



*Annual Review of Neuroscience*

# A Theoretical Framework for Human and Nonhuman Vocal Interaction

Gregg A. Castellucci,<sup>1</sup> Frank H. Guenther,<sup>2</sup>  
and Michael A. Long<sup>1</sup>

<sup>1</sup>NYU Neuroscience Institute and Department of Otolaryngology, New York University Langone Medical Center, New York, NY, USA; email: mlong@med.nyu.edu

<sup>2</sup>Departments of Speech, Language & Hearing Sciences and Biomedical Engineering, Boston University, Boston, Massachusetts, USA

Annu. Rev. Neurosci. 2022. 45:295-316

The *Annual Review of Neuroscience* is online at [neuro.annualreviews.org](https://neuro.annualreviews.org)

<https://doi.org/10.1146/annurev-neuro-111020-094807>

Copyright © 2022 by Annual Reviews.  
All rights reserved

## Keywords

speech, animal vocalization, nonverbal communication, social interaction, language

## Abstract

Vocal communication is a critical feature of social interaction across species; however, the relation between such behavior in humans and nonhumans remains unclear. To enable comparative investigation of this topic, we review the literature pertinent to interactive language use and identify the superset of cognitive operations involved in generating communicative action. We posit these functions comprise three intersecting multistep pathways: (a) the Content Pathway, which selects the movements constituting a response; (b) the Timing Pathway, which temporally structures responses; and (c) the Affect Pathway, which modulates response parameters according to internal state. These processing streams form the basis of the Convergent Pathways for Interaction framework, which provides a conceptual model for investigating the cognitive and neural computations underlying vocal communication across species.



## Contents

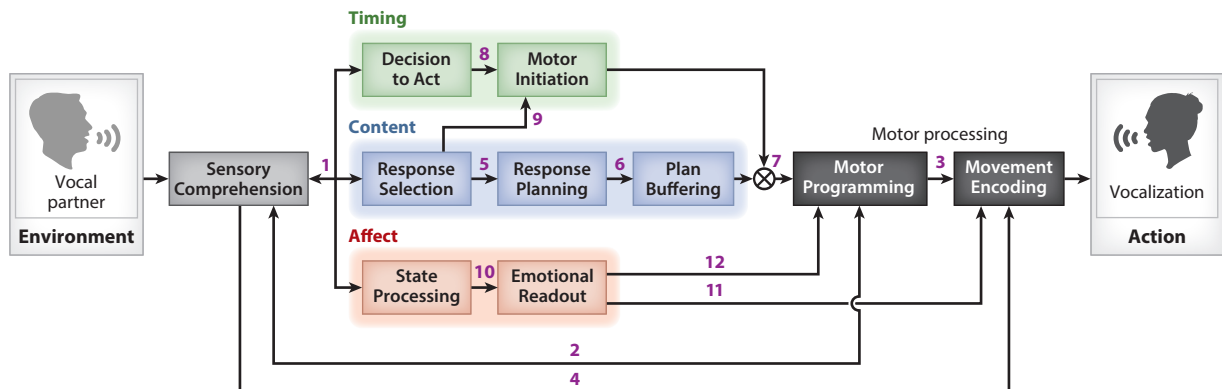
INTRODUCTION .....	296
A FRAMEWORK FOR HUMAN SPOKEN INTERACTION .....	297
SENSORY COMPREHENSION .....	298
MOTOR PROCESSING .....	298
Motor Programming .....	299
Movement Encoding .....	299
CONVERGENT COGNITIVE PATHWAYS .....	300
THE CONTENT PATHWAY .....	300
Response Selection .....	300
Response Planning .....	301
Plan Buffering .....	301
THE TIMING PATHWAY .....	302
Decision to Act .....	302
Motor Initiation .....	303
THE AFFECT PATHWAY .....	304
State Processing .....	304
Emotional Readout .....	304
APPLICATION OF THE CPI FRAMEWORK TO NONHUMAN	
VOCAL INTERACTIONS .....	305
Affective Vocal Production ( <i>Mus musculus</i> ) .....	305
Forebrain Vocal Motor Processing ( <i>Taeniopygia guttata</i> ) .....	307
Coordinated Vocal Timing ( <i>Scotinomys teguina</i> ) .....	307
Response Selection (Nonhuman Primates and Beyond) .....	308
CONCLUSION .....	309

## INTRODUCTION

In humans, conversation is central to social interaction. Although speaking with a partner may seem effortless, this behavior requires complex sensory, cognitive, and motor function, and devastating disorders can result from brain damage affecting any aspect of these processes (Pedersen et al. 2004). However, the neural and behavioral mechanisms of interactive language use remain poorly understood due to two main factors. First, most speech production studies are performed in noninteractive contexts using controlled behaviors, and relatively few investigations focus on conversation (Castellucci et al. 2022). Consequently, the extent to which established models of speech production (Guenther 2016, Levelt et al. 1999, Redford 2015, Saltzman & Munhall 1989, Turk & Shattuck-Hufnagel 2019) can be applied to naturalistic language use is unclear. Second, the methodologies available for studying the human brain are limited in their ability to uncover circuit- and cellular-level insights into language generation. Meanwhile, a wider array of techniques can be leveraged to study nonhuman species that engage in flexible interactive communication (Pika et al. 2018), which are often mediated by neural substrates displaying similarities to those relevant for human speech (Benichov et al. 2016, Miller et al. 2015, Okobi et al. 2019, Roy et al. 2016, Tagliatela et al. 2008). However, we lack a systematic method for relating vocal and social interaction across species; therefore, the specific parallels between these behaviors and human communication remain largely unknown.

To address these lacunae, we review the scientific literature concerning interactive language use to develop a conceptual framework constituting the cognitive processes required for human





**Figure 1**

Diagram of the Convergent Pathways for Interaction framework, which depicts the set of cognitive functions involved in interactive vocal behaviors. Sensory information generated by a vocal partner or the environment is processed in a Sensory Comprehension module, and communicative action is executed through a two-part Motor Processing stage. Between the sensory and motor components of the framework are three parallel pathways that encode an organism's response (i.e., the Timing, Content, and Affect Pathways). Each pathway is composed of multiple discrete processing modules, and links between modules are labeled with numbers to facilitate discussion within the text.

communication. We then demonstrate how various interactive behaviors across a range of species can be mapped onto individual components of the framework, thus providing direct connections between features of interactive speech and nonhuman vocal communication. We intend for this model to increase cross-talk between the fields of animal behavior, linguistics, and neuroscience and to enable the sophisticated technological and analytical tools of each discipline to advance our neurobiological understanding of language and social behavior.

## A FRAMEWORK FOR HUMAN SPOKEN INTERACTION

In the context of human conversation, interactants produce a spoken response in reaction to sensory cues generated by a vocal partner. Enabling this motor response is a network of multistep cognitive pathways that formulate a speaker's intended message. We schematize this network with the Convergent Pathways for Interaction (CPI) framework (**Figure 1**), a conceptual model comprised of the cognitive modules we hypothesize to be critical for vocal communication.

The CPI framework is organized around three pathways that mediate different features of an interactant's response. First, the Content Pathway encodes a response's conceptual substance as action—enabling both linguistic and nonverbal communication (Abner et al. 2015, Kendon 1997, Pezzulo et al. 2013). Second, the Timing Pathway controls response initiation, thus allowing for a high degree of coordination between speakers (Heldner & Edlund 2010, Levinson & Torreira 2015, Stivers et al. 2009) while also temporally structuring the individual elements within a response (e.g., syllables and words) (Shattuck-Hufnagel & Turk 1996). Third, the Affect Pathway modulates response production according to emotional and motivational state; in spoken language, this concept most notably manifests as affective prosody (Banse & Scherer 1996, Hammerschmidt & Jurgens 2007). In the CPI framework, these pathways converge on motor processing operations to generate communicative actions.

The bulk of our review is dedicated to explaining the structure of the CPI framework and its relevance to interactive language use. We begin with an overview of the framework's sensory and motor components, as these processes (e.g., sensory perception, articulation, motor control) have been longstanding topics of research in multiple disciplines. We then introduce the three

**Linguistic communication:** communication through language, including spoken language and sign language

**Nonverbal communication:** nonlinguistic communication used to convey information, including nonvocal behaviors (e.g., co-speech gestures, body language) and nonspeech vocalizations (e.g., crying, laughing)

**Affective prosody:** acoustic cues, such as modulations in pitch, amplitude, and voice quality, which reflect a speaker's emotional state

**Communicative action:** all vocal and nonvocal motor behaviors used by an interactant to communicate

aforementioned cognitive pathways and provide human behavioral and neuropsychological evidence for their component modules and associated interconnections. Finally, we conclude by considering specific nonhuman vocal behaviors through the lens of the CPI framework to identify potential parallels across species (Nieder & Mooney 2020, Petkov & Jarvis 2012).

## SENSORY COMPREHENSION

To generate an appropriate response during spoken conversation, an interactant must first decode the communicative actions produced by their vocal partner. In this context, communication primarily occurs through acoustic means (i.e., via speech), but information is also conveyed through other channels such as vision. For example, a speaker's co-speech gestures (e.g., pointing and other hand movements) provide additional semantic or contextual information about the spoken message (Goldin-Meadow & Sandhofer 1999) and can indicate the proper timing for their partner to respond (Holler et al. 2018). Therefore, while perception of communicative actions can occur using a single sensory modality (e.g., sign language or telephone calls), it typically relies on multiple sensory streams (Gick & Derrick 2009, McGurk & MacDonald 1976).

