Seyfarth, Cheney and Marler (1980) played recorded alarm calls (in the absence of actual predators) and filmed the monkeys’ responses to the calls. They found that subjects looked in the direction of the concealed loudspeaker and responded to each type of call with an appropriate escape response. Analysis of the filmed material also revealed that individuals responded largely independently of one another. Moreover, alarm call specific responses were elicited regardless of the sender’s or responder’s age or sex and response type was not affected by manipulation of the length or the amplitude of the playback calls. Thus, the conclusion that some vervet monkey calls have semantic qualities concerning external objects or events seems difficult to escape.

This sort of “representational”-like behavior has been reported in other nonhuman primate species including gibbons (Tenaza and Tilson, 1977), ringtail and ruffed lemurs (Macedonia, 1990), and Goeldi’s monkeys (Masataka, 1983a). As already noted, in free-ranging Goeldi’s monkeys, freezing and emission of warning calls are recognized as the two most consistent types of responses to two different, naturally occurring types of alarm calls. Playback of synthesized versions of these calls varying in frequency range actually produced differential behavioral responding with a slight change of the acoustic parameter, suggesting an underlying perceptual boundary. Similar findings have also been reported for macaque vocalizations and for pygmy marmoset calls that are thought to represent external objects as well. The phenomena appear closely analogous to the manner in which humans perceive speech. Taken together with findings on the semantic quality of calls that are perceived categorically, these findings reinforce the emerging view that at least some monkey vocalizations possess "phonetic-like" features.

Admittedly, the four sources of variability discussed here do not represent all possible sources of variability within monkey calls. However, the findings obtained so far suggest that all or some of the sources of variability between calls can usually be identified in any call type and that when identified, the calls are separated from one another acoustically. For example, if cues (phonetic-like, individual, populational and localization) are perceived in a call by humans, they are also perceived by monkeys based on independent acoustical features that correspond to each cue. This is analogous to the acoustic relations between phonetic features and paralinguistic properties of speech sounds.

The ethological perspective on the evolution of vocal communication in primates

Scientists working with sounds or vocalizations of nonhuman species point out that the phonatory apparatus, as it evolved toward its human form, is paralleled not only by an increase in the vocal repertoire but also by an overall increase in voluntary control over vocal or sound production. At the simplest level of vocal communication a subject reacts innately to a specific stimulus with a specific call. In classical ethological terms, this could be called a vocal "fixed action pattern," evoked by an innate releasing mechanism. Under such circumstances, neither the vocalization, which represents a genetically preprogrammed motor pattern, nor the eliciting stimulus, which elicits vocalization without any prior experience, has to be learned. At this level of vocal communication, voluntary control is hardly recognized. In cases where voluntary control is completely wanting, vocalizations from the vocal repertoire correspond to elicited reactions, comparable to isolation calls in response to separation from conspecifics. Nonetheless, individual variability in vocalizations is apparent even in the absence of voluntary control, due to individuality in the morphological characteristics of the vocal apparatus.

In cases where responses are elicited involuntarily, animals would exhibit antiphonal calling immediately on hearing a particular call. However, in cases where responsive vocal production is somehow under voluntary control, as described next, perception of individuality in the hearing call, if it could develop, would allow the call receiver to take action on a more socially relevant basis. Moreover, if vocalizations uttered by animals, as a consequence of heightened arousal caused by being separated from other group members by great distances, are modified involuntarily so that acoustic properties that could serve to localize the sound are exaggerated, the evolution of the voluntary modification could be facilitated also.

The next more complex level of communication involves the situation where a subject reacts with a genetically preprogrammed vocal motor pattern but the eliciting stimulus is learned. In other words, the subject has to learn the appropriate context in which to perform a particular vocal utterance that, until then, had been used more or less indiscriminately. Most of the monkey calls and a number of the nonverbal emotional vocal utterances of humans seem to belong to this category. The alarm calls of vervet monkeys and squirrel monkeys are among the most intensively investigated examples of this sort. With respect to the alarm calls of vervet monkeys, their acoustic quality has been found to be genetically preprogrammed. However, Seyfarth and Cheney (1986) have shown evidence of observational learning and social reinforcement in the comprehension and usage of the three types of calls. Juvenile animals give the calls in response to a variety of objects. For example, they might emit eagle alarm calls to starlings, to hawks, to falling leaves, etc. On the other hand, adults make alarm calls only to martial eagles, their only aerial predator, and infants often wait to call until after an adult has given an alarm.

