Why Beatboxing?

Reed Blaylock
University of Southern California

1 Introduction

Beatboxing is a type of non-linguistic vocal percussion that can be performed as an accompaniment to linguistic music (e.g. rapping or a cappella singing) or as a standalone performance—the latter being primarily the focus here. Beatboxers are increasingly recognized in both scientific and popular literature as artists who push the limits of the vocal tract with sounds that have only recently been captured with modern imaging technology. Scientific study of beatboxing is valuable on its own merits, especially for beatboxers hoping to teach and learn beatboxing more effectively. But beatboxing science can also serve as an avenue of Linguistics.

Similar to a control condition in an experiment, beatboxing can be a system against which to compare speech. Beatboxing and speech both have their own set of discrete vocal sound units which are temporally organized into an infinite number of complex, rapid, hierarchically structured utterances. The greatest difference between them is the message: speech carries a linguistic message, and beatboxing does not. There are few other vocal behaviors—perhaps none—that on their face bear such a similarity to speech without also carrying a linguistic message. This makes beatboxing useful for addressing questions about how language is situated in human cognition more generally. Can there be phonology without features encoding contrastive meaning? Which sounds of beatboxing are unattested in speech, and why are they suitable for one system but not the other? What parts of language remain when the message is taken away? How much of speech and language are special, and how much are shared by nonlinguistic human behaviors? That is, what is domain-specific, and what is domain-general?

2 The art and science of beatboxing

2.1 *Beatboxing art* The foundation of beatboxing lies in hip hop. The "old school" of beatboxing began as human mimicry of the sounds of a beat box, a machine that synthesizes percussion sounds and other sound effects. The beat box created music that an MC could rap over; when a beat box wasn't available, a human could perform the role of a beat box by emulating it vocally. The last four decades have given human beat boxes (beatboxers) plenty of time to innovate artistically and mechanically. Modern beatboxing performances often stand alone: if there are any words, they are only occasional and woven by the beatboxer into the beat pattern rather than said by a second person. There are also related art forms like beatrhyming where singing/rapping and beatboxing are fully integrated, but this is a different vocal behavior. (Combining words or other vocal behaviors into beatboxing is sometimes called multi-vocalism.) Beat patterns in the "new school" of beatboxing reflect contemporary popular music styles.

Beatboxing evolves through innovation of new sounds or sound variations, patterns (e.g. combinations of sounds or styles of breathing), and integration with other behaviors (e.g. beatboxing flute, beatboxing cello, beatrhyming, beatboxing with other beatboxers). For expert beatboxers, the goal is to create art through innovation while keeping up with trends. This innovation is constrained by both physical and cultural forces. The major physical constraint is the vocal tract itself which limits the speed and quality (i.e. constriction degree and location) of possible movements; new beatboxing sounds and patterns are thought to arise from testing these physical limitations. As for cultural forces, both the musical genres that inspire beatboxing and the preferences of beatboxers themselves have a role.

Since the initial role of beatboxing was to provide a clear beat by emulating drum sounds, non-continuant stops and affricates became very common while vowel-like sounds are almost never used. When drawing on inspiration from other musical sources, related genres like electronic dance music would have been appealing for their percussive similarities. Contemporary beat patterns keep the percussive backbone, though some sustained sounds including different types of phonation man be used concurrently. More broadly, beatboxing shares musical

* Thanks to the AJL 7 organizers and Seunghun Lee for the invitation to speak and write about beatboxing. Thanks also to Louis Goldstein and Mairym Llorens for their insightful feedback on the many drafts that led to this one.

Proceedings of *AJL 7*

properties with a broad range of (Western) genres, resulting in common patterns. One common property is 4/4 time, which signifies that the smallest musical phrases each contain four main events. Another common property is the placement of emphasis on the backbeat via snare sounds (Greenwald 2002). These types of properties, together with the vocal modality, shape the musical style and evolution of beatboxing. Innovation in beatboxing occurs within these constraints.

Common advice in beatboxing pedagogy is to learn incrementally. New aspiring beatboxers are encouraged to start by drilling the fundamentals: basic sounds like Kick Drums {B} [p'], Closed Hi-Hats {t} [t'], and PF Snares {PF} [pf'] should become familiar first in isolation, then in combos and beat patterns to practice them in a rhythmic context. (Curly brackets a beatboxing sound in Standard Beatbox Notation [TyTe & Splinter 2002/2014, Stowell 2003], while square bracket notation indicates an approximation in International Phonetic Alphabet notation.) Once the relatively small set of sounds is secure, it is time to learn new sounds that facilitate breath management—this is important for performing progressively more complex and intensive beat patterns that demand more air regulation. At the same time, new beatboxers also need to focus on "technicality", a jargon word in the beatboxing community that refers to how accurately and precisely a sound is performed. Reference to and imitation of other beatboxers is common for establishing ideals and task targets. All of these basics are the foundations from which a beatboxer can start to innovate by making novel sounds and beat patterns, and beatboxers continue to revisit these different facets of their art to make improvements at multiple time scales (i.e. improving one sound, improving a combination of sounds, developing a flow or a new style). Beatboxers are therefore often overtly aware of some facet of their beatboxing, contrary to how fluent speakers of a language may not be aware of their own performance. Beatboxers may also beatbox very differently depending on the sounds they know and which facet of beatboxing they are practicing.