The neuropsychological basis for sensory processing, especially as it relates to speech perception and language comprehension, is an active area of investigation in neuroscience and linguistics. Several influential models for speech processing exist (Fowler 1986, Hickok & Poeppel 2007, Kotz & Schwartze 2010, Marslen-Wilson 1975, McClelland & Elman 1986), and we direct readers to numerous informative reviews on this topic (Diehl et al. 2004, Fowler et al. 2016, Hickok & Poeppel 2016, Samuel 2011, Skipper et al. 2017). It is important to note that the processing of language relies on multiple sensory and comprehension-related operations (Davis et al. 2007, Obleser et al. 2007, Phillips 2001), which function in a highly interactive fashion (Heald & Nusbaum 2014, McClelland et al. 2006, Samuel 2011). Therefore, we model these operations with the combined Sensory Comprehension module in the CPI framework, which subsumes the processes required to decode a vocal partner's message and relevant information from the environment. Therefore, the specific function of this module is to translate sensory stimuli into percepts in the brain of an interactant, which provide the necessary input for the downstream mechanisms enabling response generation.

Sensory comprehension is strongly influenced by top-down processes (Ganong 1980, McClelland et al. 2006). For instance, listeners can understand highly degraded speech (Miller et al. 1951, Stickney & Assmann 2001) and perceive phonemes that are not present in an acoustic signal if their occurrence is contextually predictable (Samuel 1996, Warren 1970). Such phenomena indicate that sensory processing is strongly biased by its expected conceptual content. Likewise, perception of sensory signals can be influenced by an interactant's internal state (Kelley & Schmeichel 2014, Niedenthal & Wood 2019) as well as fluctuations in attention related to their decision to execute a response (Bronkhorst 2015, Cherry 1953, Talluri et al. 2018). These modulatory effects on perception are modeled in the CPI framework with a bidirectional connection between Sensory Comprehension and the Content, Timing, and Affect pathways (see Link 1 in **Figure 1**). The specific role of the framework's individual modules in the comprehension process will be further discussed as each pathway is presented.

## MOTOR PROCESSING

During interactive language use, the motor system controls the execution of communicative actions. Motor-related neural circuitry is organized into two subnetworks related to vocalization: a direct motor control system generating articulatory movements and an upstream system enabling volitional control over vocal production (Hage & Nieder 2016, Jurgens 2009, Nieder & Mooney

2020). Humans possess both circuits, which control the articulation of innate vocalizations—such as laughs and cries—and learned speech in a largely independent fashion (Jurgens 2009). The CPI framework models this functional and anatomical division within the vocal motor system as a two-step Motor Processing operation consisting of the Motor Programming and Movement Encoding modules.

### Motor Programming

In humans, representations of the speech articulators are observed in both motor cortex and brainstem motor nuclei (Bouchard et al. 2013, Zemlin 1998), and damage to either region can result in disordered articulation (Duffy 1995). The cortical aspects of the speech motor system are specifically involved in programming articulatory movements rather than activating musculature directly (Guenther 2016). For example, damage to speech-related premotor cortex can result in a selective impairment in the motor planning and sequencing of speech movements that largely spares the execution of nonverbal vocal tract movement (Deger & Ziegler 2002, Graff-Radford et al. 2014). However, this division between the programming and execution of action is not speech specific but instead appears to be common to learned motor behaviors—a notion illustrated by the symptomatology of ideomotor apraxia (Wheaton & Hallett 2007). We model this higher-level aspect of motor execution in the CPI framework with the Motor Programming module, which translates response plans—both vocal and nonvocal—into commands within the motor system.

Input to the Motor Processing modules is provided through the cognitive systems responsible for response formulation. In addition, several lines of evidence demonstrate that the sensory system can bypass these processing steps and feed into motor planning operations directly. For example, both infants and adults can imitate spoken material without understanding its linguistic content (Kuhl & Meltzoff 1996, Vaquero et al. 2017). Furthermore, because brain lesions can disrupt speech repetition but largely spare other language functions and vice versa (Davis et al. 1978, Gorno-Tempini et al. 2008), it is likely that speech imitation and propositional speech rely on distinct neural circuitry. Likewise, speech output is unconsciously monitored to correct errors in real time (Abbs & Gracco 1984, Burnett et al. 1998, Tourville et al. 2008) and to adapt speech motor programs when repeated sensory errors are detected (Houde & Jordan 1998, Lametti et al. 2012). These motor control processes rely heavily on interactions between the sensory and motor planning operations (Golfinopoulos et al. 2011, Guenther 2016) and further reinforce the tight link of these systems during speech production. In the CPI framework, we model this sensorimotor interfacing as a bidirectional connection between the Sensory Processing and Motor Programming modules (see Link 2 in **Figure 1**).

### Movement Encoding

Prior to innervation of the musculature, the final processing step for centrally generated motor actions in vertebrates occurs in brainstem and spinal cord circuits (Jurgens 2009, Nielsen 2016). In the case of human speech production, neural signals from motor cortex converge along monosynaptic projections to vocal tract motor nuclei in the pons and medulla whose outputs drive articulatory movement (Zemlin 1998). Processing in these low-level motor circuits can operate independently from those in higher-order motor centers; for instance, motor cortical activity continues to reflect articulatory parameters in patients who are unable to speak due to brainstem stroke (Guenther et al. 2009). Conversely, brainstem vocal motor circuitry can drive the articulation of innate vocalizations in patients displaying disordered speech production due to lesions of motor cortical regions (Groswasser et al. 1988, Mao et al. 1989). In the CPI framework, we model this low-level motor processing step with the Movement Encoding module, which acts to

#### Innate vocalizations:

vocalizations whose production is genetically programmed in a species and thus do not need to be learned

#### Ideomotor apraxia:

an acquired neurological condition typified by deficits in the motor planning of learned skills and communicative gestures

#### Propositional speech:

meaningful linguistic output whose content is generated voluntarily and spontaneously



transform the motor commands output by the Motor Programming step (see Link 3 in **Figure 1**) into patterned muscle movements.

Sensory information can directly modulate subordinate motor execution processes as well as the higher-level operations modeled with the Motor Programming module. For example, humans unconsciously increase vocal intensity in the presence of background noise. This compensatory reflex—known as the Lombard effect (Lombard 1911)—is observed widely in learned and innate vocalizations across species and is thought to be mediated by brainstem mechanisms (Luo et al. 2018). The CPI framework accounts for such phenomena by allowing the Sensory Comprehension module to influence the Motor Encoding step via a direct connection (see Link 4 in **Figure 1**). Together with the feedback connection targeting the Motor Programming step (see Link 2 in **Figure 1**), these sensorimotor interfaces provide a theoretical mechanism for interactants to communicate robustly by compensating for both environmental and internal perturbations (Golfinopoulos et al. 2011, Tourville et al. 2008).

## CONVERGENT COGNITIVE PATHWAYS

Having reviewed the sensory and motor components of the CPI framework, we now turn to the pathways that interconnect these processes and formulate communicative action. In the context of human conversation, interactants produce responses to their vocal partners that display three key features: They relay conceptual information represented by articulatory movements and concurrent nonvocal actions (Content Pathway), they are timed appropriately (Timing Pathway), and their execution is modulated by affective state (Affect Pathway).

### THE CONTENT PATHWAY

Once a partner's message is processed by the Sensory Comprehension module, an interactant must select an appropriate response, prepare the actions comprising it, and potentially hold those actions in working memory. In the CPI framework, this translation from abstract meaning into intended movement proceeds via the Content Pathway, which is subdivided into three separate modules: Response Selection, Response Planning, and Plan Buffering.

#### Response Selection

To react to a vocal partner, an interactant must first select and formulate a conceptual (i.e., nonlinguistic) response (Levelt et al. 1999). Evidence for this abstract level of representation is provided by both neuropsychology and psycholinguistics. For example, lesions in lateral regions of frontal cortex can result in severe deficits to linguistic output while largely sparing the ability to comprehend language and respond nonverbally (Corina et al. 1992, Graff-Radford et al. 2014, Groswasser et al. 1988, Marshall et al. 2004). In addition, humans have been found to process events independently of their native language's word order, indicating the existence of a nonlinguistic level of representation (Papafragou & Grigoroglou 2019, Trueswell & Papafragou 2010), which furthermore appears to be accessed during speech production (Bunger et al. 2013).

In the CPI framework, we model the abstract formulation of a response's conceptual content with the Response Selection module. This module receives input from the Sensory Comprehension step and also provides feedback to it (see Link 1 in **Figure 1**), consistent with previous research suggesting that action selection is highly interconnected with sensory perception via top-down and bottom-up mechanisms (Cisek & Kalaska 2005, 2010). In the context of interactive language use, we hypothesize that an interactant collects information from their partner's communicative actions via Sensory Comprehension until the most appropriate response is selected for generation

via separate downstream planning operations (de Ruyter 2000, Fuchs et al. 2013, Indefrey 2011, Lee et al. 2013, Meyer 1996, Smith & Wheeldon 1999). At the same time, an interactant's expectations regarding the content of an appropriate response guides sensory processing via predictive mechanisms (Ganong 1980, McClelland et al. 2006, Miller et al. 1951, Samuel 1996, Stickney & Assmann 2001, Warren 1970).

