

CHAPTER 7.2

The polyvagal hypothesis: common mechanisms mediating autonomic regulation, vocalizations and listening

Stephen W. Porges* and Gregory F. Lewis

Brain-Body Center, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA

Abstract: This chapter applies hypotheses and principles derived from the polyvagal theory (Porges, 1995, 2001, 2007) to interpret mammalian vocalizations. The theory emphasizes the parallel phylogenetic shift in both the neural regulation of the autonomic nervous system and the evolutionary emergence of an integrated social engagement system that includes features optimizing conspecific mammalian vocal communication. The chapter presents the polyvagal hypothesis as a new way to interpret adaptive functions and acoustic features of mammalian vocalizations. The chapter emphasizes both the neural mechanisms involved in social communication, including the reciprocal relationship between production and reception of vocalizations, and the perceptual advantage that mammals have by vocalizing within a frequency band optimized by the physics of the middle ear.

Keywords: autonomic nervous system; polyvagal theory; middle ear muscles; heart rate variability; vagus; social engagement system; perceptual advantage

I. Introduction

Vocalizations are an intricate component of the complex biobehavioral repertoire of mammals. Since vocalizations are involved in coordinating behaviors for mammals that are both prosocial and survival oriented, the predominance of research on mammalian vocalizations has focused on describing the adaptive functions of specific vocal features. In contrast, there is a paucity of research investigating the neurophysiological linkage between autonomic regulation and both the production of vocalizations and the processing of acoustic information. This chapter applies hypotheses and principles derived from the polyvagal

theory (Porges, 1995, 2001, 2007) to interpret mammalian vocalizations. The theory emphasizes the parallel phylogenetic shift in both the neural regulation of the autonomic nervous system and the evolutionary emergence of an integrated social engagement system that includes features optimizing conspecific vocal communication. The chapter will present the polyvagal hypothesis as a new way to interpret the adaptive functions of mammalian vocalizations. The chapter will emphasize neural mechanisms involved in social communication, including the reciprocal relationship between production and reception of vocalizations.

II. The polyvagal theory: three phylogenetic response systems

The polyvagal theory (Porges, 1995, 2001, 2007) links parallel phylogenetic changes in the neural regulation

*Corresponding author.

Stephen W. Porges, PhD, Brain-Body Center (MC 912), Department of Psychiatry, University of Illinois at Chicago, 1601 W. Taylor Street, Chicago, IL 60612, USA.

of the autonomic nervous system with adaptive behaviors. The theory emphasizes the phylogenetic shift in the features of the vagus, the primary parasympathetic cranial nerve involved in regulating visceral state. The theory is named polyvagal, since only in mammals does the vagus contain two distinct efferent pathways. In addition to an unmyelinated pathway originating in the dorsal motor nucleus of the vagus shared with other vertebrates (i.e., reptilians, amphibians, teleosts and elasmobranchs), mammals have a myelinated pathway originating in the nucleus ambiguus and this circuit communicates in the brainstem with source nuclei regulating the striated muscles of the face and head that are involved in several of the features of social communication (e.g., facial expression, vocalizations, listening).

The polyvagal theory provides a physiological basis to link uniquely mammalian anatomical and physiological circuits involved in visceral state regulation to both the expressive and receptive domains of social communication. The development of these features reflects the phylogenetic distinction between reptiles and mammals, and includes a cluster of behaviors dependent on the function of structures that phylogenetically emerged with mammals. These changes include detached middle ear bones, the advent of a diaphragm, and a myelinated vagal system regulating supradiaphragmatic organs that is distinct from an unmyelinated vagal system regulating subdiaphragmatic organs. The classic definition of mammals focuses on the observation of mammary glands and hair. However, the fossil record, since it cannot be explored through these features, relies on the identification of detached middle ear bones as the defining feature of mammals. Coincident with the separation of the middle ear bones, other phylogenetic transitions resulted in brainstem areas regulating the vagus becoming intertwined with the areas regulating the striated muscles of the face and head. The result of this transition was a dynamic social engagement system with social communication features (e.g., facial expression, head movements, vocalizations and listening) interacting with visceral state regulation.

The polyvagal theory articulates how each of three phylogenetic stages of the development of the vertebrate autonomic nervous system is associated with a distinct autonomic subsystem that is retained and expressed in mammals. These autonomic subsystems are phylogenetically ordered and behaviorally linked to social communication (e.g., facial expression, vocalization, listening), mobilization (e.g., fight-flight

behaviors), and immobilization (e.g., feigning death, vasovagal syncope and behavioral suppression). Social communication (i.e., social engagement system, see Section III) involves the myelinated vagus, which serves to foster calm behavioral states by inhibiting the sympathetic influences to the heart and dampening the hypothalamic-pituitary axis (HPA) (e.g., [Bueno et al., 1989](#)). The mobilization system is dependent on the functioning of the sympathetic nervous system. The most phylogenetically primitive component, the immobilization system, is dependent on the unmyelinated vagus, which is shared with most vertebrates. With increased neural complexity due to phylogenetic development, the organism's behavioral and affective repertoire is enriched. The three circuits can be conceptualized as dynamically adjusting to provide adaptive responses to safe, dangerous and life-threatening events and contexts.