All of this leads to a few premises for scientific investigation of beatboxing. The fact that beatboxers are aiming for particular sound qualities and flow patterns means that we should expect to find beatboxing patterns that balance specific aesthetics with motor efficiency. The lack of words in beatboxing, the interest in imitating instruments/sound effects, and the drive to innovate through the use of new sounds are all hints that whatever beatboxing phonology may exist is not a variation of speech phonology but a sound organization system in its own right. And while there is a growing consensus about the names and production strategies for different beatboxing sounds within the beatboxing community, the fact that beatboxers are practicing artists who focus on different parts of their beatboxing and cultivate different styles tells us that differences among beatboxing performances may be tied to the beatboxers' identities.

2.2 *Beatboxing science* A guiding theme in beatboxing science is the study of vocal agility and capability (Dehais-Underdown 2021). The complex sounds and patterns of beatboxing inform our understanding of what kinds of vocal sound-producing movements and patterns can be performed efficiently—and sometimes surprise us when we see articulations that we didn't think were possible. This in turn offers a better general phonetic framework for studying the relationship between linguistic tasks, cognitive limitations, physical limitations, and motor constraints in the evolution of speech.

Likewise, knowing more about the physical abilities of the vocal tract also informs our understanding of disordered or otherwise non-normative speech production strategies. Some researchers advocate for using beatboxing for speech therapy (Pillot-Loiseau et al. 2021), for example the BeaTalk strategy which has been used to improve speech in adults (Icht 2018, 2021, Icht & Carl 2022). Beatboxers also use beatboxing in therapy with children (Swissbeatbox 2019; Martin & Mullady, n.d.). (See also Himonides et al. 2018, Moors et al. 2020.) Although beatboxing interventions for therapeutic purposes are still quite new, the tantalizingly obvious connection between beatboxing and speech via the vocal tract they share has generated interest within the beatboxing and academic communities.

Crucial to both these branches of inquiry but almost completely undeveloped within the field is a theory of beatboxing cognition. The literature offers just three claims about beatboxing cognition so far, none of which are firmly established: one about the intentions of beatboxers, and two about the fundamental units of beatboxing. There is a general consensus that, based on the origins of beatboxing as a tool for supporting hip hop emcees, a beatboxer's primary intention is to imitate the sounds of a drum kit, electronic beat box, and a variety of other sound effects (Lederer 2005, Stowell & Plumbley 2008, Pillot-Loiseau et al. 2020). But treating beatboxing as simple imitation is reductive and disingenuous to the primacy of the art form (Woods, 2012). Even in the earliest days, old school beatboxers established distinctive vocal identities that were surely not just attempts to mimic different electronic beat boxes. The new school of beatboxing has come a long way since then and shows rapidly evolving preferences in artistic expression that a drive for pure imitation seems unlikely to motivate.

As for the cognitive representations of the sounds themselves, Evain et al. (2019) and Paroni et al. (2021) posit the notion of a "boxeme" by analogy to the phoneme—an acoustically and articulatorily distinct building block of a beatboxing sequence. While they imply that boxemes are meant to be a hypothesis of cognitive units, they do not address other questions begged by the phoneme analogy (Dehais-Underdown 2021): Are boxemes the smallest compositional units or are they composed of even smaller elements, as phonemes are thought to be composed of features? Does beatboxing exhibit phonological patterns that require a theory with some degree of abstraction? And are boxemes symbolic units, action units, or something else? Separately, Guinn and Nazarov (2018) argue for the active role of phonological features in beatboxing based on evidence from variations in beat patterns and phonotactic place restrictions (an absence of beatboxing coronals in prominent metrical positions). They do not link features back to larger (i.e. segment-sized) units; while they offer the possibility that speech and beatboxing features are linked, it remains unclear whether or how speech representations and beatboxing representations are connected.

The lack of work on beatboxing cognition is understandable: the field of beatboxing science is still in its infancy with less than 20 years of research, and the few scientists involved in the field have had their hands full with other more immediate and tractable questions. But it will be difficult to use beatboxing to inform an account of the physical and cognitive factors that shape speech without both physical and cognitive accounts of beatboxing. And while the viability of beatboxing as a tool for speech therapy is ultimately an empirical question, an explicit theory of beatboxing cognition and the relationship between beatboxing and speech should help predict which interventions are more likely to work.

3 Beatboxing as a lens for linguistic inquiry

There are plenty of ways that beatboxing could be useful for linguists. The question addressed here is one of domain-specificity: Does the human capacity for language consist only of a specialized composite of other cognitive systems, or is there some component that is unique to language and cannot be attributed to other cognitive systems (e.g. Anderson, 1981)? Section 3.1 provides a brief overview of this question in Linguistics, and section 3.2 focuses on the role of domain-specific and domain-general models just within phonology. Section 3.3 sketches out an approach for analyzing speech and beatboxing within the same partially domain-general model based in Articulatory Phonology.

3.1 *The question of domain-specificity* The question of domain-specificity has been central in the development of major linguistic paradigms over the last several decades, including the Minimalist program that views the human language faculty as only minimally domain-specific (the language faculty in the narrow sense) and otherwise composed of a unique assembly of other cognitive functions (e.g. Hauser et al. 2002, Collins 2017 provides an overview).