### Response Planning

Following the selection of a response, an interactant must encode its conceptual content into action. During conversation, these actions primarily consist of articulatory vocal tract movements but also include a variety of associated nonvocal actions (Abner et al. 2015, Goldin-Meadow & Sandhofer 1999, Kendon 1997, Pezzulo et al. 2013). For example, speakers use nonverbal communication such as co-speech gestures (e.g., head nods, manual signs, and facial expressions) alongside their spoken output and can generate responses that are completely nonvocal (Kogure 2007). To enable this flexible usage of multimodal communicative actions, we posit a general-purpose Response Planning module in the CPI framework, which computes action plans from the conceptual-level output generated by the Response Selection step (see Link 5 in **Figure 1**).

Neuropsychological evidence suggests that the Response Planning module consists of distinct subprocesses. Specifically, language planning is at least partly compartmentalized from general premotor mechanisms, as demonstrated in patients with expressive aphasia (Corina et al. 1992, Graff-Radford et al. 2014, Groswasser et al. 1988, Marshall et al. 2004). Furthermore, different linguistic features are likely planned via distinct operations (Guenther 2016, Levelt et al. 1999). For example, both speech and sign language exhibit linguistic structure and linguistic prosody (Dachkovsky & Sandler 2009, Shattuck-Hufnagel & Turk 1996); disorders such as fluent aphasia can selectively affect a patient's ability to generate linguistic content while leaving their prosody relatively unaffected (Seddoh 2004, Van Lancker Sidtis et al. 2010), thus suggesting these features are planned separately. Furthermore, linguistic structure is thought to be generated via serially organized processes encoding semantic, syntactic, and phonological structure (Fuchs et al. 2013, Indefrey 2011, Lee et al. 2013, Meyer 1996, Smith & Wheeldon 1999). The number and identity of language planning processes remains an active area of investigation in psycholinguistics and neurolinguistics, and we refer readers to several influential models and reviews on this topic (Guenther 2016, Levelt et al. 1999, Redford 2015, Saltzman & Munhall 1989, Turk & Shattuck-Hufnagel 2019). However, for the purpose of the CPI framework, we hypothesize that these speech planning functions as well as nonvocal planning operations are performed by the Response Planning module, which generates multimodal action plans that are then translated into low-level motor commands via Motor Programming.

### Plan Buffering

Once the actions comprising a response have been planned, they can be executed as movement. However, an interactant may be required to temporarily buffer their planned response in order to properly time its execution (Levinson & Torreira 2015). For example, gaps between speakers engaged in conversational turn-taking are typically less than 200 ms in duration (Heldner & Edlund 2010, Stivers et al. 2009), which is considerably shorter than reaction times typically observed in single-word production tasks (Bates et al. 2003). To achieve this degree of temporal coordination, interactants must plan their responses prior to the completion of their partner's turn and store this plan until the appropriate time to respond (Bögels et al. 2015, Levinson & Torreira 2015). In the CPI framework, this working memory process is modeled with the Plan Buffering module, which is immediately downstream of Response Planning (see Link 6 in

---

**Expressive aphasia:** an acquired neurological condition characterized by severely disrupted spoken and sign language but largely preserved nonlinguistic action generation

**Linguistic structure:** grammatical features of an utterance's words, including semantics (meaning), syntax (hierarchically organized word order), morphology (word-internal structure), and phonology (abstract sound structure)

**Linguistic prosody:** features of an utterance that relate to its linguistic meaning, including amplitude, pitch, and timing (Timing Pathway)

**Fluent aphasia:** an acquired neurological condition typified by the production of fluent sounding speech which lacks proper linguistic structure

**Semantics:** subfield of linguistics dedicated to the study of meaning

**Syntax:** subfield of linguistics dedicated to the study of the hierarchical rules governing word order

**Phonology:** subfield of linguistics dedicated to the study of language's abstract sound structure rather than its articulatory realization

---



**Turn-taking:** the behavior exhibited during conversation in which interactants attempt to respond to vocal partners quickly while avoiding overlap

**Turn-end prediction:** a predictive process used by human interactants to anticipate the end of their partner's spoken turn

**Figure 1**) and acts to temporarily hold action plans before Motor Processing commences. As we discuss below, the Plan Buffering module is also necessary following the initiation of a spoken response, where it acts to maintain the complete outgoing message while the motor system serially reads out its individual components (e.g., syllables within a word) (Guenther 2016). The Plan Buffering module in the CPI framework therefore resembles working memory steps in other models (Baddeley 1992, Bohland et al. 2010); however, we hypothesize this module also stores nonvocal communicative actions (e.g., co-speech gestures) and therefore subsumes other domain-specific buffers (Poletti et al. 2008, Smyth & Pendleton 1989).

## THE TIMING PATHWAY

During interactive language use, both the timing of response initiation (Levinson & Torreira 2015) and the internal temporal structure of a response (Shattuck-Hufnagel & Turk 1996) are used for communication (Breen et al. 2010, Mücke & Grice 2014, Pomerantz & Heritage 2013, Snedeker & Trueswell 2003). However, speech timing is controlled by mechanisms that appear largely distinct from the processes encoding its linguistic structure. For example, prelinguistic children can time vocal responses with near adult-like latencies (Jasnow & Feldstein 1986), and aphasics with severely disordered speech production can use laughter as responses during conversational turn-taking (Norris & Drummond 1998, Rohrer et al. 2009). These results indicate that linguistic content is not required for temporally coordinated vocal interactions. Furthermore, cerebellar lesions can result in speech timing abnormalities but largely spare other features of linguistic structure (Ackermann et al. 1992), thus demonstrating that neurological damage can selectively disrupt the temporal structure of speech. Similarly, while the perturbation of speech-related motor cortex in neurosurgical patients via mild focal cooling results in articulatory disruptions, cooling upstream frontal areas instead modulates speech rate (Long et al. 2016). To account for these observations, the CPI framework includes a Timing Pathway separate from the Response Planning operations of the Content Pathway. The Timing Pathway is divided into two modules—Decision to Act and Motor Initiation—which are described below.

### Decision to Act

During rapid conversational turn-taking, a high degree of temporal coordination across interactants is achieved via turn-end prediction, which relies on several linguistic features (de Ruiter et al. 2006, Holler et al. 2018). For example, articulation becomes systematically longer in duration as a speaker nears the end of a fluent utterance (Wightman et al. 1992), and interactants monitor their partner's speech for this cue to precisely time their responses (Bögels & Torreira 2015). Similarly, speakers can rapidly suppress ongoing speech (Tilsen 2011), and they can use this ability to prevent overlapping talk when faced with an interruption (Schegloff 2000).

In the CPI framework, the processes important for deciding when to initiate a response—such as turn-end prediction—are modeled with the Decision to Act module, which we posit to also include general-purpose decision-making operations (Rangel et al. 2008, Rilling & Sanfey 2011) in addition to language-specific ones. This hypothesis is supported by the existence of disordered turn-taking behavior in neurodevelopmental conditions typified by increased impulsivity, such as attention deficit hyperactivity disorder (Giddan 1991). Furthermore, as decision-making and predictive mechanisms are well known to modulate perceptual processes (Ganong 1980, Samuel 1996, Sohoglu et al. 2012), we model the connection between the Decision to Act and Sensory Comprehension modules as bidirectional (see Link 1 in **Figure 1**). We likewise postulate a bidirectional connection between the Decision to Act and Response Selection modules (see Link 1



in **Figure 1**) since active control of response onset can carry communicative value—for example, a speaker may delay response onset to signal that their response might be met with disapproval (Pomerantz & Heritage 2013).

### Motor Initiation

After deciding to respond to a partner, an interactant must rapidly execute their planned response to achieve the precise interspeaker coordination observed in human conversation (Levinson & Torreira 2015). Action initiation is thought to be controlled by cognitive operations that are separate from those underlying planning (Klapp & Maslovat 2020); likewise, the initiation of motor plans has been shown to rely on distinct neural circuitry (Bohland et al. 2010, Dacre et al. 2021, Dick et al. 2019, Guenther 2016, Zimmik et al. 2019). In the CPI framework, we model these processes with the Motor Initiation module, which we hypothesize acts to gate the release of response plans stored in Plan Buffering to the downstream Motor Programming module (see the  $\otimes$  in Link 7 in **Figure 1**). To accurately time response onset, we therefore posit that the Decision to Act module provides direct input to the Motor Initiation module (see Link 8 in **Figure 1**).

During conversation, spoken responses typically consist of connected speech (Farnetani & Recasens 1997), and interactants can modulate the temporal structure of these multiword responses to communicate (Breen et al. 2010, Mücke & Grice 2014). For example, speakers can produce the same series of words at a variety of rates (Goldstein et al. 2007) and can modify pause and word durations for emphasis or to signal differences in linguistic structure (Snedeker & Trueswell 2003). These behavioral features suggest that the mechanisms controlling speech timing receive input from the processes encoding the content of a response. In the CPI framework, we model this interaction between timing and planning functions as a direct connection between the Response Selection module in the Content Pathway and Motor Initiation (see Link 9 in **Figure 1**). This link therefore allows for the conceptual content of an interactant's response to be signaled through a series of actions as well as its timing. Specifically, connected speech consists of multiple speech planning units (e.g., syllables) (Guenther 2016). During articulation, these planning units must be buffered as the motor system sequentially executes the vocal tract movements corresponding to each unit. We hypothesize the Motor Initiation module releases individual planning units stored in Plan Buffering to the Motor Programming step and thereby acts to encode the timing of each unit's articulation. Through control of this mechanism by the Response Selection module, our model therefore provides a theoretical account for the generation of linguistic prosody as it relates to speech timing (Turk & Shattuck-Hufnagel 2019).