Unlike reptiles, the mammalian nervous system did not evolve solely to survive in dangerous and life-threatening contexts, but it evolved to promote social interactions and social bonds in safe environments. To accomplish this adaptive flexibility, the mammalian nervous system evolved a new neural strategy for safe environments, while retaining two more primitive neural circuits to regulate defensive strategies (i.e., fight-flight and death-feigning behaviors). It is important to note that social behavior, social communication and visceral homeostasis are incompatible with the neurophysiological states and behaviors promoted by the two neural circuits that support defense strategies. Thus, via evolution, the mammalian nervous system retains three neural circuits, which are in a phylogenetically organized hierarchy. In this hierarchy of adaptive responses, the newest circuit associated with social communication is used first, and if that circuit fails to provide safety, the older survival-oriented circuits are recruited sequentially.

By investigating the phylogeny of the regulation of the vertebrate heart ([Morris and Nilsson, 1994](#); [Porges, 1995, 1997](#); [Taylor et al., 1999](#)), four principles can be extracted that provide a basis for speculation regarding emergent behaviors and social communication. These principles lead to testable hypotheses relating the specific neural mechanisms supporting social engagement, fight-flight and death-feigning behaviors:

1. There is a phylogenetic shift in the regulation of the heart from endocrine communication, to unmyelinated nerves and finally to myelinated nerves.

2. There is a development of opposing neural mechanisms of excitation and inhibition to provide rapid regulation of graded metabolic output.
3. A face–heart connection evolved as the brainstem source nuclei of vagal pathways shifted ventrally from the older dorsal motor nucleus to the nucleus ambiguus. This resulted in an anatomical and neurophysiological linkage between the neural regulation of the heart via the myelinated vagus and the special visceral efferent pathways that regulate the striated muscles of the face and head, forming an integrated social engagement system (see Fig. 1) (for more details see [Porges, 2001, 2003, 2007](#)).
4. With increased cortical development, the cortex exhibits greater control over the brainstem via direct (e.g., corticobulbar) and indirect (e.g., corticoreticular) neural pathways originating in motor cortex and terminating in the source nuclei of the myelinated motor nerves emerging from the brainstem (e.g., specific neural pathways embedded within cranial nerves V, VII, IX, X and XI), controlling visceromotor structures (i.e., heart, bronchi) and somatomotor structures (muscles of the face and head).

III. The social engagement system

The study of comparative anatomy, evolutionary biology, and embryology provide important hints regarding the functional relationship between the neural control of the striated muscles of the face and head and emergent behaviors, including facial expressions, head movements and vocalizations. The nerves that control the muscles of the face and head share several common features. Pathways from five cranial nerves control the muscles of the face and head. Collectively, these pathways are labeled as special visceral efferent. Special visceral efferent nerves innervate striated muscles, which regulate the structures derived during embryology from the ancient gill arches ([Truex and Carpenter, 1969](#)). The special visceral efferent pathways regulate the muscles of mastication (e.g., ingestion), muscles of the middle ear (e.g., listening to conspecific vocalizations), muscles of the face (e.g., emotional expression), muscles of larynx and pharynx (e.g., prosody and intonation), and muscles controlling head tilt and turning (e.g., gesture). In fact, the neural pathway that enables the eyelids to be closed in a graded manner (i.e., winking) in humans is also involved in tensing the stapedius muscle in

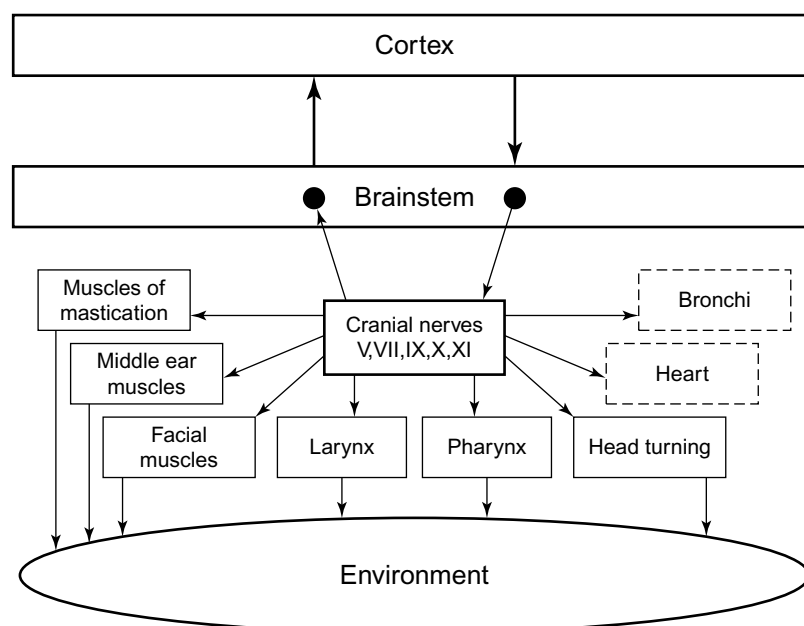


Fig. 1. The social engagement system. The social engagement system consists of a somatomotor component (i.e., special visceral efferent pathways that regulate the striated muscles of the face and head) and a visceromotor component (i.e., the myelinated vagus that regulates the heart and bronchi). Solid blocks indicate the somatomotor component. Dashed blocks indicate the visceromotor component.

the middle ear to facilitate hearing the human voice (Djupestrand, 1976).