One of the strongest theories of domain-specificity in cognition comes from Fodor (1983) who offers a modular approach in which a cognitive domain constitutes its own system. In the original conception, modules are low-level (mostly sensory input) systems which are likely to be encapsulated, automatic, innate, and which perform computations exclusively over inputs relevant to their domain—hence, domain-specific. Modules are distinct from the non-specific handling of general cognitive processing. Liberman & Mattingly's (1985) Motor Theory couched speech perception as a linguistic module built around the relationship between intended phonetic gestures and their acoustic output. The Motor Theory proposes that speech perception is a parallel system to domain-general auditory processing, a claim supported by duplex perception tasks (Liberman et al. 1981, Mann & Liberman 1983). Modularity has been conceived of in many different ways by now, and whether or not a system like language shows all of the typical traits (e.g. encapsulation, innateness) are open to empirical testing, but domain-specificity remains key to the modular theory (Coltheart 1999). Even when phonology is not considered a module in the strictest sense, it is still common to make reference to the module-inspired "interface" between phonetics and phonology which implies that the linguistic system of sounds is distinct from the physical implementation of sounds (cf. Ohala 1990).

One of the key arguments in favor of domain-specificity is tied up with innateness: there are substantial barriers for the infant attempting to learn language, including lack of invariance and lack of segmentability in the acoustic signal (Liberman et al. 1967, see also Fowler et al. 2016). Given how quickly and effectively newborns learn speech production and perception, it stands to reason that humans may be born with a language faculty that provides a universal starting point for the acquisition process. This language faculty is domain-specific insofar as the innate cognitive scaffolding is tailored to address linguistic issues. Werker and Tees (1984) and related work demonstrated that infants can distinguish speech sounds across the same categorical boundaries that adults use.

Others argue in favor of accounting for speech patterns using only language-independent, domain-general information, without relying on an innate, species-specific language capacity (Universal Grammar) (e.g. Lindblom 1983, Archangeli & Pulleyblank 2015, 2022). This approach has foregrounded major questions in phonology over the last few decades, all shaped around developing an understanding of how phonetics shapes phonology. Quantal Theory (Stevens 1989, Stevens & Keyser 2010) derives common phonological categories from quantal regions in the vocal tract where coarticulation is less likely to interfere with perception. The Theory of Vowel Dispersion (Liljencrants & Lindblom 1972, Lindblom et al. 1979) generates typologically common vowel patterns using the principle of maximal contrast but without presupposing any particular phonological categories. Likewise, proponents of the Auditory Enhancement Hypothesis (Diehl & Kluender 1989, Diehl et al. 1991) argue that the common covariation of certain phonological features is explained by their mutual compatibility in enhancing perceptual contrasts. The frame/content theory (MacNeilage 1998) posits that the origins of speech come not from a spontaneous mutation but rather evolved from homeostatic motor functions; in this case, phonological syllable structure (the frame) descended from the chewing action.

The question of domain-specificity is an undercurrent of much research in cognitive science and evolutionary psychology and often involves comparing speech and language to other types of human or non-human cognition (Hauser et al., 2002). Categorical perception has been found in chinchillas (Kuhl & Miller 1978) and crickets (Wyttenbach et al. 1996), as well as in human perception of non-speech sounds (Fowler & Rosenblum 1990) and faces (e.g. Beale & Keil 1995). Language and music share certain rhythmic (see Ravignani et al. 2017 for a recent discussion), syntactic, (Lerdahl & Jackendoff 1983) and neurological (Maess et al. 2001) qualities, with other apparently cross-domain ties (Feld & Fox 1994, Bidelman et al. 2011). Comparison of neurotypical speech and disordered speech contributes to a neurological aspect of the discussion such as whether the motor planning in speech uses specialized or domain-general circuitry (Ballard et al. 2003, Ziegler 2003a, 2003b).

3.2 *Domain-specificity in phonological models* Models of a theory help scientists describe and explain natural phenomena, and in doing so also predict what related phenomena we should expect to find. Despite the evidence suggesting that language may not have a domain-specific component, domain-specific generative models remain the norm in much of phonological theory. Domain-specific models are meant to describe and predict only phenomena within their own domain: in a domain-specific computational phonological model, for example, the inputs and outputs are exclusively linguistic and the grammar operates only over those linguistic elements. If the same model were used to try to account for the inputs and outputs of a different cognitive domain, then by definition the model would either fail or be subject to alterations that make it no longer domain-specific—thus undermining the theoretical premise of the model. And when the model predicts phenomena that are not observed within its domain, the model is said to be imperfect because it overgenerates. As a consequence, domain-specific models are unable to describe, explain, or predict phenomena outside of their domain.

The domain-specificity of computational phonological models was entrenched at least as early as the divorcing of phonetics from phonology (de Saussure 1916, Baudouin de Courtenay 1972) which led to interest in only those aspects of phonology which are essentially linguistic (Sapir 1925, Hockett 1955, Ladefoged 1989, see section 4 for a more thorough discussion). In programs descended from this tradition, the representations and grammar of phonological theory are domain-specific because they deal exclusively with phonological inputs, outputs, and processes. The inputs and outputs are typically expressed as phonological features: atomic representations of linguistic information defined by their relationship with each other, whose purpose it is to encode meaningful contrast, and which are the basis of phonological change (Dresher 2011, Mielke 2011). Phonological features are meant to be representations of linguistic meaning and organization. They are, obviously and crucially, not meant to be representations of any other domain.

Depending on the strictness of a model's commitment to domain-specificity, sometimes explanation in phonology can come from outside the domain (Diehl 1991). Widespread interest in the relationship between phonetics and phonology was renewed with the advent of acoustically-grounded distinctive features (Jakobson et al. 1951) and the mapping of gradient phonetic features to scalar phonetic (phonological) features in SPE (Chomsky & Halle 1968, see Keating 1996 for the dual role of phonetics in SPE). Phonological grammars commonly use phonetic grounding to constrain their outputs (Prince & Smolensky 1993/2004, Hayes, Kirchner & Steriade 2004). On the other hand, programs based on strict domain-specific modularity argue that phonetics should have no role in the makeup of the grammar (e.g. Hale & Reiss 2000). But in neither case is phonology expected to explain anything about phonetics—except perhaps at the phonetics-phonology interface where outputs from the phonological system are transduced into the inputs of the phonetic system (Keating 1996, Cohn 2007). Even then, the interface is not intended to account for any phonetic phenomenon that is not clearly the result of a linguistic intent, nor is it capable of doing so without becoming a domain-general model. Regardless of whether

there is an overt commitment to an innate Universal Grammar, the resulting phonological systems are domain-specific by design.