In the CPI framework, Motor Initiation operates on both vocal and nonvocal action plans in the Plan Buffering module; therefore, our framework predicts a multimodal initiation process for all communicative action plans. This hypothesis is supported by several experimental and clinical studies that have found a tight link between the timing of speech and nonspeech movements. For example, perturbation of cortical loci important for the initiation of speech (Ferpozzi et al. 2018) via direct electrical stimulation can result in the inhibition or slowing of both articulation and hand movements (Breshears et al. 2019). Likewise, lesions of the supplemental motor area may result in slowed speech and nonvocal movements (Laplaine et al. 1977). Finally, the kinematic properties of natural speech usage also support a domain-general initiation mechanism, as the execution of nonvocal movements has been found to be temporally coordinated to articulation and linguistic prosody in many contexts (Krivokapic et al. 2017, Mendoza-Denton & Jannedy 2011, Parrell et al. 2014, Shattuck-Hufnagel & Ren 2018). For example, prolongations in word duration that signal a prominent word are also realized on co-occurring manual gestures (Krivokapic et al. 2017).

---

**Connected speech:**  
spoken series  
consisting of multiple  
words

**Planning unit:**  
linguistic units used  
during speech  
planning; speech  
production models  
posit different  
planning units such as  
syllables, words, and  
fluent phrases

---



## THE AFFECT PATHWAY

During speech communication, the affective and physiological state of an interactant can strongly influence their behavior. For instance, a speaker's internal state—which is informed by the content of their partner's message—can directly affect their choice of response (Saslow et al. 2014). Likewise, a speaker's emotional state is signaled by the affective prosody of their spoken responses (Banse & Scherer 1996, Hammerschmidt & Jurgens 2007). The processing of emotion and arousal occurs in limbic brain regions (LeDoux 2000) that are distinct from those underlying the volitional motor production (Hage & Nieder 2016, Jurgens 2009, Mao et al. 1989, Pichon & Kell 2013). For instance, damage to language-critical neural circuitry can severely disrupt speech production but largely spare the ability to communicate affective state using innate vocalizations (Fruhholz et al. 2015, Groswasser et al. 1988). This separation between the emotional and linguistic aspects of human spoken communication is modeled in the CPI framework with the Affect Pathway. During interactive language use, this distinct pathway influences processing in the Content and Timing pathways via its State Processing and Emotional Readout modules.

### State Processing

The internal state of an organism modulates a wide array of processes related to sensory perception (Kelley & Schmeichel 2014, Niedenthal & Wood 2019), motor behavior (Barrett & Paus 2002, Reilly et al. 1992), and decision-making (Lerner et al. 2015). In the context of human vocal communication, a speaker's affective and physiological state can therefore be expected to influence how they perceive their partner's communicative actions as well as whether and how they respond. While it remains unclear how internal state specifically affects each of these processes individually, human interaction is known to be heavily modulated by factors related to emotion (Lopes et al. 2005, Weightman et al. 2014). Likewise, previous research has demonstrated that increased stress levels result in higher rates of disfluencies in stutterers (Blood et al. 1997) and reduced linguistic complexity in typical speakers (Saslow et al. 2014). These results indicate that language generation itself can be specifically influenced by a speaker's affective state.

To account for the effects of emotional and physiological state on interactive behavior, the CPI framework models the processes involved in monitoring and regulating internal state with the State Processing module. While physiological state likely affects the brain and behavior globally (e.g., stress can affect working memory capacity) (Sorg & Whitney 1992), we explicitly model the aspects of this modulation that are particularly important for communicative action generation with connections from the State Processing module (see Link 1 in **Figure 1**) to (a) Sensory Comprehension, allowing for the communicative message of a vocal partner(s) to affect an interactant's state; (b) Response Selection, enabling an interactant's state to influence the content of their response; and (c) Decision to Act, which provides a mechanism for emotional state to influence whether and when an interactant initiates their response. Finally, this four-way link is hypothesized to be bidirectional, allowing for internal state to shape comprehension processes (Kelley & Schmeichel 2014, Niedenthal & Wood 2019) and enabling language generation itself to affect an interactant's own internal state (Lerner et al. 2015).

### Emotional Readout

An interactant's affective status can result in communicative action via two general mechanisms: (a) production of discrete behaviors and (b) modification of ongoing movement. In the first case, specific emotional or physiological states are biologically preprogrammed to result in innate vocalizations such as a groan (Hage & Nieder 2016, Jurgens 2009). These behaviors are generated through a circuit consisting of phylogenetically ancient limbic structures that directly project to

brainstem motor nuclei (Jurgens 2009, Nieder & Mooney 2020). However, this system also mediates the execution of innate nonvocal behaviors such as defensive posturing (Deng et al. 2016, Evans et al. 2018), thus indicating that it is not specific to vocalization. To model the production of these affective vocalizations and motor routines, the CPI framework allows for specific internal states to be read out as distinct behavioral units via a direct link between the Emotional Readout and Movement Encoding modules (see Link 11 in **Figure 1**).

While innate vocalizations can be used communicatively during interaction (Vettin & Todt 2004), the more prevalent reflection of internal state in human communication is affective prosody, which is realized as modulations in prosodic features such as pitch, duration, and voice quality according to a speaker's emotional status (Banse & Scherer 1996, Barrett & Paus 2002, Hammerschmidt & Jurgens 2007). Although superficially related to linguistic prosody, the generation of affective prosody appears independent of operations encoding the linguistic structure of speech. For example, akinetic mutism can resolve over longer timescales such that patients recover their ability to speak, albeit with disordered affective prosody (Jurgens & von Cramon 1982). Similarly, the speech of schizophrenic individuals has been observed to display abnormal affective prosody, while linguistic prosody is generally unaffected (Lucarini et al. 2020, Murphy & Cutting 1990).

While most studies of affective prosody focus on its vocal properties, a speaker's internal state is also signaled during natural language production by nonverbal movements such as facial expressions (Attardo et al. 2011, Gironzetti et al. 2016). Furthermore, a speaker's affective status can also influence the execution of ongoing nonvocal actions. For example, co-speech gestures of the head display increased velocity when a speaker is experiencing stress (Giannakakis et al. 2018). Likewise, affective prosody is realized in sign language through modulations of linguistically meaningful movements of the head and face (Hietanen et al. 2004, Reilly et al. 1992). Such phenomena demonstrate that the emotional processes that generate affective prosody influence both vocal and nonvocal aspects of communicative motor behavior. In the CPI framework, affective state exerts modulatory effects on all communicative actions through the Emotional Readout module, which influences the generation of motor commands by the Motor Programming step (see Link 12 in **Figure 1**).

## APPLICATION OF THE CPI FRAMEWORK TO NONHUMAN VOCAL INTERACTIONS

In the CPI framework, the generation of communicative action is schematized as a set of modules constituting distinct yet interacting pathways. These modules and their connections were motivated using behavioral and neuropsychological features of human linguistic communication, but this model can also be used to contextualize the vocal behaviors of other organisms (**Figure 2**). The CPI framework therefore provides a means for systematically identifying analogous cognitive processes across species, which may in turn rely on shared neural mechanisms. To demonstrate this important feature of the CPI framework, we now discuss how various communicative behaviors across species can be related to specific model components and pathways.

### Affective Vocal Production (*Mus musculus*)

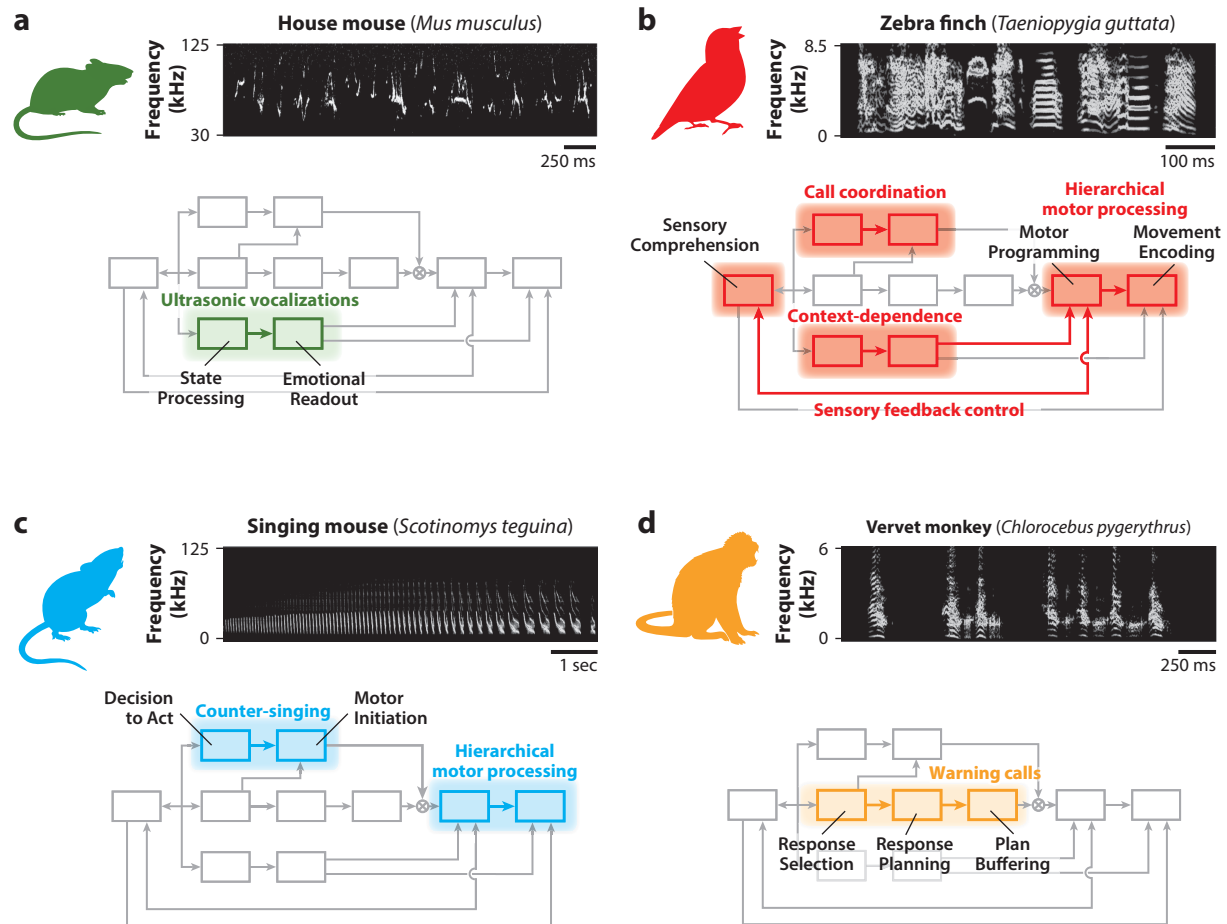
The production of innate vocalizations can be automatically driven by factors related to an organism's affective state (Hage & Nieder 2016, Jurgens 2009). These affective vocalizations are widespread among vertebrates and have been well studied in laboratory mice (*Mus musculus*), which generate a variety of ultrasonic and audible vocalizations (Portfors 2007). While mice have not been observed to use their calls in structured vocal interactions (e.g., turn-taking),