The source nuclei of the circuits regulating the striated muscles of the face and head interact in the brainstem with the source nucleus of the myelinated vagus, forming an integrated social engagement system. This system (as illustrated in Fig. 1) provides the neural structures involved in social and emotional behaviors. The social engagement system has a control component in the cortex (i.e., upper motor neurons) that regulates brainstem nuclei (i.e., lower motor neurons) to control eyelid opening (e.g., looking), facial muscles (e.g., emotional expression), middle ear muscles (e.g., extracting vocalizations from background noise), muscles of mastication (e.g., ingestion), laryngeal and pharyngeal muscles (e.g., vocalization) and head turning muscles (e.g., social gesture and orientation). Collectively, these muscles function as filters that limit social stimuli (e.g., observing facial features and listening to vocalizations) and determine engagement with the social environment. The neural control of these muscles determines social experiences by changing facial features (especially in humans and other primates), modulating laryngeal and pharyngeal muscles to regulate intonation of vocalizations, and coordinating both facial and vocal motor tone with respiratory actions (see also Smotherman et al., Chapter 9.2 in this volume). In addition, the frequency of breathing is encoded into the phrasing of vocalizations and may convey urgency by expressing short phrases associated with short exhalations (i.e., rapid breathing) or convey calmness by expressing long phrases associated with long exhalations (i.e., slow breathing).

The source nuclei (i.e., lower motor neurons) of the special visceral efferent pathways are located in the brainstem and communicate with the source nuclei of the mammalian myelinated vagus, which functions as a neural pathway inhibiting visceral functions (e.g., slowing heart rate, lowering blood pressure, etc.) to actively reduce arousal to promote calm states consistent with the metabolic demands of growth and restoration. Direct corticobulbar pathways reflect the influence of frontal areas of the cortex (i.e., upper motor neurons) on the regulation of this system. Moreover, afferent feedback, through the sensory vagus to medullary areas (e.g., nucleus of the solitary tract) influences forebrain areas and regulates states of arousal and alertness. In addition, the anatomical structures involved in the social engagement system have neurophysiological interactions with the HPA axis, the neuropeptides of oxytocin and vasopressin, and the immune system (for overview see Porges, 2001).

IV. Phylogeny of the middle ear

As vertebrates evolved from reptiles to mammals, the structures at the end of the mandible (i.e., jaw bone) that define components in the middle ear became detached (Rowe, 1996; Luo, 2007; Luo et al., 2001; Wang et al., 2001) and formed the small bones (auditory ossicles) of the middle ear. For humans and other mammals, sound in the environment impinges on the eardrum and is transduced from the eardrum to the inner ear via the ossicles in the middle ear. Bárány (1938) argued that the main purpose of the ossicular chain in land mammals was to reduce the effect of low-frequency sounds via bone conduction. In addition to the filtering imposed by bone separation, further attenuation can be achieved when the stapedius muscle (stabilizing stapes and innervated via a branch of the facial nerve) and the tensor tympani muscle (innervated via a branch of the trigeminal nerve) are contracted. Tension reduces the compliance of the ossicular chain and dampens the amplitude of the low-frequency acoustic stimulation from the environment reaching the inner ear. This process is similar to tightening the skin on a kettledrum. When the skin is tightened, the pitch of the drum is higher. When the ossicular chain is tightened, similar to the stretched skin, the movement of the eardrum is reduced and only higher frequencies bouncing against the eardrum are transmitted to the inner ear and to the auditory processing areas of the brain. The impact of these muscles on the perceived acoustic environment is to markedly attenuate low-frequency sounds, which facilitates the extraction of sounds in a higher frequency band associated with human voice and other mammalian vocalizations.

The detachment of the middle ear bones from the mandible was paralleled by two phylogenetic shifts: (1) the jawbone no longer restricted the cranium to expand and fostered the cortical development that characterizes modern mammals (Rowe, 1996); and (2) the mammalian middle ear enabled low-amplitude relatively high-frequency airborne sounds (i.e., sounds in the frequency of vocalizations) to be heard, even when the acoustic environment was dominated by low-frequency sounds.

The evolution of the mammalian middle ear enabled mammals to communicate in a frequency band that could not be detected by reptiles that, due to a dependence on bone conduction, were able to hear predominantly lower frequencies. This ability to hear low-amplitude high-frequency airborne sounds in an acoustic environment dominated by loud low-frequency sounds is accomplished when the middle ear

muscles are tensed to create rigidity along the ossicular chain. This mechanism unmasks the high-frequency sounds associated with mammalian vocalizations from background sounds. Without stiffening the ossicular chain, mammals would lose this advantage and the soft airborne sounds of vocalizations would be easily lost in the acoustic background (see [Borg and Counter, 1989](#)). In fact, individuals who can voluntarily contract middle ear muscles exhibit an attenuation of approximately 30 dB at frequencies below 500 Hz, while there is no or minimal attenuation at frequencies above 1,000 Hz (see [Kryter, 1985](#)).