A domain-specific model can of course be of great practical benefit in the interest of developing a scientific account of phonology. But the domain-specificity of phonology is only a hypothesis about part of the relationship between language and the rest of human cognition. If we were to discover that phonological phenomena that are typically described with a domain-specific model also exist in another nonlinguistic behavior, then a single model that encompasses both domains may be preferable to two domain-specific models that provide separate accounts of their shared phenomena. Enter beatboxing.

Beatboxing is particularly useful in the search for the nonlinguistic presentation of phonology because beatboxing and speech seem to have many attributes in common. For both beatboxing and speech, sound is produced when the vocal tract articulators make constrictions that manipulate air pressure. Many beatboxing sounds resemble speech sounds in their constriction locations, degrees, and between-articulator coordination. Like speech sounds, beatboxing sounds have a domain-specific classification system, in this case based on their musical function (e.g. "snare", "bass", "kick") and (in part) their articulation. The sounds of beatboxing can be combined and recombined into an unlimited number of different beat patterns—hierarchically structured phrases of beatboxing sounds produced sequentially—but with certain phonotactic restrictions as discussed earlier (e.g. "the backbeat must have a snare sound"). And some common beatboxing sounds resemble speech sounds enough that they can replace speech sounds in a beatrhyming utterance (Fukuda et al. 2022, Blaylock 2022). Given the articulatory and organizational similarities between beatboxing and speech, beatboxing is an ideal nonlinguistic behavior against which to compare speech in the search for phenomena that are unique to phonology (if any).

3.3 *Phonological models for speech and beatboxing* There is plenty more work to be done to understand how beatboxing sounds are organized and the extent to which beatboxing has language-like phonology. For the purpose of illustrating some of the potential implications, assume for the moment that there *is* a distinct beatboxing phonology: like a phonological system of a language in form, but without any linguistic function. The approaches to domain-specificity in phonology described above offer different explanations for how beatboxing ended up looking phonological.

Phonology could be a domain-specific system from which beatboxing copies cognitive properties. In this view, beatboxing is "parasitic" on phonology. The most rigid sense of this copying would be for beatboxing to have taken the phonological representations and grammar from phonology without any change. This cannot be true: though beatboxing sounds and speech sounds can be similar in their composition and apparent phonological behavior, the sounds of beatboxing can also be very different from the sounds of any language the beatboxer knows. Some beatboxing sounds are unattested in any known language of the world, and therefore not in the linguistic phonological inventories of beatboxers (Blaylock et al. 2017).

The more plausible copying hypothesis is that beatboxing might take certain qualities of phonological units and grammar—like the combinatorial nature of the representations and the framework of a computational grammar (e.g. Optimality Theory)—and adapt them into a new set of beatboxing representations and beatboxing grammar. Beatboxing phonology would then not have to be identical to speech, but beatboxing-phonological phenomena would be constrained by the limitations of the representations and grammar whose form it borrowed. Some evidence points in this direction: neophyte beatboxers commonly learn beatboxing sound patterns from adaptations of speech phrases like repeating the phrase "boots and cats" to generate a basic beatboxing sequence {B t ^K t}. Using the physical vocal apparatus to perform speechlike maneuvers could in some sense "unlock" access to phonological potential. (Hauser, Chomsky, & Fitch (2002) suggest that recursion may have similarly been adopted into speech from a different cognitive domain like navigation.) Regardless of the details, an account of beatboxing phonology that starts from the premise of a domain-specific phonological system requires two distinct models—one for speech, and one for beatboxing once the important pieces have been copied and modified.

An alternative explanation starts with domain-generality as a baseline assumption: speech and beatboxing are grounded in the same physical and cognitive capabilities, so whatever their shared capacities provide as a publicly available resource will automatically be available to them both. In the domain-general view, linguistic phonology and beatboxing phonology would be two faces of the same underlying cognitive system. An explanation of beatboxing phonology that starts from this domain-general assumption requires a single model that generates the two systems' shared behavior but that also generates their distinctive domain-defining differences. In this case, a framework like Articulatory Phonology may be appropriate.

Articulatory Phonology (Browman & Goldstein 1986, 1989) is the hypothesis that the fundamental cognitive units of phonology are not symbolic features, but actions units called "gestures". In this view, rather than having

a phonological module that interfaces with phonetics, phonology and phonetics are continuous and inseparable. Gestures unite the discrete, context-invariant properties usually attributed to abstract phonological units with the dynamic, continuous, context-dependent properties observed in running speech. These two sides of gestures are encoded together in the language of dynamical systems: the system parameters are invariant during the execution of a speech action, but the state of the system changes continuously (Fowler 1980).