---

**Akinetic mutism:** an acquired neurological condition characterized by a dramatic reduction in the motivation to speak and generate movements

---





**Figure 2**

Using the Convergent Pathways for Interaction framework to investigate vocal communication in nonhuman species. Behaviors across four organisms are highlighted: (a) house mouse, (b) zebra finch, (c) singing mouse, and (d) vervet monkey. An example sonogram depicting the vocal behavior of each species is shown at the top of each panel. Below the sonograms, the aspects of the framework we hypothesize to be pertinent to each vocal behavior are highlighted. The framework shown below each sonogram is illustrated in greater detail in **Figure 1**. The audio data in panel *a* were collected by Gregg Castellucci and in panels *b* and *c* by Michael Long. The finch and singing mouse images were adapted from artwork created by Julia Kuhl. The vervet monkey image and accompanying sonogram in panel *d* were adapted with permission; copyright Julia Fischer, German Primate Center.

these vocalizations are used communicatively and elicit complex behavioral responses from conspecifics (Ehret 1992, Ehret & Bernecker 1986, Hammerschmidt et al. 2009, Liu et al. 2013). Mice therefore represent a valuable species for the study of innate calls reflecting emotional state, which are modeled in the CPI framework as arising through the Affect Pathway.

Recent studies have demonstrated that stimulation of vocalization-related neurons in the periaqueductal gray (PAG) of male mice results in the production of stereotyped courtship-related calls (Chen et al. 2021, Tschida et al. 2019), while stimulation of upstream centers in the lateral preoptic area leads to vocalizations that vary naturalistically in amplitude and duration via modulation of PAG dynamics (Chen et al. 2021). Likewise, stimulation of amygdalar regions projecting

to PAG can suppress vocal output (Michael et al. 2020). In the CPI framework, modulation of PAG by these upstream regions may reflect State Processing inputs to the Emotional Readout module; as PAG in turn projects to the brainstem pattern generators that drive vocal production (Jurgens 2009), we likewise hypothesize that this circuit corresponds to the direct input from the Emotional Readout module to Movement Encoding (**Figure 2a**). Further research is needed to determine whether these connections represent universal mechanisms enabling an organism's internal state to influence the production of preprogrammed behaviors. However, we anticipate such investigations to be broadly relevant given that affective state drives and modulates innate vocal output in a range of organisms, including humans (Liao et al. 2018, Stewart et al. 2013).

---

**Antiphonal calling:** a behavior in which multiple organisms produce alternating vocalizations, usually over a distance

---

### Forebrain Vocal Motor Processing (*Taeniopygia guttata*)

Human speech production relies on a two-step motor system consisting of a cortical network converging onto brainstem motor regions (Hage & Nieder 2016, Jurgens 2009). Unlike mice (Hammerschmidt et al. 2015, Nieder & Mooney 2020), humans and songbirds—such as the zebra finch (*Taeniopygia guttata*)—exhibit this hierarchical relationship. In humans and finches, lesions of this vocal forebrain system lead to selective dysfunction in the production of speech and courtship song, respectively, while the ability to produce other innate vocalizations is largely unaffected (Groswasser et al. 1988, Mao et al. 1989, Simpson & Vicario 1990). Furthermore, experimental perturbations of these forebrain regions with electrical stimulation (Penfield & Boldrey 1937, Vu et al. 1994) and focal cooling (Long & Fee 2008, Long et al. 2016) can result in similar behavioral outcomes during vocalization. In the CPI framework, this forebrain control over speech and song is hypothesized to correspond to the Motor Programming module, which outputs motor commands to the Motor Encoding module to drive articulatory musculature.

The common organization of the zebra finch and human vocal motor pathway is further reflected by several intriguing parallels. For example, zebra finch courtship song—like human speech—is learned primarily through vocal imitation (Tchernichovski et al. 2001), which is enabled in the CPI framework via direct input from Sensory Comprehension to the Motor Programming module (**Figure 2b**). The connection between these modules also provides a theoretical account for the ability of finches to execute compensatory articulatory movements when sensory feedback is perturbed (Leonardo & Konishi 1999, Sober & Brainard 2009, Tumer & Brainard 2007), which we hypothesize to underlie similar speech motor control processes in humans (Abbs & Gracco 1984, Burnett et al. 1998, Houde & Jordan 1998, Lametti et al. 2012, Tourville et al. 2008). Furthermore, involvement of the Motor Programming module in birdsong production allows for vocal output to be modulated by the animal's internal state in a manner parallel to the generation of affective prosody (**Figure 2b**). Consistent with this prediction, finches exhibit context-dependent changes in singing behavior; for example, songs directed to females differ in acoustic structure from those produced in isolation (Sossinka & Böhner 1980). Taken together, these features of the zebra finch song system reinforce its utility for examining how forebrain motor circuits control vocal behavior and are modulated by sensory information and affective processes.

### Coordinated Vocal Timing (*Scotinomys teguina*)

A key feature of human vocal interaction is rapid turn-taking between speakers (Stivers et al. 2009). However, such vocal turn-taking is not exhibited by house mice or observed during zebra finch courtship song production, although it is common in various forms across other species (Pika et al. 2018). For example, a form of antiphonal calling in the neotropical singing mouse



(*Scotinomys teguina*) called counter-singing displays a high degree of temporal coordination between vocal partners (Okobi et al. 2019). Specifically, singing mice produce extended advertisement calls (~10 s in duration) that can elicit a vocal response from conspecific males within hundreds of milliseconds. This vocal behavior is mediated by forebrain motor circuits, and an animal's ability to rapidly respond to a partner is abolished when this system is inactivated. However, perturbations of the forebrain vocal regions do not eliminate spontaneous calling or disrupt their articulation, indicating that these circuits provide a higher-level motor signal controlling response timing but brainstem motor circuits encode the vocal tract movements required for advertisement call production (Okobi et al. 2019).

Forebrain control of vocalization timing in singing mice can be modeled with the Motor Programming module, which provides commands to the Motor Encoding step (i.e., brainstem structures) to drive call articulation. According to the CPI framework, engagement of the Motor Programming module enables the Timing Pathway to provide input to the vocal motor system (**Figure 2c**), thus allowing for response timing to be controlled by decision-making operations (i.e., Decision to Act). For example, during counter-singing, the singing mouse must recognize and monitor a partner's call for acoustic features (e.g., elapsed time, song termination) relevant for response initiation. Once these cues have been detected, an affirmative Decision to Act can be made and Motor Initiation of downstream Motor Processing steps can commence. Similarly, recent work examining the usage of short, largely innate calls in the zebra finch has demonstrated that the forebrain song system is also required to finely time the initiation of these calls during interaction (Benichov et al. 2016). Therefore, Timing Pathway control over innate vocal behavior appears to be present in zebra finches (**Figure 2b**) and may represent a widespread phenomenon for precise response timing during interaction.

### Response Selection (Nonhuman Primates and Beyond)

Human language is symbolic, with individual words linked to conceptual information through experience (Woodward & Markman 1998). The extent to which singing mice and zebra finches learn to associate their vocalizations to distinct concepts is unknown; however, a limited version of this phenomenon is observed in nonhuman primates. Specifically, some primate species possess vocal systems acquired through usage learning—where animals learn to produce a call in a specific context (Gultekin & Hage 2017, Seyfarth et al. 1980, Wedgell et al. 2019). While words are further linked by semantic and syntactic relations in human language, these indexical calls may represent the precursors of symbolic vocal use (Nieder 2009). For example, vervet monkeys produce distinct warning calls for different types of predators (Seyfarth et al. 1980); while the warning calls themselves are preprogrammed, juvenile vervets learn to produce them in response to the appropriate stimulus (Seyfarth & Cheney 1986). This behavior suggests that vervets possess distinct conceptual representations of predator classes and can select the warning call corresponding to each class. The neurophysiology underlying vocal usage learning and the generation of calls acquired through this process is poorly understood; however, juvenile macaques have been trained to produce their innate vocalizations in response to an arbitrary cue (Coude et al. 2011, Gavrillov et al. 2017, Hage & Nieder 2013, Hage et al. 2013). In this behavior, premotor activity is observed prior to the production of cued calls but not before the same calls are produced spontaneously (Coude et al. 2011, Gavrillov et al. 2017), suggesting that cortical input mediates the volitional control of call production (Hage & Nieder 2016). Similar premotor activity has been observed in marmosets during antiphonal calling (Miller et al. 2009b, 2015; Roy et al. 2016), indicating that this forebrain pathway is also important for vocal control during interaction. In the CPI framework, volitional selection and regulation of specific vocal outputs is hypothesized to occur through



the Response Selection module; therefore, nonhuman primates may provide a valuable model to examine the neural correlates of the Content Pathway (**Figure 2d**), including during vocal interaction (Lemasson et al. 2011, Miller et al. 2009a).