The earliest mammals were small and vocal communication outside the acoustic range of their predominant predators (i.e., reptiles) was crucial to survival. The physics of the middle ear resulted in a specific range of frequencies requiring a lower sound pressure level to be detected. In small mammals this frequency band was noticeably higher than the frequency band that large reptiles, due to a dependency on bone conduction, could easily detect. However, as mammals evolved, selective pressure resulted in larger mammals surviving. The size of the middle ear's structures increased with increased body size and the resonant frequency of the middle ear became lower. Thus, with large mammals (e.g., elephants and whales) the frequencies enhanced by middle ear structures foster infrasound communication with wavelengths that can travel over long distances and overlap with the lower frequencies that reptiles detect via bone conduction.

V. Impact of middle ear structures on sensitivity to conspecific vocalizations

The perception of sound is not equal at all frequencies. We hear sounds at low frequencies as if they were softer than their actual physical energy. In contrast, humans are relatively accurate in estimating the acoustic energy of frequencies associated with voice. This phenomenon initially reported as the Fletcher–Munson equal loudness contours ([Fletcher and Munson, 1933](#)), illustrated how human perception attenuated the “loudness” of low-frequency sounds. As measurement technologies improved, researchers refined the perceived loudness contours, and sound meters were modified to include a scale known as dB(A), which adjusted for the perceived differences in loudness as a function of frequency (i.e., the acoustic energy of lower frequencies had to be greatly increased to be perceived at the equivalent loudness of higher frequencies). This contrasts to sound pressure level, which describes the

physical energy of the signal and does not apply any perceptually based weighting to the frequencies that constitute the acoustic stimulation.

The perceptual process of detecting conspecific vocalizations in background noise, illustrates the antimasking function of the middle ear muscles (attenuating the sounds at low frequencies). In addition to the antimasking function of the middle ear muscles, the middle ear structures act as a natural amplifier and contribute to the advantage in detecting conspecific vocalizations. Amplification occurs when the acoustic energy of vocalizations is characterized by a frequency band that overlaps with the resonance frequencies of the middle ear structures. Thus, due to the selective active antimasking by the middle ear muscles and the passive amplification by the middle ear structures, there are species-specific equal loudness contours. As a general rule, conspecific vocalizations occupy this frequency band of perceptual advantage.

[Dammeijer and colleagues \(2007\)](#) evaluated the effect of noise exposure on the stapedius muscle in the rat. Their data suggest that even in the absence of loud noise the stapedius is active, with contraction observed at sound pressure levels much lower than those needed to elicit the acoustic reflex threshold ([Pilz et al., 1997](#)). In addition, the data were consistent with the assumed purpose of the stapedius in unmasking high-frequency signals by attenuating low-frequency low-level constant noise in everyday life ([Pang and Guinan, 1997](#)). The middle ear muscles are composed primarily of relatively small fast-twitch fibers ([de Jong et al., 1988](#)). Since the muscles are characterized, at least in the rat, by relatively high anaerobic, glycolytic and aerobic oxydative enzyme activity, there is a natural resistance to fatigue. In addition, the numerous motor end plates, in combination with the numerous axon bundles, are indicative of small motor units, and support the assumption that the middle ear muscles are able to perform finely graded contractions. In our laboratory, we are demonstrating a similar intensity-tuned function of the middle ear muscles in humans.

The antimasking role of the middle ear muscles in auditory processing is especially relevant when considering the impact of low-frequency background sounds on cochlear mechanisms. Standing waves from pure tone stimuli occupy larger regions of the basilar membrane as the intensity increases, reducing the sensitivity of cochlear filtering for intense stimuli. This process is observed in the flattening of the equal loudness profile. Thus, by attenuating low-frequency sound waves, even below the level of the acoustic reflex threshold, the graded contractions of the middle

ear muscles would improve the frequency sensitivity and selectivity mediated by the outer hair cells in the frequency band of mammalian vocalizations.

The physics of the middle ear structures impose other filter characteristics. Although the stiffening of the ossicular chain functions as a highpass filter by contracting the middle ear muscles and dampening the influence of low-frequency sounds on the inner ear, the physical characteristics of the ossicular chain also influence the acoustic energy reaching the inner ear. Ossicle inertia determines the highest frequencies that can pass through the middle ear (Hemila et al., 1995). The high-frequency limit is inversely proportional to the ossicular mass. Although, in general, larger mammals have greater ossicular mass, there are instances of larger mammals being able to detect higher frequencies due to adaptive shifts in ossicle mass to facilitate the detection of sounds associated with predator, prey and cohort. For example, cats have excellent sensitivity at very high frequencies and may benefit from hearing high-frequency sounds produced by small rodents (Rosenzweig and Amon, 1955; Forsman and Malmquist, 1988).