Gestures also simultaneously contain both domain-specific and domain-general properties. This is because, as actions, gestures are not unique to speech but they are specialized for speech: by design, the dynamical equations in the task dynamic framework of motor control can characterize any goal-oriented action from any domain (Saltzman & Munhall, 1989). This means that gestures are on the one hand domain-general because the dynamical system that defines them can serve as the basis for any goal-oriented action, but on the other hand domain-specific because a given gesture is specialized for a speech-specific (and language-specific) goal (Browman & Goldstein 1991:314-315):

> Second, we should note that the use of dynamical equations is not restricted to the description of motor behavior in speech but has been used to describe the coordination and control of skilled motor actions in general (Cooke, 1980; Kelso, Holt, Rubin, & Kugler, 1981; Kelso & Tuller, 1984a, 1984b; Kugler, Kelso, & Turvey, 1980). Indeed, in its preliminary version the task dynamic model we are using for speech was exactly the model used for controlling arm movements, with the articulators of the vocal tract simply substituted for those of the arm. Thus, in this respect the model is not consistent with Liberman and Mattingly's (1985) concept of language or speech as a separate module, with principles unrelated to other domains. However, in another respect, the central role of the task in task dynamics captures the same insight as the "domain-specificity" aspect of the Modularity hypothesis—the way in which vocal tract articulators is yoked is crucially affected by the task to be achieved (Abbs, Gracco, & Cole, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984).

This is similar in spirit to what has been called the "anthropophonic perspective". The modern meaning of anthropophonics comes from Catford (1977), who defined anthropophonics as a person's total sound-producing potential—referring to all the vocal sound possibilities that can be described (general phonetics) of which the whole set of speech possibilities is only a subset (linguistic phonetics). (The term anthropophonics originated with Jan Baudouin de Courtenay as part of the distinction between the physical or *anthropophonic* and the psychological or *psychophonetic* properties of speech sounds.) Following Catford, Lindblom (1990) suggested an anthropophonic perspective as a non-circular strategy for defining what could be a possible sound of speech (cf Ladefoged 1989). The anthropophonic perspective is to start from domain-general vocal potential—all of the possible vocal sound-making strategies and configurations—and consider how domain-specific tasks (for Lindblom, "selection constraints") filter that potential into a coherent speech system.

Gestures as phonological units are in this sense a formalization of the anthropophonic perspective: speech actions are domain-general abilities of the vocal tract leveraged for domain-specific speech purposes (Figure 1). And although gestures and the anthropophonic perspective have mostly been applied to the realm of speech, there are no restrictions against other behaviors using the same strategy: if there is evidence for gesture-like phonological units in beatboxing, then the sounds of beatboxing can be accounted for just like speech sounds—as the result of interaction between domain-specific tasks and the domain-general potential of the vocal tract (Figure 2). Thus, one model generates two distinct behaviors from a common ground.

The primary advantage of such an approach is the ability to make direct comparisons and predictions grounded in theory. If beatboxing patterns and speech utterances are both composed from gestures, then the kinematic consequences of both types of gestures can be measured and compared, providing insights about the dynamical makeup of speech gestures. Linguists could compare the sounds of beatboxing with the sounds of different languages to identify if and how linguistic goals uniquely shape phonological inventories. The same goes for phonological patterns: for example, if beatboxing utterances exhibit phonological processes with sound alternations just like speech does (e.g. assimilation), then phonological patterns that we treat as belonging to language may be more parsimoniously analyzed as belonging to domain-general vocal control that can be leveraged for both speech and beatboxing needs. These lines of inquiry are justified in the context of a model that treats speech and beatboxing as two distinct behaviors related indirectly through their shared vocal tract.

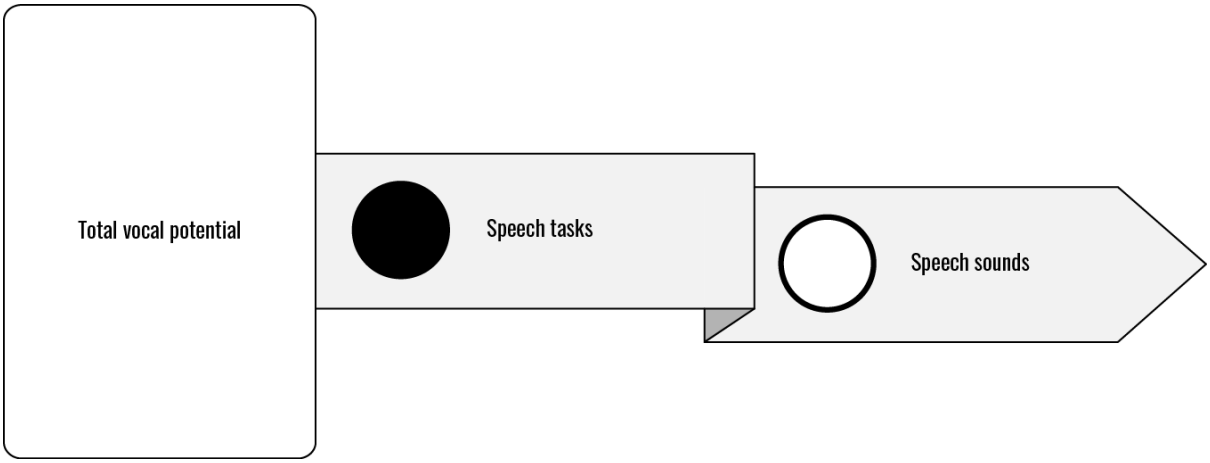

Figure 1: The anthropophonic perspective, adapted from Lindblom (1990). "Total vocal potential" refers to any and every sound and sonic dimension of the voice. Speech tasks, represented by empty circles, are constriction goals like changes in lip aperture and tongue position that are meaningful within a language. Each speech sound (or component speech act), represented by a filled-white circle, leverages a small portion of the total vocal potential to implement a speech task. in Articulatory Phonology, this occurs when gestures activate and influence the vocal articulators. In a broader sense, the speech tasks are also more holistic "selection constraints"—forces that shape speech systems at many levels (i.e. parity, intelligibility, motor efficiency).