Humans can learn new words and speech sounds even as adults (Brybaert et al. 2016), however vocal usage in nonhuman primates is limited to a species-specific repertoire of innate calls (Hage & Nieder 2016). Consequently, additional model organisms are required to study how conceptual information can be linked to novel vocal elements. For example, several avian species are capable of both vocal usage learning and imitation, including corvids (Brecht et al. 2019) and parrots (Osmanski et al. 2021, Pepperberg 2002). Parrots in particular can associate numerous human words to objects (Pepperberg 2002), demonstrating their ability to signal an object's mental representation with learned vocal output. The neural mechanisms enabling this behavior remain unknown, but investigation of this topic may yield important insights regarding how conceptual processes related to meaning (i.e., Response Selection) interact with forebrain motor processes (i.e., Motor Programming)—which is crucial for human language.

## CONCLUSION

In this review, we synthesize research from the fields of linguistics, neuropsychology, animal behavior, and systems neuroscience with the goal of uncovering a species-general cognitive architecture underlying communicative action. The CPI framework represents the first step in this effort, and we anticipate that future findings will update this model by providing additional detail to specific modules or uncovering modules and connections that were not originally included. Overall, the development of the CPI framework is intended to foster communication across disciplines by facilitating comparative investigations of communicative behaviors and their neural dynamics. By integrating studies from a variety of model organisms, we can achieve a more mechanistic understanding of the circuit elements (e.g., cell types, synaptic connectivity) underlying vocal interaction that is presently unavailable for the human brain. The CPI framework therefore provides an essential device for organizing and contextualizing these efforts.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We thank Arkarup Banerjee, Jelena Krivokapić, and members of the Long laboratory for comments on earlier versions of this manuscript. This review was supported by R01 DC019354 (M.A.L. and F.H.G.) and the Simons Collaboration on the Global Brain (M.A.L.).

## LITERATURE CITED

- Abbs JH, Gracco VL. 1984. Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. *J. Neurophysiol.* 51:705–23
- Abner N, Cooperrider K, Goldin-Meadow S. 2015. Gesture for linguists: a handy primer. *Lang. Linguist. Compass* 9:437–51
- Ackermann H, Vogel M, Petersen D, Poremba M. 1992. Speech deficits in ischaemic cerebellar lesions. *J. Neurol.* 239:223–27
- Attardo S, Pickering L, Baker AA. 2011. Prosodic and multimodal markers of humor in conversation. *Pragmat. Cogn.* 19:224–47
- Baddeley A. 1992. Working memory. *Science* 255:556–59



- Banse R, Scherer KR. 1996. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70:614–36
- Barrett J, Paus T. 2002. Affect-induced changes in speech production. *Exp. Brain Res.* 146:531–37
- Bates E, D'Amico S, Jacobsen T, Szekely A, Andonova E, et al. 2003. Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10:344–80
- Benichov JI, Benezra SE, Vallentin D, Globerson E, Long MA, Tchernichovski O. 2016. The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.* 26:309–18
- Blood IM, Wertz H, Blood GW, Bennett S, Simpson KC. 1997. The effects of life stressors and daily stressors on stuttering. *J. Speech Lang. Hear. Res.* 40:134–43
- Bögels S, Magyari L, Levinson SC. 2015. Neural signatures of response planning occur midway through an incoming question in conversation. *Sci. Rep.* 5:12881
- Bögels S, Torreira F. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *J. Phon.* 52:46–57
- Bohland JW, Bullock D, Guenther FH. 2010. Neural representations and mechanisms for the performance of simple speech sequences. *J. Cogn. Neurosci.* 22:1504–29
- Bouchard KE, Mesgarani N, Johnson K, Chang EF. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–32
- Brecht KF, Hage SR, Gavrilov N, Nieder A. 2019. Volitional control of vocalizations in corvid songbirds. *PLOS Biol.* 17:e3000375
- Breen M, Fedorenko E, Wagner M, Gibson E. 2010. Acoustic correlates of information structure. *Lang. Cognit. Process.* 25:1044–98
- Breshears JD, Southwell DG, Chang EF. 2019. Inhibition of manual movements at speech arrest sites in the posterior inferior frontal lobe. *J. Neurosurg.* 85:E496–501
- Bronkhorst AW. 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* 77:1465–87
- Brybaert M, Stevens M, Mander P, Keuleers E. 2016. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Front. Psychol.* 7:1116
- Bunger A, Papafragou A, Trueswell JC. 2013. Event structure influences language production: evidence from structural priming in motion event description. *J. Mem. Lang.* 69:299–323
- Burnett TA, Freedland MB, Larson CR, Hain TC. 1998. Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* 103:3153–61
- Castellucci GA, Kovach CK, Howard MA, Greenlee JDW, Long MA. 2022. A speech planning network for interactive language use. *Nature* 602(7895):117–22
- Chen J, Markowitz JE, Lilascharoen V, Taylor S, Sheurpukdi P, et al. 2021. Flexible scaling and persistence of social vocal communication. *Nature* 593:108–13
- Cherry EC. 1953. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25:975–79
- Cisek P, Kalaska JF. 2005. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45:801–14
- Cisek P, Kalaska JF. 2010. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33:269–98
- Corina DP, Poizner H, Bellugi U, Feinberg T, Dowd D, O'Grady-Batch L. 1992. Dissociation between linguistic and nonlinguistic gestural systems: a case for compositionality. *Brain Lang.* 43:414–47
- Coode G, Ferrari PF, Roda F, Maranesi M, Borelli E, et al. 2011. Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLOS ONE* 6:e26822
- Dachkovsky S, Sandler W. 2009. Visual intonation in the prosody of a sign language. *Lang. Speech* 52:287–314
- Dacre J, Colligan M, Clarke T, Ammer JJ, Schiemann J, et al. 2021. A cerebellar-thalamocortical pathway drives behavioral context-dependent movement initiation. *Neuron* 109:2326–38.e8
- Davis L, Foldi NS, Gardner H, Zurif EB. 1978. Repetition in the transcortical aphasias. *Brain Lang.* 6:226–38
- Davis MH, Coleman MR, Absalom AR, Rodd JM, Johnsrude IS, et al. 2007. Dissociating speech perception and comprehension at reduced levels of awareness. *PNAS* 104:16032–37





- de Ruiter JP. 2000. The production of gesture and speech. In *Language and Gesture*, ed. D McNeill, pp. 284–311. Cambridge, UK: Cambridge Univ. Press
- de Ruiter JP, Mitterer H, Enfield NJ. 2006. Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82:515–35
- Deger K, Ziegler W. 2002. Speech motor programming in apraxia of speech. *J. Phon.* 30:321–35
- Deng H, Xiao X, Wang Z. 2016. Periaqueductal gray neuronal activities underlie different aspects of defensive behaviors. *J. Neurosci.* 36:7580–88
- Dick AS, Garic D, Graziano P, Tremblay P. 2019. The frontal aslant tract (FAT) and its role in speech, language and executive function. *Cortex* 111:148–63
- Diehl RL, Lotto AJ, Holt LL. 2004. Speech perception. *Annu. Rev. Psychol.* 55:149–79
- Duffy JR. 1995. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, MO: Mosby
- Ehret G. 1992. Categorical perception of mouse-pup ultrasounds in the temporal domain. *Anim. Behav.* 43:409–16
- Ehret G, Bernecker C. 1986. Low-frequency sound communication by mouse pups (*Mus musculus*): Wriggling calls release maternal behaviour. *Anim. Behav.* 34:821–30
- Evans DA, Stempel AV, Vale R, Ruehle S, Lefler Y, Branco T. 2018. A synaptic threshold mechanism for computing escape decisions. *Nature* 558:590–94
- Farnetani E, Recasens D. 1997. Coarticulation and connected speech processes. In *The Handbook of Phonetic Sciences*, ed. WJ Hardcastle, J Laver, FE Gibbon, pp. 316–52. West Sussex, UK: Blackwell Publ.
- Ferpozzi V, Fornia L, Montagna M, Siodambro C, Castellano A, et al. 2018. Broca's area as a pre-articulatory phonetic encoder: gating the motor program. *Front. Hum. Neurosci.* 12:64
- Fowler CA. 1986. An event approach to the study of speech perception from a direct realist perspective. *J. Phon.* 14:3–28
- Fowler CA, Shankweiler D, Studdert-Kennedy M. 2016. "Perception of the speech code" revisited: Speech is alphabetic after all. *Psychol. Rev.* 123:125–50
- Fruhholz S, Klaas HS, Patel S, Grandjean D. 2015. Talking in fury: the cortico-subcortical network underlying angry vocalizations. *Cereb. Cortex* 25:2752–62
- Fuchs S, Petrone C, Krivokapić J, Hoole P. 2013. Acoustic and respiratory evidence for utterance planning in German. *J. Phonetics* 41:29–47
- Ganong WF 3rd. 1980. Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6:110–25
- Gavrilov N, Hage SR, Nieder A. 2017. Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell Rep.* 21:2393–406
- Giannakakis G, Manousos D, Simos P, Tsiknakis M. 2018. Head movements in context of speech during stress induction. In *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 710–14. Piscataway, NJ: IEEE
- Gick B, Derrick D. 2009. Aero-tactile integration in speech perception. *Nature* 462:502–4
- Giddan JJ. 1991. Communication issues in attention-deficit hyperactivity disorder. *Child Psychiatry Hum. Dev.* 22:45–51
- Gironzetti E, Pickering L, Huang M, Zhang Y, Menjo S, Attardo S. 2016. Smiling synchronicity and gaze patterns in dyadic humorous conversations. *Humor* 29:301–24
- Goldin-Meadow S, Sandhofer CM. 1999. Gestures convey substantive information about a child's thoughts to ordinary listeners. *Dev. Sci.* 2:67–74
- Goldstein L, Pouplier M, Chen L, Saltzman E, Byrd D. 2007. Dynamic action units slip in speech production errors. *Cognition* 103:386–412
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH. 2011. fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage* 55:1324–38
- Gorno-Tempini ML, Brambati SM, Ginex V, Ogar J, Dronkers NF, et al. 2008. The logopenic/phonological variant of primary progressive aphasia. *Neurology* 71:1227–34
- Graff-Radford J, Jones DT, Strand EA, Rabinstein AA, Duffy JR, Josephs KA. 2014. The neuroanatomy of pure apraxia of speech in stroke. *Brain Lang.* 129:43–46
- Groswasser Z, Korn C, Groswasser-Reider I, Solzi P. 1988. Mutism associated with buccofacial apraxia and bihemispheric lesions. *Brain Lang.* 34:157–68



