Birdsong as model for infants' emergent speech – a brief introduction

Axel G. Ekström
KTH Speech, Music & Hearing, Sweden
axeleks@kth.se

Abstract

Songbirds have long and widely been considered a model species for the development of human speech capacities. Modelling efforts are dependent on parallels and similarities between emergent song and speech behavior. The present text describes eight such parallels, including, among others, neural lateralization, critical periods of development, and a dependency on auditory and perceptual feedback for normal development. The text takes as its unit of comparison patterns of speech observed in developing infants and patterns of song observed in juvenile songbirds, and serves at once as general summary of classic and contemporary research on the two phenomena, as well as a brief introduction to the topic.

Introduction

The study of birdsong has primarily been of interest to researchers in neuroscience (Konishi, 1985), language acquisition (Goldstein et al., 2003), and learning (see Simonyan et al., 2012). In particular, its relationship as a model for emergent human speech and language has been at the center of much influential research. Human speech and birdsong are both learned behaviors not present in the organism at birth. Though features of oscine vocal anatomy and physiology (Greenwalt, 1968) differ significantly from those of humans (Ladefoged, 1996) – and though such differences have distinct and obvious acoustic consequences – the two can be usefully conceptualized as comparable. Systems of vocalization in both species are *a priori* free (there is no objectively more beneficial system of vocalization) and subject to relatively well defined constraints, including the limitations resulting from the musculature of the larynx and syrinx (the avian vocal organ), respectively (see Nottebohm, 1970).

Intraspecies social vocalization represents an ancient evolutionary heritage (Bass et al., 2008), but vocal learning – the capacity to repeat vocally that which has been previously heard (Janik & Slater, 2000) – is an ability shared by only a few disparate lineages, including humans, pinnipeds, bats, and cetaceans among mammals; and parrots, hummingbirds, and oscine birds among avian species (see e.g., Tyack et al., 2020). Among primates, only humans consistently exhibit sophisticated vocal learning (see Egnor & Hauser, 2004). While vocal learning has been observed in many nonhuman species, oscines (hereafter songbirds) have been most studied, with pioneering work reported by Metfessel (1935) and Thorpe (1954) laying the ground for more contemporary work (e.g., Prather et al., 2017; Yazaki-Sugiyama, 2019).

Such efforts have identified substantial similarities between the processes by which a developing songbird acquires song, and those by which human infants learn to speak. There exist also surprising similarities between songbird song and human speech neural systems and circuitry, resulting from convergent evolution (Colquitt et al., 2021). Accordingly, various researchers have drawn on behavioral parallels between birdsong and human speech to guide theoretical (Marler, 1970; Doupe & Kuhl, 1999; Bolhuis et al., 2010; Prather et al., 2017) and empirical (Goldstein et al., 2003) work on language development. The present text summarizes and updates, in a brief and accessible format, this research tradition. It is intended both as a general summary, as well as an introductory text to anyone with an interest in the relationship between the two phenomena.

Behavioral parallels

Lateralization of function

In humans, both speech production (e.g., Wildgruber et al., 1996) and speech perception (e.g., Zatorre et al., 1992) exhibit significant neural lateralization; that is, domain-specific activity tends to be concentrated in one hemisphere (typically, though not exclusively, the left). Songbirds exhibit similar lateralization of vocal control. For example, severing the left hypoglossus (a small neural structure innervating the syrinx) in adult chaffinches (*Fringilla coelebs*) with established ("crystallized") song patterns, results in the loss of a majority of song elements from the established pattern, or a modification of all elements. In comparison, effects of severing the right hypoglossus are much more limited, sometimes with no effect on song structure at all (Nottebohm, 1970). Crucially, if severing takes place prior to any song learning, the chaffinch may still develop normal song – a sequence of events that distinctly parallels that of humans, where in the event of damage or lesioning early in life, speech and language functions typically lateralized to the dominant hemisphere may still be assumed by the non-dominant hemisphere and develop normally (e.g., Lenneberg, 1967).