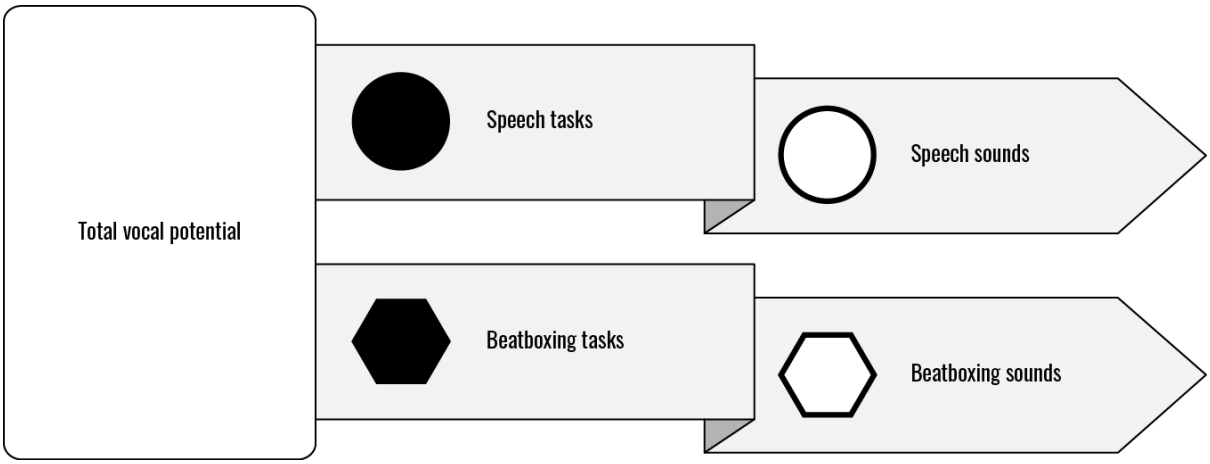

Figure 2: Beatboxing is grounded in the same vocal potential as speech, but beatboxing has its own tasks and selection constraints. A beatboxing sound is produced when the vocal tract is leveraged to satisfy the goals for a particular beatboxing sound or act. Speech and beatboxing share a similar relationship with the vocal tract and its potential; the difference is that they use the vocal tract to satisfy different aims (represented in the figure by the shaping of the vocal potential into circles for speech tasks and hexagons for beatboxing tasks).

4 Conclusion

We have considered the issue of domain-specificity of language in two different ways. The first line of thinking is about cognitive theory: there is a dichotomy between modular theories, in which each domain has its own computations, and integrated theories in which computations/mental abilities are shared across domains and assembled differently depending on the domain. The second line of thinking is a specific instantiation of this dichotomy in phonological theory: there are domain-specific theories which by design are only capable of generating phonological data, and there are frameworks that unify domain-specific tasks and domain-general abilities to explain phonological data. The latter type of account offers the potential to understand nonlinguistic behaviors like beatboxing as relatives of speech in a family of organized vocal behaviors, opening the possibility of theory-grounded comparisons of the two systems to learn more about each.

We conclude with a re-examination of what the domain of inquiry is in the first place. Section 4.1 briefly considers how narrow definitions of language risk excluding perspectives and data that could otherwise be quite

valuable for linguistic inquiry. Section 4.2 offers evidence for phonological flexibility—the ability for phonology to interact with musical (and therefore nonlinguistic) goals to create musical speech. Finally, section 4.3 recapitulates that if beatboxing is found to have its own phonology, we will see evidence that phonology is not just compatible with musical aims but actually shares crucial foundations with other uniquely human but nonlinguistic behaviors like beatboxing and other forms of musical expression.

4.1 *The scope of inquiry* A substantial part of linguistic inquiry involves searching for the properties that uniquely define language as language (e.g. Hockett 1959, 1960, Chomsky 2005). A predominant way of thinking about this question has been to distill language to its core elements. In this sense, developing a theory of language requires defining the narrowest possible scope of linguistic inquiry by excluding data that do not seem to be relevant to the essence of language. As there are different conceptions within Linguistics of what language is, there are therefore different attempts to define what counts as relevant data and what does not (Santana 2016). For example, Chomsky (1986, 1995) distinguishes between I-language (the individual, internal language of thought) and E-language (related to, among other things, the externalized and communicative function of language), and dismisses E-language as uninteresting for a linguist—thereby also dismissing certain fields of Linguistics like sociolinguistics and phonetics as not valid parts of linguistic inquiry insofar as they are concerned with the communicative function of language.

The division between phonetics and phonology predominant in the 20th century is itself based in a history of separating what is cognitive and therefore linguistic from what is physical and therefore not related to the essence of language (for a summary, see Diehl 1991). Baudouin de Courtenay (1972) separated the *psychophonetic* (phonological) from the *anthropophonic* (phonetic), and Ferdinand de Saussure (1916:31) likewise distinguished between *langue* (language) and *parole* (speech) in order to separate "ce qui est essentiel de ce qui est accessoire et plus ou moins accidentel" ("what is essential from what is incidental and more or less accidental").