- Guenther FH. 2016. *Neural Control of Speech*. Cambridge, MA: MIT Press
- Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, et al. 2009. A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* 4:e8218
- Gultekin YB, Hage SR. 2017. Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat. Commun.* 8:14046
- Hage SR, Gavrilov N, Nieder A. 2013. Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* 25:1692–701
- Hage SR, Nieder A. 2013. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* 4:2409
- Hage SR, Nieder A. 2016. Dual neural network model for the evolution of speech and language. *Trends Neurosci.* 39:813–29
- Hammerschmidt K, Jurgens U. 2007. Acoustical correlates of affective prosody. *J. Voice* 21:531–40
- Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J. 2009. Female mice respond to male ultrasonic ‘songs’ with approach behaviour. *Biol. Lett.* 5:589–92
- Hammerschmidt K, Whelan G, Eichele G, Fischer J. 2015. Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. *Sci. Rep.* 5:8808
- Heald SLM, Nusbaum HC. 2014. Speech perception as an active cognitive process. *Front. Syst. Neurosci.* 8:35
- Heldner M, Edlund J. 2010. Pauses, gaps and overlaps in conversations. *J. Phonetics* 38:555–68
- Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8:393–402
- Hickok G, Poeppel D. 2016. Neural basis of speech perception. In *Neurobiology of Language*, ed. G Hickok, S Small, pp. 299–310. London: Academic
- Hietanen JK, Leppänen JM, Lehtonen U. 2004. Perception of emotions in the hand movement quality of Finnish sign language. *J. Nonverbal Behav.* 28:53–64
- Holler J, Kendrick KH, Levinson SC. 2018. Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychon. Bull. Rev.* 25:1900–8
- Houde JF, Jordan MI. 1998. Sensorimotor adaptation in speech production. *Science* 279:1213–16
- Indefrey P. 2011. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255
- Jasnow M, Feldstein S. 1986. Adult-like temporal characteristics of mother-infant vocal interactions. *Child Dev.* 57:754–61
- Jurgens U. 2009. The neural control of vocalization in mammals: a review. *J. Voice* 23:1–10
- Jurgens U, von Cramon D. 1982. On the role of the anterior cingulate cortex in phonation: a case report. *Brain Lang.* 15:234–48
- Kelley NJ, Schmeichel BJ. 2014. The effects of negative emotions on sensory perception: Fear but not anger decreases tactile sensitivity. *Front. Psychol.* 5:942
- Kendon A. 1997. Gesture. *Annu. Rev. Anthropol.* 26:109–28
- Klapp ST, Maslovat D. 2020. Programming of action timing cannot be completed until immediately prior to initiation of the response to be controlled. *Psychon. Bull. Rev.* 27:821–32
- Kogure M. 2007. Nodding and smiling in silence during the loop sequence of backchannels in Japanese conversation. *J. Pragmat.* 39:1275–89
- Kotz SA, Schwartze M. 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14:392–99
- Krivokapic J, Tiede MK, Tyrone ME. 2017. A kinematic study of prosodic structure in articulatory and manual gestures: results from a novel method of data collection. *Lab. Phonol.* 8:3
- Kuhl PK, Meltzoff AN. 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100:2425–38
- Lametti DR, Nasir SM, Ostry DJ. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J. Neurosci.* 32:9351–58
- Laplane D, Talairach J, Meininger V, Bancaud J, Orgogozo JM. 1977. Clinical consequences of corticectomies involving the supplementary motor area in man. *J. Neurol. Sci.* 34:301–14
- LeDoux JE. 2000. Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23:155–84
- Lee EK, Brown-Schmidt S, Watson DG. 2013. Ways of looking ahead: hierarchical planning in language production. *Cognition* 129:544–62



- Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M, et al. 2011. Youngsters do not pay attention to conversational rules: Is this so for nonhuman primates? *Sci. Rep.* 1:22
- Leonardo A, Konishi M. 1999. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399:466–70
- Lerner JS, Li Y, Valdesolo P, Kassam KS. 2015. Emotion and decision making. *Annu. Rev. Psychol.* 66:799–823
- Levelt WJ, Roelofs A, Meyer AS. 1999. A theory of lexical access in speech production. *Behav. Brain Sci.* 22:1–38
- Levinson SC, Torreira F. 2015. Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731
- Liao DA, Zhang YS, Cai LX, Ghazanfar AA. 2018. Internal states and extrinsic factors both determine monkey vocal production. *PNAS* 115:3978–83
- Liu HX, Lopatina O, Higashida C, Fujimoto H, Akther S, et al. 2013. Displays of paternal mouse pup retrieval following communicative interaction with maternal mates. *Nat. Commun.* 4:1346
- Lombard E. 1911. Le signe de l'elevation de la voix. *Ann. Mal. l'Oreille Larynx* 37:101–19
- Long MA, Fee MS. 2008. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456:189–94
- Long MA, Katlowitz KA, Svirsky MA, Clary RC, Byun TM, et al. 2016. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* 89:1187–93
- Lopes PN, Salovey P, Côté S, Beers M, Petty RE. 2005. Emotion regulation abilities and the quality of social interaction. *Emotion* 5:113–18
- Lucarini V, Grice M, Cangemi F, Zimmermann JT, Marchesi C, et al. 2020. Speech prosody as a bridge between psychopathology and linguistics: the case of the schizophrenia spectrum. *Front. Psychiatry* 11:531863
- Luo J, Hage SR, Moss CF. 2018. The Lombard effect: from acoustics to neural mechanisms. *Trends Neurosci.* 41:938–49
- Mao CC, Coull BM, Golper LA, Rau MT. 1989. Anterior operculum syndrome. *Neurology* 39:1169–72
- Marshall J, Atkinson J, Smulovitch E, Thacker A, Woll B. 2004. Aphasia in a user of British Sign Language: dissociation between sign and gesture. *Cogn. Neuropsychol.* 21:537–54
- Marslen-Wilson WD. 1975. Sentence perception as an interactive parallel process. *Science* 189:226–28
- McClelland JL, Elman JL. 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18:1–86
- McClelland JL, Mirman D, Holt LL. 2006. Are there interactive processes in speech perception? *Trends Cogn. Sci.* 10:363–69
- McGurk H, MacDonald J. 1976. Hearing lips and seeing voices. *Nature* 264:746–48
- Mendoza-Denton N, Jannedy S. 2011. Semiotic layering through gesture and intonation: a case study of complementary and supplementary multimodality in political speech. *J. Engl. Linguist.* 39:265–99
- Meyer AS. 1996. Lexical access in phrase and sentence production: results from picture–word interference experiments. *J. Mem. Lang.* 35:477–96
- Michael V, Goffinet J, Pearson J, Wang F, Tschida K, Mooney R. 2020. Circuit and synaptic organization of forebrain-to-midbrain pathways that promote and suppress vocalization. *eLife* 9:e63493
- Miller CT, Beck K, Meade B, Wang X. 2009a. Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 195:783–89
- Miller CT, Eliades SJ, Wang X. 2009b. Motor planning for vocal production in common marmosets. *Anim. Behav.* 78:1195–203
- Miller CT, Thomas AW, Nummela SU, de la Mothe LA. 2015. Responses of primate frontal cortex neurons during natural vocal communication. *J. Neurophysiol.* 114:1158–71
- Miller GA, Heise GA, Lichten W. 1951. The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psychol.* 41:329–35
- Mücke D, Grice MT. 2014. The effect of focus marking on supralaryngeal articulation—Is it mediated by accentuation? *J. Phon.* 44:47–61
- Murphy D, Cutting J. 1990. Prosodic comprehension and expression in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* 53:727–30
- Niedenthal PM, Wood A. 2019. Does emotion influence visual perception? Depends on how you look at it. *Cogn. Emot.* 33:77–84