VI. The frequency band of perceptual advantage

In very small mammals, the middle ear and inner ear structures can convey acoustic information in a range well above the audible sounds that humans can reliably detect. The upper frequency limit for audible sounds in humans is approximately 20,000 Hz. Thus, 20,000 Hz is often used to delineate audible from ultrasound in the acoustic spectrum. However, the terminology is misleading, since acoustic stimuli characterized as ultrasonic are within the “audible” range of several mammalian species.

When audiograms of mammals are plotted on a logarithmic x-axis (for review see Fay, 1988), the lowest threshold is observed in a band of frequencies functionally defined by the high pass of the middle ear muscles and the low pass features of the combined effect of olivary-cochlear mechanisms and the inertia of the ossicular chain. This frequency band of perceptual advantage (i.e., lower thresholds to hear) is specific to each mammalian species with the smaller mammals, in general, having an advantage to hear higher frequencies. However, since low-frequency sounds dominate most acoustic environments this frequency band of perceptual advantage is optimized only when lower frequencies do not overwhelm the acoustic apparatus.

Thus, the importance of contracting the middle ear muscles is to reduce acoustic energy in the lower frequencies. In general, it is within this frequency band of perceptual advantage that mammalian species produce most of their relevant conspecific vocalizations. For example, with the human, although the generally accepted frequencies for human hearing are between 20 and 20,000 Hz, the human frequency band of perceptual advantage includes a band of frequencies from approximately 500 Hz to about 4,000 Hz. Within these frequencies, the second and third formant in both male and female human speech always occur and in many cases so does the first formant. This select band, which conveys the information of human voice, is functionally amplified by the antimasking mechanisms of the middle ear muscles that dampen low-frequency activity and olivary-cochlear mechanisms that dampen high-frequency activity. Similar mechanisms in the rat provide a frequency band of perceptual advantage from approximately 5 kHz to 50 kHz (see Bjork et al., 1999).

Auditory information can be weighted within this frequency band to improve the extraction of human speech. Two such methods are known as the “index of articulation” (Kryter, 1962) and the more recent “speech intelligibility index” (American National Standards Institute, ANSI, 1997). These indices emphasize the relative importance of specific frequencies in conveying speech-related information embedded in human vocalizations. In the normal ear, acoustic energy within the primary frequencies of these indices is not attenuated, as it passes through the middle ear structures to the inner ear. The frequency band defining the index of articulation is similar to the frequency band that composers have historically selected to express melodies. It is also the frequency band that mothers have used to calm their infants by singing lullabies. Modulation of the acoustic energy within the frequencies of human voice that characterize music, similar to vocal prosody, will recruit and modulate the neural regulation of the middle ear muscles, functionally calm the behavioral and physiological state by increasing vagal regulation of the heart, and promote more spontaneous social engagement behaviors. Vocal music duplicates the effect of vocal prosody and triggers neural mechanisms that regulate the entire social engagement system with the resultant changes in facial affect and autonomic state. Basically, we start to look and feel better when we listen to melodies.

The auditory system is capable of compressing the range of acoustic stimuli. Much of this compression

occurs at the periphery. For example, stapedius muscle contraction reduces sound transmission at low frequencies. Zwislocki (2002), said: "placement of stimulus compression in the auditory periphery must have been an important evolutionary adaptation through which the remaining system can operate within a biologically more easily achievable range" (p. 14,601). As Zwislocki suggests, by compressing the energy of loud low-frequency sounds, the inner ear and higher brain structures can process the higher frequency content of the acoustic signal in the range of vocalizations. From an engineering point of view, placing a mechanism to compress the nonlinearity at the periphery of a system reduces the dynamic range required of the remaining parts of the system. This compression mechanism would function as an automatic gain control to filter the acoustic energy at frequencies lower than the vocalizations and would allow higher brain structures to extract meaning and syntax by processing the acoustic energy in this narrower frequency range.

Although humans and other mammals can vocalize outside the frequency band of perceptual advantage, the within-species social communication is usually characterized by frequency modulated vocalizations within this frequency band. In contrast, danger and pain signals may be shrill cries (i.e., high pitch with diminished frequency modulation) at the upper edge of this frequency band. In addition, aggressive signaling may push vocalizations to lower frequencies outside this band (e.g., the roar of a lion). This preference to vocalize in a social context within the frequencies most easily detected by conspecifics has clear adaptive features, but it also creates challenges. In particular, the frequencies of these vocalizations are dependent on the processing of airborne acoustic energy and are above the frequencies easily conveyed through bone conduction. The higher frequencies of airborne mammalian vocalizations (audible and ultrasound) are characterized by very short wavelengths that dissipate rapidly with distance from the source. In contrast, low frequencies have long wavelengths that travel over long distances.