A predisposition to learning

In both humans and songbirds, developing individuals must solve similar adaptive problems in ontogeny, that is, adapting repertoires of vocal output to ambient sounds as observed in mature conspecifics. However, it is not given how this problem should be resolved. An endlessly prepared imitating brain is due to learn all sounds in the immediate environment, including those of objects and nature – as opposed to being concentrated on learning species-typical vocal behavior. However, human infants and children develop language with remarkable rapidity; they develop sophisticated language, even without apparent direct stimulation, as has been observed in cultures where children are rarely or never directly addressed by adults (Ochs & Schieffelin, 1984; Pinker, 2003). Comparably, in songbirds, song development apparently follows ready-set constraints, such that it develops predictably with various experimental conditions (see Nottebohm, 1970) including naturally occurring song in wild birds (Thorpe, 1954), song in captive birds deafened at hatching (Konishi, 1964), and song in birds reared in isolation from conspecifics (Marler, 1981).

Critical periods

Like human childrens' language acquisition, songbirds' song learning is subject to a developmental period of attrition of possible expression, critical periods, during which domain-specific learning is possible – and after which, barring necessary influences, adequate learning cannot take place (Marler & Tamura, 1964; Boettiger & Doupe, 2001; Yazaki-Sugiyama, 2019; Nottebohm, 1970; Ruben, 1997; DeKeyser, 2000). Neurologically, speech is not lateralized at birth but allocated space throughout development; evidence to this effect comes from observations that children that acquired deafness before three years of age shows atypical cerebral representation of speech functions (Marcotte & Morere, 1990). Again, this process of lateralization appears mirrored in songbirds, where auditory song functions may be similarly constrained, appearing emergent in the right caudomedial nidopallium in birds that have begun to sing, compared to pre-singing ones (Vahaba et al., 2017).

A basis in central pattern generators

Central pattern generators (CPGs) are groupings of neurons responsible for predictable rhythmic output (Grillner & Wallen, 1985). Humans utilize innate CPGs for a variety of behaviors, including walking and chewing. For present purposes, CPGs are involved in the execution of suckling (Barlow, 2009) and later in life for mastication – processes which are ontogenetically coopted for speech (Barlow et al., 2010; Lund & Kolta, 2006). Suboscines (species belonging to the Passerine suborder Tyranni) utilize a CPG for song (Kroodsma & Konishi, 1991; see also Konishi, 2010), and thus develop species-typical vocalization even when deafened (see also *Dependence on auditory feedback* below). By comparison, songbirds can apparently modify innate CPGs according to an internal proprioceptive template. Thus, both humans and songbirds make use of CPGs for the basic productive motor patterns involved in species-typical vocal communication and build on top of them using learned experience.

Periods of early vocalization

Human infants are born without language, but predictably begin to babble – reliably producing syllabic speech-like utterances – around the age of six months (e.g., Oller, 2000). Crucially, birds exhibit similar ontogenetic development toward proper song. For example, wild chaffinches begin producing stable song patterns around 10 months of age – but this period is proceeded in ontogeny by subsong, described by Nottebohm (1970, p. 951) as "a rather soft and rambling vocalization", without apparent direction. Both behaviorally and acoustically, then, there is apparent similarity between subsong and babbling, as periods characterized by species-typical vocal practice. This relationship and parallels between the practice in songbirds and human infants was first discussed at length by Marler (1970) (see also Kuhl, 2003).

Songbird subsong appears to be controlled by neural architecture and circuitry distinct from that representing adult vocalization (Aronov et al., 2008). Similarly, while human vocalizations in general are controlled by distinct brain regions, including the laryngeal motor cortex (Brown et al., 2008; Simonyan & Horwitz, 2011), pre-speech vocalizations such as crying are neurologically differentiable from later-in-life speech production (Okanoya et al., 2002), and possibly characterized by distinct patterns of involved neural circuitry (Jürgens, 2009). Indeed, crying, unlike speech, may be produced via neural circuits conserved across the mammalian lineage (for an overview of related works and theories on mammalian crying and its neural underpinnings, see Newman, 2007).