Ring-tailed lemurs can learn to respond to the alarm calls of other species. Oda and Masataka (1996) have shown a gradual development of infant lemur responsiveness to alarm calls of sympatrically living sifakas, another species of prosimian. Infants living in groups with considerable exposure to sifaka alarm calls respond to the sifaka calls whereas infants living in groups with no exposure to the sifaka do not. The study demonstrated that animals living in groups with considerable exposure to these alarm calls actually comprehend the meaning of the calls. The development of the categorical perception of calls also requires experience with hearing the sounds. Japanese macaques with no exposure to alarm calls only perceive their variations continuously while those with abundant exposure to alarms perceive the same variations in a categorical manner (Masataka, 1983b).

For some mammalian species, such as the cat, the dog, the sea lion, the dolphin and several species of primates, it has been experimentally demonstrated in captivity that they can be trained to master vocal conditioning tasks. That is, they can learn to emit a species-specific vocalization for a food reward when a conditioned stimulus is presented (and to refrain from vocalizing during presentation of a different stimulus). Such species clearly have some voluntary control over vocalization. This control, however, is limited to the initiation and suppression of vocalization; it does not extend to the acoustic structure, which is still genetically determined.

The third and most complex level of vocal communication involves learned vocal motor patterns uttered in response to learned stimuli. In this case, there is not only voluntary control over the initiation and suppression of an utterance, but there is also voluntary control over the acoustic structure of the utterance. The possibility of population variability emerging in vocalizations arises. This level of communication is the common communicatory mode of humans. Among nonhuman primates, Japanese and rhesus macaque coo calls are typical examples of this level; Masataka and Fujita (1989) found learning of allospecific vocalizations in cross-fostered monkeys. In the study, one Japanese macaque was cross-fostered by rhesus parents and two rhesus macaques were cross-fostered by Japanese monkey parents. The cross-fostered monkeys imitated food calls of their foster parents. Other monkeys tested in a playback paradigm responded to the calls of their cross-fostered conspecifics as they would to the calls of the foster species. Moreover, the brain structures involved were found to differ depending on the levels at which vocal communication took place. Producing learned vocal motor patterns requires a number of brain structures that are not necessary to produce innate vocal utterances. The capacity to voluntarily initiate or suppress vocalization depends upon brain structures that are not required for the production of unconditioned vocal reactions. In parallel with the hierarchy of levels of complexity in vocal communication, there is a hierarchy of brain structures underlying the different levels of vocal communication.

However, species' brain volumes have been shown to positively correlate with the size of the group in which the species lives, at least among nonhuman primates (Dunbar and Bever, 1998). Based upon this finding, a hypothetical scenario has been presented concerning the evolution of human language. According to Dunbar's argument, language has become progressively more complex in tandem with the increasingly pressing demands of larger group sizes. In Old World monkeys and apes, particularly, contact calling functions as a kind of grooming-at-a-distance. As time-budgets became increasingly squeezed, the animals might have kept up a steady flow of vocal chatter. Eventually, the content in these communications would have been irrelevant: rather along the lines of those formulaic greetings so common in our own conversations.

As group sizes began to drift upward, beyond the sizes to which living species of monkeys and apes are currently limited, vocal grooming began increasingly to supplement physical grooming. Dunbar and Bever (1998) hypothesized that this process would have begun around two million years ago with the appearance of *Homo erectus*. Increasing emphasis was being placed on vocal as opposed to physical grooming for group cohesion. Eventually, even this form of communication would have exhausted its capacity to maintain group cohesion. A more efficient mechanism for bonding might be required to allow group size

Table 1.1 *Comparison of three-stage-evolution of vocal communication in animals and three-stage-development of language in human children*