The short wavelength of mammalian vocalizations evolved with convergent mechanisms to aid in adaptive social engagement behaviors (see Porges, 2007). For many mammals, including humans, facial expressivity and behavioral gestures (e.g., use of the hands by primates) are coordinated with the shifts in prosody (intonation) to reduce ambiguity of the acoustic message (Corballis, 2003). Thus, the signals of distress and danger often require concordant facial cues

and hand gestures (see also Eberl, Chapter 11.1 in this volume). Areas in the temporal cortex are sensitive to this cross-modal binding of auditory–visual inputs during vocalizations. Congruent speech-related visual input activates supra-additive multisensory neurons in the superior temporal cortex. In contrast, reduced activation in these areas has been reported during tasks requiring the integration of auditory and visual language inputs in schizophrenics, a disorder frequently associated with auditory hallucinations (Surguladze et al., 2001). Functionally, the simultaneous observation of facial and head movements, while listening to human vocalizations, improves speech intelligibility (e.g., McGurk and MacDonald, 1976; Munhall et al., 2004) and has been reported to increase the ability to extract speech from background sounds by approximately 10–20 dB (Sumby and Polack, 1954; Chen and Rao, 1998).

One of the consequences of depending on high frequencies for social communication is that infants cannot stray far from the protection of their mother. In many small mammalian species (e.g., rats, mice) the predominance of ultrasound vocalizations by infants further restricts the distance that the caregiver can move from the litter. In rats the frequencies used to communicate change developmentally. As rat pups develop and express exploratory behaviors, vocalizations shift from infantile ultrasound to adult type communication, which may also include audible sounds (Takahashi, 1992). As the rat pups mature, well-organized mobilization behaviors support exploration and the pups extend the distance they explore from the mother. Paralleling this shift towards adult vocalizations are increases in the neural regulation of the larynx and pharynx, structures involved in the production and articulation of vocalizations, with a parallel increase in the neural regulation of the heart via the myelinated vagus (Larson and Porges, 1982).

VII. Adaptive cost for active listening to vocalizations

There may be a cost for actively dampening the sensitivity to low-frequency sounds and engaging the neural mechanisms involved in listening to the frequency band of perceptual advantage. Listening to the frequency band of perceptual advantage requires the neural implementation of an active filter that reduces the acoustic information at low frequencies that can reach the brain. Since the sounds associated

with predators, especially the movements of larger animals, are characterized by low-frequency sounds, engaging in this active “listening” process has maladaptive consequences by reducing the ability to detect predators. Thus, the advantage of listening to conspecific vocalization also comes at a cost. In the “wild” the potential cost of social communication is reduction of predator detection. The adaptive consequence of this vulnerability is to restrict or to limit listening to vocalizations within the frequency band of perceptual advantage (an important component of social engagement and social communication) primarily in safe environments such as nests and burrows.

VIII. The social engagement system and the polyvagal “vocalization” hypothesis

As proposed by the polyvagal theory, the functioning and development of the striated muscles of the face and head involved in listening and in production of vocalizations parallel the maturation of the myelinated vagus (see [Larson and Porges, 1982](#)). This developmental convergence, among several neural circuits that constitute an integrated functional social engagement system (see [Fig. 1](#)), results in facilitating several adaptive behaviors including: (1) an improved ability to regulate physiological state via myelinated vagal pathways (i.e., vagal brake) to both self-soothe and maintain calm states, as well as to mobilize by withdrawing the vagal brake to explore, forage and defend; (2) increased neural regulation of the larynx and pharynx to promote conspecific vocalizations in the frequency band of perceptual advantage for the species to selectively signal peers and caregivers with a vocal mechanism that produces vocalizations with the highest adaptive value; and (3) improved thermoregulatory activity that reduces the need for the caregiver, as the autonomic nervous system matures.

The polyvagal theory emphasizes a phylogenetic parallel in the changing neural regulation of the autonomic nervous system and the neural regulation of the striated muscles of the face and head. This point is relevant to the study of mammalian vocalizations, since the striated muscles of the face and head are involved in both the detection of vocalizations (during listening) and in the production of these sounds through the coordination of the laryngeal and pharyngeal muscles with respiratory mechanisms.

The convergent phylogenetic changes in the neural regulation of the structures involved in the production

and detection of mammalian vocalizations leads to the polyvagal hypothesis. Specifically, only mammals have a diaphragm to coordinate vocalizations with respiratory effort and volume. Consistent with the dependence on the mammalian diaphragm is the uniquely mammalian distinction between the two branches of the vagus: one dealing with supradiaphragmatic organs and the other dealing with subdiaphragmatic organs. The neural regulation of the subdiaphragmatic vagus is involved in abdominal breathing, while the neural regulation of supradiaphragmatic vagus is coordinated with the laryngeal and pharyngeal muscles that shape the acoustic features and provide facial expressions consistent with prosodic features of vocalizations. Moreover, slow exhalation, the respiratory process associated with expressive social vocalizations, enhances the impact of the myelinated vagus on the heart, promoting calm states.