Within phonology, a variety of behaviors and properties that might clearly count as linguistic expression for a community are largely ignored. Hockett (1955:17-18) claims that the aim of phonology is to investigate only the properties of speech that are unique to the domain of language, but excludes surrogate languages, singing, and whispering from phonological inquiry based on their redundancy with more normative speech:

> Perhaps no hard-and-fast line can be drawn, but there are two criteria which seem useful: discreteness of contrast and duality of pattern. … A combination of these two criteria serves, at least in a practical way, to construct a working boundary to what we shall call "linguistic behavior"; this does not necessarily imply that the term "language" must be restricted in this way for efficient use. Yet even these combined criteria leave certain marginal cases untouched. These marginal cases [Mazateco whistle-talk, African drum signals, whispering, singing words] can be referred to collectively as linguistic by-systems: they derive from a language, and are closely related to it, yet are not perhaps to be classed as language in the primary sense….
>
> That such by-systems are in general easy for a speaker of a language to learn, and that messages transmitted in them rather than in ordinary speech are usually intelligible, are facts tied up with the redundancy of the speech signal… For the bulk of our discussion we regard by-systems as not properly a part of the object of our analysis.

Likewise, Ladefoged (1989:6) argues for a narrow scope of linguistic inquiry that excludes sociolinguistic and other types of information usually present in the speech signal:

> To be more precise about what should be included within language (and hence within phonology), I suggest that we should consider language to be just the system that we need for modeling our known world. This would be very much like considering spoken language as the direct counterpart of written language. From this point of view it would be inappropriate to speak of a language being reduced to writing, implying that some part of spoken language is not present in the written form. It would be better to say that (virtually) all that is language can be expressed in speech or in writing—and all the sociolinguistic, attitudinal, emotional, stylistic and personal information, that is left out is not part of what we want to define as language.

Narrowing the definition of language seems reasonable enough: language is complex, and a smaller scope of inquiry can yield more tractable questions. But the question of "what counts" is political, not objective, and is tied up with the power structures that shape scientific paradigms and that affect who is allowed to determine "what

counts" in the first place. At least in U.S. Linguistics, this is related to a legacy of colonialism and racism (e.g. Errington 2007, Charity Hudley et al. 2020). These issues have major ramifications for the development of linguistic theory and lead us to question the narrow definitions on which so much work is predicated.

For example, the search for cognitive universals is based on a body of data and research heavily biased toward Western, Educated, Industrialized, Rich, Democratic (WEIRD) people (Henrich, Heine & Norenzayan 2010) and Written, Institutionally supported, Standardized, Prestige (WISPy) language varieties (Sedarous & Namboodiripad 2020). The resulting theories may work better for languages of more powerful communities than for languages of less powerful communities. For another example, the documentation of languages by linguists from colonial traditions has sometimes enforced dominant language biases onto transcriptions and grammars, resulting in documentation choices that the community of language users disagree with (Errington, 2007, Leonard 2017) as well as inaccurate or unreliable characterizations of linguistic patterns that are taken as fact in the construction of linguistic theory (Shih & de Lacy 2019). Narrow stipulations for which facets of language are more worthy of study can exclude valuable data and perspectives about the unique nature of language—and can also exclude the people interested in studying language from those perspectives, disproportionately affecting researchers from minoritized groups (miles-hercules 2020, Muwwakkil 2020, Calhoun et al. 2021).

4.2 *Phonology is flexible* The same constraints on what counts as language limit the inclusion of data from linguistic behaviors other than unmarked speech—that is, other than conversational, read, or laboratory-situated speech. The general absence of musical speech from the literature precludes us from understanding how flexible phonology can be when it interacts with musical goals.

The unmarked speech that is the subject of the bulk of linguistic research is continuous with many other speech behaviors and at different levels of phonological structure. Unmarked speech is prosodically continuous with poetry, rapping, chanting, and singing: just a few small adjustments to rhythm or intonation transform conversational speech into any of an abundance of genres of vocal linguistic art. A non-musical speech utterance can even become perceived as musical when it is heard a few times in a row (the speech to song illusion; Deutsch et al., 2011). Acoustic cues that express phonological contrast can be produced without the normative vocal articulator strategies: surrogate speech like talking drums (Beier 1954, Akinbo 2019), xylophones (McPherson 2018), and whistle speech (Rialland 2005) shift phonological expression to different physical systems or different uses of the vocal tract. And phonological units and grammar of unmarked speech are not only used in speech contexts: scat singing is utterly non-linguistic but follows phonological restrictions of language anyway (Shaw 2008).

These different behaviors are collaborations between phonology and various non-linguistic (e.g. musical) tasks. The collaborations are often organized to satisfy all the linguistic and non-linguistic tasks involved within the constraints of the physical systems through which they operate. In singing, unmarked speech prosody cannot manifest at the same time as sung musical melody because the vocal tract isn't flexible enough to allow the same vowel to have two different pitches or durations simultaneously; sustaining a note during a song therefore requires selecting between a musical and speech-prosodic pitch and rhythm. But while intonation and duration may be compromised, the contrastive information and structure of the speech sound units are often unperturbed: syllable structure, sound selection, and relative sound order largely remain intact because they do not compete with the rhythmic and melodic tasks of music. Compositionally there may be text-to-tune alignment where musical pitch and rhythm reflect the likely prosody of the utterance if it had been spoken non-musically (Hayes & Kaun 1996). Similar tone-to-tune alignment is found in languages with lexical tone, with tone contours exerting greater influence on the musical melody to avoid producing unintended tones (Schellenberg 2013, McPherson & Ryan 2018). And in beatrhyming, the speech and beatboxing tasks leverage their shared vocal apparatus to maximize their compatibility through place of articulation matching (Fukuda et al. 2022, Blaylock 2022).