Dependence on auditory feedback

In songbird species such as the Zebra finch (*Taeniopygia guttata*), auditory feedback is necessary for matching explorative vocal output against intended sounds (i.e., a sensory template; Konishi, 2010; 1965a, 1965b, 1985; Price, 1979; Brainard & Doupe, 2000). This was most clearly made evident through the work of Masakazu Konishi in his studies of deafened songbirds, that subsequently failed to develop adequate songs (Konishi, 1964; 1965b); and the results of his experiments were later replicated in other songbird species (Waser & Marler, 1977). Birds such as chickens (*Gallus domesticus*), that utilize a CPG for song produce species-typical vocalizations, even when deafened (Konishi, 1963a). By comparison, the speech in deaf-born humans is often underarticulated (e.g., Hudgins & Numbers, 1942) and monotone (e.g., Smith, 1975), seemingly equivalent to the song of birds deafened in early development. Further, deaf-born infants exhibit impaired development of babbling behavior, such that hearing infants may begin producing differentiated syllables as early as 10 months after birth, while deaf-born infants do not (Oller & Eilers, 1988).

In his doctoral work, Konishi (1963b) posited "template theory", according to which a young bird will memorize the song of a conspecific tutor individual, using that song as points of reference in future own song development and elaboration. According to the theory, a young bird hears its own song and compares it to that of its sensory template; in the event of a mismatch between the two, the bird continually adjusts its song until the song matches the template. Experimental results on juvenile songbirds' emergent songs have largely been interpreted as aligning with the theory (see Konishi, 2010). While invasive surgical procedures are not available for research on human infant speech ontogenesis for ethical reasons, empirical evidence suggests similar mechanisms of imitation in the context of infants' vowel sound matching (Kuhl & Meltzoff, 1996).

Development requires improvisation

The conversion from auditory template to proprioceptive template requires a young bird to map the auditory consequences of production to corresponding motor activity in the vocal organs. Such relationships are initially unknown, however. Consequently, various researchers have hypothesized as to the importance of motor exploration in song learning (Poulsen, 1959; Waser & Marler, 1977). While direct empirical evidence are as of yet unavailable, clues as to the validity of the theoretical assumptions are not. For one, even in adulthood some songbird species are capable of adaptively shifting the fundamental frequency of some targeted portion of a song to avoid disruption, consistent with some degree of flexibility across the lifespan (Tumer & Brainard, 2007). Similarly, in humans, Kuhl and Meltzoff (1996) have shown infants that are apparently capable of storing speech sound patterns in memory, which subsequently help guide the process of imitation. Work by Sober et al. (2008) further suggests that individual features of acoustic vocal production output are subject to separate loci of neural ensemble control. Finally, recent work on zebra finches by Tchernichovski et al. (2021) suggests that individual birds with low-diversity song tutors compensate by producing greater variability in own vocal production. Though no equivalent evidence exists for human infant vocal production, such findings suggest intriguing possibilities for future modelling and empirical work.

Learning generates dialects

In human speakers, accents and dialects – and over time, languages – develop via sound change, through perception (Ohala, 2012) and subsequent imitation (Hockett, 1965) by new speakers. Pioneering work by Marler and Tamura (1962) documented how populations of white-crowned sparrows (*Zonotrichia leucophrys nutalli*), living only miles apart, also exhibited distinct song patterns (see Kroodsma, 2004). There is as of yet no consensus as to the mechanisms of formation of songbird dialects (for an overview, see Podos & Warren, 2007), and the topic as such is outside the scope of the present text. For present purposes, however, it is sufficient to state that while the phenomenon of birdsong dialects is not of immediate relevance to an *individual* human or songbird's learning of speech or song respectively (but rather a population-level phenomenon), the fact that analogous population-level differences in both species result from mechanisms underlying both systems bears mentioning, as it suggests further mechanistic convergence.

Noteworthy discrepancies

Finally, similarities between emergent birdsong and speech are not unlimited, and some notable differences should be noted. First, birdsong is primarily a sexually selected trait, and while females of most songbird species are equipped with mechanisms for evaluating the songs of males, females themselves typically do not sing (but see e.g., Krieg & Getty, 2016). Second, interspecies differences in birdsong learning and crystallization patterns are widespread. Whereas males of some species sing continually throughout the year, others, such as canaries (*Serinus canaria*), go through seasonal song cycles with changes in testosterone (Nottebohm et al., 1986). Finally, humans everywhere are capable of learning new vocal patterns throughout the lifespan, whereas birdsong is typically either stable across the lifespan following song pattern crystallization or goes through a period of change with each coming breeding season (e.g., Nottebohm et al., 1986); again, there is significant variation between species (see Tumer & Brainard, 2007). Nevertheless, similarities between speech and birdsong are multiple and significant, and deserving of a treatise as background for more in-depth comparison on the level of neural structure and activity.