		Example	
Stage	Key characteristics	Animal	Human child
1st	Genetically preprogrammed pattern of vocalizations are elicited as a response to a specific stimulus	Isolation calls are uttered when separated from conspecifics. When lost from other group members, free-ranging lemurs are predisposed to utter this type of vocalization, whose acoustic pattern is invariate and species-specific	Due to the increase of “arousal” or “level of excitement”, species-specific pattern of vocal expression is evoked. Young infants start crying whenever it is functionally required to bring the infants to closer proximity with their caregivers.
2nd	While produced vocalizations are preprogrammed, contexts in which they are produced or responses to them are learned	While some prosimians’ alarm calls are predispositionally different according to the differences of types of predators, the difference is learned by sympatric related species. Sympatrically living lemurs can perceive two types of sifaka’s alarm calls differently, which are produced according to aerial and terrestrial predators, respectively, but lemurs with no contact with sifakas cannot.	Although the acoustic pattern of speech-like vocalizations by three-month-olds is invariate, they are capable of volitionally controlling its production. After vocalizing spontaneously, the infants, waiting for the mothers’ responses, vocalize again in bursts if the mothers are unresponsive.
3rd	Learned vocal motor patterns are produced in response to learned stimuli	In free-ranging groups, affiliated macaques exchange coo calls with one another, volitionally modifying the acoustic feature of the calls. Japanese macaques match pattern of frequency modulation of coos to that of the preceding calls of others if they attempt to respond to the calls.	Nine-month-olds learn the pattern of the use of different pitch contours as a means of signaling different communicative functions. Rising terminal contours are used by them with utterances that demand a response such as request and protests, whereas nonrise is used with functions that label external objects.

to continue its upward drift. At this point, communicative systems resembling human language are assumed to appear. Such a view of the evolution of language is partially in line with the view of language development adopted in this book where I emphasize social communication and ethologically valid models of language usage.

Combining ethological data and dynamic system approaches to the development of action

This scenario of a three-stage evolution of vocal communication appears to share common features with the ethological scenario for the development of language. Three stages are also recognized in the developmental process through which infants' prespeech sounds are transformed into intelligible speech. First, there are natural categories of sounds that emerge when the oral, facial, respiratory and ingestive apparatuses combine and activate at specific stages of anatomical and functional maturation. Second, genetically preprogrammed perceptual mechanisms, together with the input of caregivers, allow infants to respond selectively to the sounds. Through experience with the responses, infants learn to give their vocalizations voluntarily under specific circumstances. Finally, infants select from the universe of possible natural categories of sound patterns by matching their own motor output to the sounds of the ambient linguistic environments.

In this book, in order to elucidate the details of the transitional process through the first to the third stages that are depicted in the ethological scenario, the conceptual framework of a dynamic systems approach to the development of action (Fogel and Thelen, 1987; Thelen and Smith, 1994) will be combined with ethological data. This conceptual framework addresses how complex systems like the human vocal system change over time. In the dynamic systems approach, actions are regarded as a set of relationships between properties defined across child and environment. For example, actions such as walking arise in response to forces from the environment and from muscles as well as according to how skillfully the child functions in performing a specific task. In this view, actions are softly assembled, online, by marshaling the dynamic properties of the body relative to how a particular task is perceived.

The principles of dynamic systems are very general and can be applied both to the assembly of behavior in real time and to the emergence of behavior in ontogenetic time. That is, in real time, these principles speak to how articulators cooperate to produce consonant–vowel syllables as well as to how infants progress, in ontogenetic time, from vegetative to speech-like vocalizations. The dynamic systems approach is especially powerful because it focuses not only on the products or end states, but also on the processes that give rise to new forms of behavior and development. In contrast, perhaps as an historical consequence of long-lasting debates with behaviorism, ethologists are often overly preoccupied with a genetically deterministic view of behavior and discussions of how the behavior develops ontogenetically. Therefore in the book, I combine ethological data on early language development with the dynamic systems interpretation of behavioral development.

In the next chapter, I will look at one of the first steps in the process of language acquisition, the onset of vocal turn-taking. If caregiver–infant interaction is critical to infants' language acquisition, and if caregiver–infant alternations of behavior form an important component of interaction, then we must attend to the phenomenon of vocal turn-taking. A rudimentary form of vocal turn-taking behavior has been observed among nonhuman primates (Masataka and Biben, 1987; Sugiura and Masataka, 1995), and it is thought to have been phylogenetically inherited by humans. Further, I argue that in humans the practice of vocal turn-taking facilitates the acquisition of a native language even during early infancy because in order to perceive and reproduce sound patterns, infants must have good perceptual access to the material to be reproduced. Also, to respond contingently, caregivers must hear the infant's reproductions of their speech with some clarity. In fact, several investigators have found a marked increase in vocal turn-taking between twelve and eighteen weeks of age (Ginsburg and Kilbourne, 1988). Studies suggest that at this age, infants begin to inhibit their own vocalizations if their mother is speaking and to fall silent if their mother starts to speak.