JOINT SPEECH AND ITS RELATION TO JOINT ACTION

FRANK A. RUSSO
Ryerson University, Toronto, Canada

IN HIS ARTICLE “THE TERRITORY BETWEEN Speech and Song: A Joint Speech Perspective,” Cummins (2020) argues that research has failed to adequately recognize an important category of vocal activity that falls outside of the domains of language and music, at least as they are typically defined. This category, referred to by Cummins as joint speech, spans a range of vocal activity so broad that it is not possible to define it using musical or phonetic terms. Instead, the feature that draws the varied examples together is vocal activity that is coordinated across participants and embedded in a physical and social context. In this invited commentary, I argue that although joint speech adds an important thread to the discourse on the relations between speech and song by putting an emphasis on the collective, it is ultimately related to a wider class of joint action phenomena found in the animal kingdom.

Received: December 9, 2019, accepted December 16, 2019.

Key words: vocal, sensorimotor coordination, social connectedness

CUMMINS (2020) MAKES NO CLAIMS ABOUT the biological origins of joint speech and leaves open the possibility that it is an entirely human invention. However, in assessing its value as a new category of behavioral investigation, it seems prudent to further contemplate inclusion criteria, and its connection to non-human behavior. As a starting point, it is worth noting that all of the examples described by Cummins involve movements that are more or less coordinated in time across members of a group. One example he cites involves coordination in time without the presence of a discernable beat (an oath swearing ceremony where the timing of individual members is not highly coordinated). The rest of the examples he provides would appear to have a beat with varying levels of beat salience and/or metrical structure—i.e., a clear beat without meter (Gregorian chant), a weak beat in which a metrical structure emerges over time (political song), and a strong beat in which a metrical structure is evident from the outset (choral singing). In addition, the vocal activity may occur with or without the assistance of written text and notation, but it is always emergent and dynamic, influenced by the embodiment of its participants and the environment in which they are situated. Accordingly, while beat, metrical structure, and notation are not inclusion criteria, coordinated vocal activity, dynamic sharing of information, and specificity of physical and social context are inclusion criteria.

I would argue that with the exception of coordinated vocal activity, these inclusion criteria strongly resemble joint action—defined here as movement that is shaped by the physical and social context, and more or less coordinated in time and space. In the case of vocal movement, the temporal coordination may be thought of as the pacing of utterances, while the spatial coordination may be thought of as the shaping of the vocal tract, the tensing of the vocal folds, and the swaying of bodies. The collective vocal movement can be speech-like or song-like but the mere act of producing these movements collectively pushes them closer to song by regularizing timing and stretching vowels. One unique and noteworthy aspect of vocally based joint action is that it is possible to observe coordination across the group while being embedded within it as an actor. This ability for an actor to participate in, as well as observe, coordination across the group is necessarily limited when it comes to joint movements that are primarily observable through vision, as is the case in dance. Our vision provides us with a limited field of view and is further hampered by occlusion effects, while our hearing allows us to “hear in all directions.”

Nevertheless, it is not very difficult to conceive of joint speech as but one family of examples, of a wider class of joint action phenomena found not only in humans but also more broadly in the animal kingdom. For example, a flock of 400 birds moving at high speed can change its collective direction in as little as half a second without incident (Attanasi et al., 2014). This type of phenomenon depends on some form of sensorimotor information that is coordinated between co-actors. Although sensorimotor coordination can be observed at the level of the group (flock), its origin always starts at the level of local clusters where there is an adjustment to the flight path (typically in response to a threat). These
local changes then propagate in wave-like fashion to envelop the entire flock. Similar forms of sensorimotor coordination can be observed in swarming ants, schooling fish, and herding sheep.

The ability of a flock to execute joint action has been well described using a limited set of mathematical rules that do not require the need for a beat (Attanasi et al., 2014; Sumpter et al., 2018). Other collective behaviors that may be observed in the animal kingdom fall closer to rhythmic activity found in human music making. For example, coordinated oscillatory activity can be observed in some species of frogs (Jones, Jones, & Ratnam, 2014), crickets (Greenfield & Roizen, 1993; Sismondo, 1990), and fireflies (Buck & Buck, 1968). Although this coordinated oscillatory activity is executed without a leader and thought to depend on sensorimotor coordination, it does not appear to have the same level of flexibility as beat-like phenomena observed in vocal learning species such as songbirds, parrots, and humans (Patel, 2006). The beat is ultimately a psychological construct (London, 2012) that depends upon the entrainment of internal neural oscillators (e.g., the basal ganglia in humans; Grahn, 2009). Once neural entrainment has been established across the collective, sharing of sensorimotor information is greatly facilitated, which in turn allows for some flexibility in expression at the individual level.

By considering joint speech in the context of joint action (Vesper et al., 2017), it may be possible to better encompass the varied mechanisms that underpin joint speech. The most basic mechanism is a common goal (e.g., recitation of an oath; cf. flying in the same direction). In most cases there is some form of sensorimotor coordination (e.g., rate accommodation; cf. adjustments to flight path). In rarer instances, a beat is clearly present (e.g., choral singing; cf. dancing). The presence of a beat affords an additional level of flexibility in the coordinated oscillatory activity allowing for flexibility inclusive of information exchanged across the senses (Russo, 2019).

From an evolutionary standpoint, joint action has been interpreted with respect to survival. The term “selfish herd” was coined by evolutionary Biologist
William Hamilton (1971) to describe the tendency for groups of con-specifics to clump together to avoid predation. There exists experimental evidence that this behavioral tendency is effective against predator threats (Treherne & Foster, 1981) and that a predator is more likely to target prey possessing weaker coordination (Ioannou, Guttal, & Couzin, 2012). From this perspective, it is easy to think about joint speech as a means of building up group resiliency or warding away threats to the tribe. There is mounting evidence that joint action in the form of singing or drumming has the capacity to enhance trust, feelings of social connectedness, and prosociality (Cirelli, Einarson, & Trainor, 2014; Cross, Turgeon & Atherton, 2019; Good & Russo, 2016; Good, Choma, & Russo, 2017; Hove & Risen, 2009; Kirschner & Tomasello, 2010; Tungceng & Cohen, 2016; Valdesolo, Ouyang & DeSteno, 2010; Wiltermuth & Heath, 2009).

In summary, I have argued that joint speech may be better understood as but one family of examples of a wider class of joint action phenomena. Joint speech has elements of music making but dispels with the notion of performer/audience, focusing instead on the social contexts that enable coordination between co-actors and the resultant grounding of collectives. This is a refreshing and important view of music for music science to embrace. I would encourage further elaboration of this theory, with particular consideration regarding the relation of joint speech to other forms of joint action.

Author Note

Correspondence concerning this article should be addressed to Frank A. Russo, Department of Psychology, Jorgenson Hall (906), Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada. E-mail: russo@ryerson.ca

References