Summary

Available evidence suggests significant overlap between the ontogenies of human infant speech and juvenile songbird birdsong. The purpose of the present text is to present an overview of a set of the most commonly observed and most well researched of these similarities, in an accessible format. It is the hope of this author that the text can inform future work on the topic of speech development and language acquisition.

Acknowledgements

The author thanks Ghazaleh Esfandiari Baiat for comments on a previous version of the manuscript.

References

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science, 320(5876), 630-634.

Barlow, S. M. (2009). Central pattern generation involved in oral and respiratory control for feeding in the term infant. Current opinion in otolaryngology & head and neck surgery, 17(3), 187.

Barlow, S. M., Radder, J. P. L., Radder, M. E., & Radder, A. K. (2010). Central pattern generators for orofacial movements and speech. Handbook of behavioral neuroscience, 19, 351-369.

Bass, A. H., Gilland, E. H., & Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain–spinal compartment. Science, 321(5887), 417–421.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral cortex*, *18*(4), 837-845.

Boettiger, C. A., & Doupe, A. J. (2001). Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. Neuron, 31(5), 809–818.

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. Nature Reviews Neuroscience, 11(11), 747–759.

Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. Nature Reviews Neuroscience, 1(1), 31–40.

Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F., & Brainard, M. S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. Science, 371(6530), eabd9704.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. Studies in second language acquisition, 22(4), 499-533.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. Annual review of neuroscience, 22(1), 567-631.

Egnor, S. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. Trends in neurosciences, 27(11), 649–654.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. Proceedings of the National Academy of Sciences, 100(13), 8030–8035.

Greenwalt, C. H. (1968). Bird Song: Acoustics and Physiology. Smithsonian Institution Press. Washington, D.C.

Grillner, S., & Wallen, P. (1985). Central pattern generators for locomotion, with special reference to vertebrates. Annual review of neuroscience, 8(1), 233–261.

Hockett, C. F. (1965). Sound change. Language, 41(2), 185–204.

Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. Genetic psychology monographs.

Janik, V. M., & Slater, P. J. (2000). The different roles of social learning in vocal communication. Animal behaviour, 60(1), 1-11.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of Voice*, *23*(1), 1-10.

Konishi, M. (1963a). The Role of Auditory Feedback in the Vocal Behavior of the Domestic Fowl 1. Zeitschrift für Tierpsychologie, 20(3), 349–367.

Konishi, M. (1963b). The role of audition in the development and maintenance of avian vocal behavior. PhD thesis, University of California, Berkeley.

Konishi, M. (1964). Effects of deafening on song development in two species of juncos. The Condor, 66(2), 85–102.

Konishi, M. (1965a). The Role of Auditory Feedback in the Control of Vocalization in the White‐Crowned Sparrow 1. Zeitschrift für Tierpsychologie, 22(7), 770–783.

Konishi, M. (1965b). Effects of deafening on song development in American robins and black-headed grosbeaks. Zeitschrift für Tierpsychologie.

Konishi, M. (1985). Birdsong: from behavior to neuron. Annual review of neuroscience, 8(1), 125-170.

Konishi, M. (2010). From central pattern generator to sensory template in the evolution of birdsong. Brain and language, 115(1), 18–20.

Krieg, C. A., & Getty, T. (2016). Not just for males: females use song against male and female rivals in a temperate zone songbird. Animal Behaviour, 113, 39–47.

Kroodsma, D. E. (2004). The diversity and plasticity of birdsong. Nature's music: the science of birdsong, 108–131.

Kroodsma, D. E., & Konishi, M. (1991). A suboscine bird (eastern phoebe, Sayornis phoebe) develops normal song without auditory feedback. Animal Behaviour, 42(3), 477-487.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. The journal of the Acoustical Society of America, 100(4), 2425–2438.

Kuhl, P. K. (2003). Human speech and birdsong: communication and the social brain. Proceedings of the National Academy of Sciences, 100(17), 9645–9646.

Ladefoged, P. (1996). Elements of acoustic phonetics. University of Chicago Press.