The polyvagal hypothesis proposes that acoustic characteristics of vocalization not only serve to communicate to conspecifics relevant features in the environment, but also reflect the physiological state of the producer of the vocalizations. Mammals, but not reptiles or other phylogenetic antecedents to mammals, have a myelinated vagus, a diaphragm, detached middle ears, and neural circuits in the brainstem linking and coordinating the regulation of the myelinated vagus with the regulation of the striated muscles of the face and head. Specifically, engagement of this circuit conveys and expresses states of calmness and safety, and would be associated with greater vagal influences to the heart and lungs to promote calmer physiological states, increased neural tone to the middle ear muscles to optimize listening in the frequency band of perceptual advantage, and increased neural tone to the laryngeal and pharyngeal muscles to shift the acoustic energy of vocalizations to lower frequencies and increased frequency modulation within the band of perceptual advantage. In contrast, retraction of this circuit conveys and expresses states of danger and distress, and would be associated with faster heart rate and breathing and higher pitched vocalizations. Thus, in humans the features of vocal prosody are expressed during social interactions and reduced during both mental and physical illness. Similarly, the vocalizations of small mammals may have a parallel to human prosody. Rats, while playing or experiencing more positive affective states, modulate their ultrasonic vocalizations across a range of frequencies within the rat's band of perceptual advantage, while vocalizations communicating negative states such as

danger are characterized by vocalizations at a relatively constant frequency through a different neural mechanism (e.g., Brudzynski, 2007).

Consistent with the polyvagal hypothesis, medically-compromised human infants have a high-pitched cry with little frequency modulation that is articulated in short bursts (Lester and Zeskin, 1982; Porter et al., 1988). Intonations of infant cries are regulated by neural tone, via source nuclei in the nucleus ambiguus, to the laryngeal muscles and the heart. In the physiologically stressed infant, decreased neural tone theoretically reduces the inhibitory effect on the heart and bronchi and the contraction of laryngeal muscles, thus producing dramatic increases in both heart rate and respiration rate and the fundamental frequency of the cry.

Porter and her colleagues (Porter et al., 1988) reported a convergence between withdrawal of cardiac vagal tone (i.e., measured by respiratory sinus arrhythmia and mediated via the myelinated vagal pathways originating in the nucleus ambiguus) and the shift in fundamental frequency of newborn pain cries in response to circumcision. Cardiac vagal tone was significantly reduced during the severe stress of circumcision, and these reductions were paralleled by significant increases in the pitch of the infants' cries. These results document the important role that the social engagement system has in signaling and responding to "stress" and pain, and provide a demonstration that the vocalizations convey information regarding visceral and emotional state.

IX. Summary

The polyvagal theory emphasizes the phylogenetic shifts in the neural regulation of the autonomic nervous system and how this evolutionary shift in neural regulation converged with the regulation of the middle ear muscles to facilitate mammalian vocal communication. The theory emphasizes the different neural circuits that support defensive behaviors (i.e., fight-flight and freeze) and social interactions. According to the theory, during defensive states, when the middle ear muscles are not contracted, acoustic stimuli are prioritized by intensity and during safe social engagement states, acoustic stimuli are prioritized by frequency. During safe states, hearing of the frequencies associated with conspecific vocalizations is selectively being amplified, while other frequencies are attenuated. During the defensive states, the loud low-frequency sounds signaling a predator could be more easily

detected and the soft higher frequencies of conspecific vocalizations are lost in background sounds. During social engagement behaviors, an integrated social engagement system regulates a shift in autonomic state to dampen sympathetic activity and to increase parasympathetic tone, while simultaneously increasing the neural tone to the striated muscles of the face and head (i.e., facial expressions, increased "emotional" cueing of the eyes associated with increased eye contact, increased prosody and enhanced listening by contracting the middle ear muscles). During social interactions, the stiffening of the ossicular chain actively changes the transfer function of the middle ear, and functionally dampens low-frequency sounds and improves the ability to extract conspecific vocalizations. However, the selectivity to listen to conspecific vocalizations comes at a cost, and the detection of lower acoustic frequencies generated by predators becomes more difficult. Thus, the identification and construction of safe contexts (e.g., burrows, nests, or houses) plays an important role in enabling the social engagement system to promote prosocial behavior.

References

- Bárány, E., 1938. A contribution to the physiology of bone conduction. *Acta Otolaryngol. Suppl.* 26, 1–233.
- Bjork, E., Nevalainen, T., Hakumaki, M., Voipio, H.-M., 1999. R-weighting provides better estimation for rat hearing sensitivity. *Lab. Anim.* 34, 136–144.
- Borg, E., Counter, S.A., 1989. The middle-ear muscles. *Sci. Am.* 261, 74–80.
- Brudzynski, S.M., 2007. Ultrasonic calls of rats as indicator variables of negative and positive states: acetylcholine-dopamine interaction and acoustic coding. *Behav. Brain Res.* 182, 261–273.
- Bueno, L., Gue, M., Fargeas, M.J., Alvinerie, M., Junien, J.L., Fioramonti, J., 1989. Vagally mediated inhibition of acoustic stress-induced cortisol release by orally administered kappa-opioid substance in dogs. *Endocrinology* 124, 1788–1793.
- Chen, T., Rao, R.R., 1998. Audio-visual integration in multimodal communication. *Proc. IEEE* 86 (5), 837–852.
- Corballis, M., 2003. From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav. Brain Sci.* 26, 199–260.
- Dammeijer, P., Dijk, P., Chenault, M., Manni, J., Mameren, H., 2007. Stapedius muscle fibre characterization in the noise exposed and auditory deprived rat. *Hear. Res.* 233, 54–66.
- deJong, F., Kingma, H., Wirtz, P., Berge, H., Marres, E., 1988. Indications of a differentiated regulation of sound transmission by the middle ear muscles of the rat. *Am. J. Otol.* 9 (1), 70–75.