- Nieder A. 2009. Prefrontal cortex and the evolution of symbolic reference. *Curr. Opin. Neurobiol.* 19:99–108
- Nieder A, Mooney R. 2020. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375:20190054
- Nielsen JB. 2016. Human spinal motor control. *Annu. Rev. Neurosci.* 39:81–101
- Norris MR, Drummond SS. 1998. Communicative functions of laughter in aphasia. *J. Neurolinguistics* 11:391–402
- Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. 2007. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex* 17:2251–57
- Okobi DE Jr., Banerjee A, Matheson AMM, Phelps SM, Long MA. 2019. Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363:983–88
- Osmanski MS, Seki Y, Dooling RJ. 2021. Constraints on vocal production learning in budgerigars (*Melopsittacus undulatus*). *Learn. Behav.* 49:150–58
- Papafragou A, Grigoroglou M. 2019. The role of conceptualization during language production: evidence from event encoding. *Lang. Cogn. Neurosci.* 34:1117–28
- Parrell B, Goldstein L, Lee S, Byrd D. 2014. Spatiotemporal coupling between speech and manual motor actions. *J. Phon.* 42:1–11
- Pedersen PM, Vinter K, Olsen TS. 2004. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc. Dis.* 17:35–43
- Penfield W, Boldrey E. 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60:389–443
- Pepperberg IM. 2002. *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Cambridge, MA: Harvard Univ. Press
- Petkov CI, Jarvis ED. 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4:12
- Pezzulo G, Donnarumma F, Dindo H. 2013. Human sensorimotor communication: a theory of signaling in online social interactions. *PLOS ONE* 8:e79876
- Phillips C. 2001. Levels of representation in the electrophysiology of speech perception. *Cogn. Sci.* 25:711–31
- Pichon S, Kell CA. 2013. Affective and sensorimotor components of emotional prosody generation. *J. Neurosci.* 33:1640–50
- Pika S, Wilkinson R, Kendrick KH, Vernes SC. 2018. Taking turns: bridging the gap between human and animal communication. *Proc. Biol. Sci.* 285:20180598
- Poletti M, Nucciarone B, Baldacci F, Nuti A, Lucetti C, et al. 2008. Gestural buffer impairment in early onset corticobasal degeneration: a single-case study. *Neuropsychol. Trends* 4:45–58
- Pomerantz A, Heritage J. 2013. Preference. In *Handbook of Conversation Analysis*, ed. J Sidnell, T Stivers, pp. 210–28. Cambridge, UK: Cambridge Univ. Press
- Portfors CV. 2007. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J. Am. Assoc. Lab. Anim. Sci.* 46:28–34
- Rangel A, Camerer C, Montague PR. 2008. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9:545–56
- Redford MA. 2015. Unifying speech and language in a developmentally sensitive model of production. *J. Phon.* 53:141–52
- Reilly JS, McIntire ML, Seago H. 1992. Affective prosody in American Sign Language. *Sign Lang. Stud.* 75:113–28
- Rilling JK, Sanfey AG. 2011. The neuroscience of social decision-making. *Annu. Rev. Psychol.* 62:23–48
- Rohrer JD, Warren JD, Rossor MN. 2009. Abnormal laughter-like vocalisations replacing speech in primary progressive aphasia. *J. Neurol. Sci.* 284:120–23
- Roy S, Zhao L, Wang X. 2016. Distinct neural activities in premotor cortex during natural vocal behaviors in a new world primate, the common marmoset (*Callithrix jacchus*). *J. Neurosci.* 36:12168–79
- Saltzman EL, Munhall KG. 1989. A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* 1:333–82
- Samuel AG. 1996. Does lexical information influence the perceptual restoration of phonemes? *J. Exp. Psychol.* 125:28–51



- Samuel AG. 2011. Speech perception. *Annu. Rev. Psychol.* 62:49–72
- Saslow LR, McCoy S, van der Lowe I, Cosley B, Vartan A, et al. 2014. Speaking under pressure: Low linguistic complexity is linked to high physiological and emotional stress reactivity. *Psychophysiology* 51:257–66
- Schegloff EA. 2000. Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29:1–63
- Seddoh SA. 2004. Prosodic disturbance in aphasia: speech timing versus intonation production. *Clin. Linguist. Phon.* 18:17–38
- Seyfarth RM, Cheney DL. 1986. Vocal development in vervet monkeys. *Anim. Behav.* 34:1640–58
- Seyfarth RM, Cheney DL, Marler P. 1980. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210:801–3
- Shattuck-Hufnagel S, Ren A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front. Psychol.* 9:1514
- Shattuck-Hufnagel S, Turk AE. 1996. A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* 25:193–247
- Simpson HB, Vicario DS. 1990. Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.* 10:1541–56
- Skipper JI, Devlin JT, Lametti DR. 2017. The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain Lang.* 164:77–105
- Smith M, Wheeldon L. 1999. High level processing scope in spoken sentence production. *Cognition* 73:205–46
- Smyth MM, Pendleton LR. 1989. Working memory for movements. *Q. J. Exp. Psychol.* 41:235–50
- Snedeker J, Trueswell J. 2003. Using prosody to avoid ambiguity: effects of speaker awareness and referential context. *J. Mem. Lang.* 48:103–30
- Sober SJ, Brainard MS. 2009. Adult birdsong is actively maintained by error correction. *Nat. Neurosci.* 12:927–31
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH. 2012. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32:8443–53
- Sorg BA, Whitney P. 1992. The effect of trait anxiety and situational stress on working memory capacity. *J. Res. Pers.* 26:235–41
- Sossinka R, Böhner J. 1980. Song types in the zebra finch *Poephila guttata castanotis*. *Z. Tierpsychol.* 53:123–32
- Stewart AM, Lewis GF, Heilman KJ, Davila MI, Coleman DD, et al. 2013. The covariation of acoustic features of infant cries and autonomic state. *Physiol. Behav.* 120:203–10
- Stickney GS, Assmann PF. 2001. Acoustic and linguistic factors in the perception of bandpass-filtered speech. *J. Acoust. Soc. Am.* 109:1157–65
- Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, et al. 2009. Universals and cultural variation in turn-taking in conversation. *PNAS* 106:10587–92
- Tagliabata JP, Russell JL, Schaeffer JA, Hopkins WD. 2008. Communicative signaling activates ‘Broca’s’ homolog in chimpanzees. *Curr. Biol.* 18:343–48
- Talluri BC, Urai AE, Tsetsos K, Usher M, Donner TH. 2018. Confirmation bias through selective overweighting of choice-consistent evidence. *Curr. Biol.* 28:3128–35.e8
- Tchernichovski O, Mitra PP, Lints T, Nottebohm F. 2001. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291:2564–69
- Tilsen S. 2011. Effects of syllable stress on articulatory planning observed in a stop-signal experiment. *J. Phon.* 39:642–59
- Tourville JA, Reilly KJ, Guenther FH. 2008. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39:1429–43
- Trueswell JC, Papafragou A. 2010. Perceiving and remembering events cross-linguistically: evidence from dual-task paradigms. *J. Mem. Lang.* 63:64–82
- Tschida K, Michael V, Takatoh J, Han BX, Zhao S, et al. 2019. A specialized neural circuit gates social vocalizations in the mouse. *Neuron* 103:459–72.e4
- Tumer EC, Brainard MS. 2007. Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* 450:1240–44
- Turk A, Shattuck-Hufnagel S. 2019. Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. *Front. Psychol.* 10:2952



- Van Lancker Sidtis D, Kempler D, Jackson C, Metter EJ. 2010. Prosodic changes in aphasic speech: timing. *Clin. Linguist. Phon.* 24:155–67
- Vaquero L, Rodriguez-Fornells A, Reiterer SM. 2017. The left, the better: White-matter brain integrity predicts foreign language imitation ability. *Cereb. Cortex* 27:3906–17
- Vettin J, Todt D. 2004. Laughter in conversation: features of occurrence and acoustic structure. *J. Nonverbal Behav.* 28:93–115
- Vu ET, Mazurek ME, Kuo YC. 1994. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* 14:6924–34
- Warren RM. 1970. Perceptual restoration of missing speech sounds. *Science* 167:392–93
- Wegdell F, Hammerschmidt K, Fischer J. 2019. Conserved alarm calls but rapid auditory learning in monkey responses to novel flying objects. *Nat. Ecol. Evol.* 3:1039–42
- Weightman MJ, Air TM, Baune BT. 2014. A review of the role of social cognition in major depressive disorder. *Front. Psychiatry* 5:179
- Wheaton LA, Hallett M. 2007. Ideomotor apraxia: a review. *J. Neurol. Sci.* 260:1–10
- Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91:1707–17
- Woodward AL, Markman EM. 1998. Early word learning. In *Handbook of Child Psychology, Vol. 2: Cognition, Perception, and Language* ed. W Damon, pp. 371–420. Hoboken, NJ: John Wiley & Sons
- Zemlin WR. 1998. *Speech and Hearing Science: Anatomy and Physiology*. Needham Heights, MA: Allyn and Bacon.
- Zimnik AJ, Lara AH, Churchland MM. 2019. Perturbation of macaque supplementary motor area produces context-independent changes in the probability of movement initiation. *J. Neurosci.* 39:3217–33