Lenneberg, E. H. (1967). Biological foundations of language. Hospital Practice, 2(12), 59-67.

Lund, J. P., & Kolta, A. (2006). Brainstem circuits that control mastication: do they have anything to say during speech?. Journal of communication disorders, 39(5), 381–390.

Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. Science, 146(3650), 1483–1486.

Marler, P. (1970). Birdsong and speech development: Could there be parallels? There may be basic rules governing vocal learning to which many species conform, including man. American scientist, 58(6), 669-673.

Marler, P., & Waser, M. S. (1977). Role of auditory feedback in canary song development. Journal of comparative and physiological psychology, 91(1), 8.

Marler, P. (1981). Birdsong: the acquisition of a learned motor skill. Trends in Neurosciences, 4, 88-94.

Metfessel, M. (1935). Roller canary song produced without learning from external sources. Science, 81(2106), 470–470.

Marcotte, A. C., & Morere, D. A. (1990). Speech lateralization in deaf populations: evidence for a developmental critical period. Brain and Language, 39(1), 134–152.

Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behavioural brain research*, *182*(2), 155-165.

Nottebohm, F. (1970). Ontogeny of Bird Song: Different strategies in vocal development are reflected in learning stages, critical periods, and neural lateralization. Science, 167(3920), 950–956.

Nottebohm, F., Nottebohm, M. E., & Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. Behavioral and neural biology, 46(3), 445–471.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization. Culture theory: Essays on mind, self, and emotion, 276-320.

Ohala, J. The listener as a source of sound change: An update. (2012). In M. J. Solé, & D. Recasens, (Eds.). The initiation of sound change: Perception, production, and social factors (pp. 21-35). John Benjamins Publishing.

Okanoya, K., Ichii, H., Ushijima, H. (2002). Baby cry as a pre-adaptation to language. Proceedings of Fourth International Conference on the Evolution of Language, Harvard, March 2002, pp. 86

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. Child development, 441–449.0

Oller, D. K. (2000). The emergence of the speech capacity. Psychology Press.

Pinker, S. (2003). The language instinct: How the mind creates language. Penguin UK.

Podos, J., & Warren, P. S. (2007). The evolution of geographic variation in birdsong. Advances in the Study of Behavior, 37, 403-458.

Poulsen, H. (1959). Song learning in the domestic canary. Zeitschrift für Tierpsychologie, 16(2), 173–178.

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. Neuroscience & Biobehavioral Reviews, 81, 225–237.

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. journal of comparative and physiological psychology, 93(2), 260.

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. Acta oto-laryngologica, 117(2), 202-205.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, *17*(2), 197-208.

Simonyan, K., Horwitz, B., & Jarvis, E. D. (2012). Dopamine regulation of human speech and bird song: a critical review. Brain and language, 122(3), 142-150.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. Journal of speech and hearing research, 18(4), 795–811.

Sober, S. J., Wohlgemuth, M. J., & Brainard, M. S. (2008). Central contributions to acoustic variation in birdsong. Journal of Neuroscience, 28(41), 10370–10379.

Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. D. (2021). Balanced imitation sustains song culture in zebra finches. Nature communications, 12(1), 1-14.

Thorpe, W. H. (1954). The process of song-learning in the chaffinch as studied by means of the sound spectrograph. Nature, 173(4402), 465-469.

Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of 'crystallized'adult birdsong. Nature, 450(7173), 1240–1244.

Tyack, P. L. (2020). A taxonomy for vocal learning. Philosophical Transactions of the Royal Society B, 375(1789), 20180406.

Vahaba, D. M., Macedo-Lima, M., & Remage-Healey, L. (2017). Sensory coding and sensitivity to local estrogens shift during critical period milestones in the auditory cortex of male songbirds. eneuro, 4(6).

Waser, M. S., & Marler, P. (1977). Song learning in Canaries. Journal of Comparative and Physiological Psychology, 91(1), 1.

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: a fMRI study. Neuroreport, 7(15-17), 2791-2795.

Yazaki-Sugiyama, Y. (2019). Neuronal mechanisms regulating the critical period of sensory experience-dependent song learning. Neuroscience Research, 140, 53–58.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. Science, 256(5058), 84.