- Djupesland, G., 1976. Nonacoustic reflex measurement: procedures, interpretations and variables. In: Feldman, A., Wilber, L. (Eds.), *Acoustic Impedance and Admittance: The Measurement of Middle Ear Function*. Williams and Wilkins Co., Baltimore, MD, pp. 217–235.
- Fay, R.R., 1988. Comparative psychoacoustics. *Hear. Res.* 34, 295–306.
- Fletcher, H., Munson, W.A., 1933. Loudness: its definition, measurement, and calculation. *J. Acoust. Soc. Am.* 5, 82–108.
- Forsman, K.A., Malmquist, M.G., 1988. Evidence for echolocation in the common shrew, *Sorex araneus*. *J. Zool.* 216, 655–662.
- Hemilä, S., Nummela, S., Reuter, T., 1995. What middle ear parameters tell about impedance matching and high frequency hearing. *Hear. Res.* 85, 31–44.
- Kryter, K.D., 1962. Methods for the calculation and use of the articulation index. *J. Acoust. Soc. Am.* 34, 1689–1697.
- Kryter, K.D., 1985. *The Effects of Noise on Man*. Academic Press, New York, NY.
- Larson, S.K., Porges, S.W., 1982. The ontogeny of heart period patterning in the rat. *Dev. Psychobiol.* 15 (6), 519–528.
- Lester, B.M., Zeskin, P.S., 1982. A biobehavioral perspective on crying in early infancy. In: Fitzgerald, H., Lester, B., Yogman, M. (Eds.), *Theory and Research in Behavioral Pediatrics*. Plenum Press, New York, NY, pp. 133–180.
- Luo, Z.X., 2007. Transformation and diversification in early mammal evolution. *Nature* 450 (7172), 1011–1019.
- Luo, Z.X., Crompton, A.W., Sun, A.L., 2001. A new mammaliaform from the early Jurassic and evolution of mammalian characteristics. *Science* 292, 1535–1540.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264, 746–748.
- Morris, J.L., Nilsson, S., 1994. The circulatory system. In: Nilsson, S., Holmgren, S. (Eds.), *Comparative Physiology and Evolution of the Autonomic Nervous System*. Harwood Academic Publishers, Chur, Switzerland, pp. 193–246.
- Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., Vatikiotis-Bateson, E., 2004. Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* 15, 133–137.
- Pang, X.D., Guinan, J.J., 1997. Effects of stapedius-muscle contractions on the masking of auditory-nerve responses. *J. Acoust. Soc. Am.* 102 (6), 3576–3586.
- Pilz, P.K., Ostwald, J., Kreiter, A., Schnitzler, H.U., 1997. Effect of the middle ear reflex on sound transmission to the inner ear of rat. *Hear. Res.* 105, 171–182.
- Porges, S.W., 1995. Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology* 32, 301–318.
- Porges, S.W., 1997. Emotion: an evolutionary by-product of the neural regulation of the autonomic nervous system. *Ann. N.Y. Acad. Sci.* 807, 62–77.
- Porges, S.W., 2001. The polyvagal theory: phylogenetic substrates of a social nervous system. *Int. J. Psychophysiol.* 42, 123–146.
- Porges, S.W., 2007. The polyvagal perspective. *Biol. Psychol.* 74, 116–143.
- Porter, F.L., Porges, S.W., Marshall, R.E., 1988. Newborn pain cries and vagal tone: parallel changes in response to circumcision. *Child Dev.* 59, 495–505.
- Rosenzweig, M.R., Amon, A.H., 1955. Binaural interaction in the medulla of the cat. *Cell. Mol. Life Sci.* 11 (12), 498–500.
- Rowe, T., 1996. Coevolution of the mammalian middle ear and neocortex. *Science* 273, 651–654.
- Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.
- Surguladze, S.A., Calvert, G.A., Brammer, M.J., Campbell, R., Bullmore, E.T., Giampietro, V., David, A.S., 2001. Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Res.* 106, 1–14.
- Takahashi, L., 1992. Ontogeny of behavioral inhibition induced by unfamiliar adult male conspecifics in preweanling rats. *Physiol. Behav.* 52, 493–498.
- Taylor, E.W., Jordan, D., Coote, J.H., 1999. Central control of the cardiovascular and respiratory systems and their interactions in vertebrates. *Physiol. Rev.* 79, 855–916.
- Truex, R.C., Carpenter, M.B., 1969. *Human Neuroanatomy*. Williams & Wilkins, Baltimore, MD.
- Wang, Y., Hu, Y., Meng, J., Li, C., 2001. An ossified Meckel's cartilage in two Cretaceous mammals and origin of the mammalian middle ear. *Science* 294, 357–361.
- Zwislocki, J.J., 2002. Auditory system: Peripheral nonlinearity and central additivity, as revealed in the human stapedius-muscle reflex. *Proc. Natl. Acad. Sci. USA* 99, 14601–14606.