4.3 *Why beatboxing?* The flexibility of speech as described above does not extend all the way to beatboxing. Beatboxing is not a linguistic behavior in any sense because it does not communicate a linguistic message and it does not use any of the meaningful units of language. We therefore do not find in beatboxing examples of phonology systematically collaborating with separate musical tasks to create musical language.

Instead, beatboxing science offers insights into a different dimension of phonological flexibility: that beatboxing and speech may share the same phonological foundation, including basic unit structure and organizational principles, which they leverage to very different ends. It would be this shared phonological foundation that provides the common ground for beatboxing and speech tasks to collaborate with each other in beatrhyming. The existence of such a shared phonology relates to questions about the role of language in cognition more broadly, specifically with regard to questions of domain-specificity.

References

Abbs, James H., Vincent L. Gracco & Kelly J. Cole. 1984. Control of Multimovement Coordination. *Journal of Motor Behavior* 16(2). 195–232. https://doi.org/10.1080/00222895.1984.10735318.

Akinbo, Samuel. 2019. Representation of Yorùbá Tones by a Talking Drum. An Acoustic Analysis. *Linguistique et langues africaines*. Llacan UMR 8135 CNRS/Inalco (5). 11–23. https://doi.org/10.4000/lla.347.

Anderson, Stephen R. 1981. Why Phonology Isn't "Natural." *Linguistic Inquiry* 12(4). 493–539.

Archangeli, Diana & Douglas Pulleyblank. 2015. Phonology without universal grammar. *Frontiers in Psychology* 6. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01229. (6 June, 2022.)

Archangeli, Diana & Douglas Pulleyblank. 2022. *Emergent phonology* (Conceptual Foundations of Language Science 7). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.5721159.

Ballard, K.J., D.A. Robin & J.W. Folkins. 2003. An integrative model of speech motor control: A response to Ziegler. *Aphasiology* 17(1). 37–48. https://doi.org/10.1080/729254889.

Baudouin de Courtenay, Jan. 1972. *Selected writings of Baudouin de Courtenay*. Edward Stankiewicz (ed.). Bloomington: Indiana University Press.

Beale, James M. & Frank C. Keil. 1995. Categorical effects in the perception of faces. *Cognition* 57(3). 217–239. https://doi.org/10.1016/0010-0277(95)00669-X.

Beier, Ulli. 1954. The talking drums of the Yoruba. *African Music: Journal for the International Library of African Music* 1(1). 29–31.

Bidelman, Gavin M., Jackson T. Gandour & Ananthanarayan Krishnan. 2011. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience* 23(2). 425–434. https://doi.org/10.1162/jocn.2009.21362.

Blaylock, Reed. 2022. *Beatboxing phonology*. Los Angeles, CA: The University of Southern California Ph.D. https://www.proquest.com/pqdtglobal/docview/2728251686/abstract/88C1F9FA1FE44957PQ/1. (22 January, 2023.)

Blaylock, Reed, Nimisha Patil, Timothy Greer & Shrikanth S Narayanan. 2017. Sounds of the human vocal tract. P*roc. Interspeech 2017*. 2287–2291. https://doi.org/10.21437/Interspeech.2017-1631.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an Articulatory Phonology. *Phonology* 3. 219–252.

Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2). 201–251. https://doi.org/10.1017/S0952675700001019.

Browman, Catherine & Louis Goldstein. 1991. Gestural Structures: Distinctiveness, Phonological Processes, and Historical Change. In Ignatius G. Mattingly & Michael Studdert-Kennedy (eds.), *Modularity and the motor theory of speech perception: proceedings of a conference to honor Alvin M. Liberman*. 313–338. Hove: Psychology Press.

Calhoun, Kendra, Anne H. Charity Hudley, Mary Bucholtz, Jazmine Exford & Brittney Johnson. 2021. Attracting Black students to linguistics through a Black-centered Introduction to Linguistics course. *Language* 97(1). e12–e38. https://doi.org/10.1353/lan.2021.0007.

Catford, J C. 1977. *Fundamental problems in phonetics*. Bloomington and London: Indiana University Press.

Charity Hudley, Anne H., Christine Mallinson & Mary Bucholtz. 2020. Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language* 96(4). e200–e235. https://doi.org/10.1353/lan.2020.0074.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Westport, CT: Greenwood Publishing Group.

Chomsky, Noam. 1995. Language and nature. *Mind* 104(413). 1–61.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1). 1–22. https://doi.org/10.1162/0024389052993655.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Cohn, Abigail C. 2007. Phonetics in Phonology and Phonology in Phonetics. In *Working Papers of the Cornell Phonetics Laboratory*, vol. 16, 1–31.

Collins, John. 2017. Faculties and Modules: Chomsky on cognitive architecture. In James McGilvray (ed.), *The Cambridge companion to Chomsky*, 2nd edn, 217–234. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316716694.011.

Coltheart, Max. 1999. Modularity and cognition. *Trends in Cognitive Sciences* 3(3). 115–120. https://doi.org/10.1016/S1364-6613(99)01289-9.

Cooke, J. D. 1980. The organization of simple, skilled movements. In George E. Stelmach & Jean Requin (eds.), *Advances in Psychology* 1, 199–212. https://doi.org/10.1016/S0166-4115(08)61946-9.

De Saussure, Ferdinand. 1916. *Cours de linguistique générale (1916)* (études et documents de payot). Paris: Payot.

Dehais-Underdown, Alexis, Paul Vignes, Lise Crevier-Buchman & Didier Demolin. 2021. In and out: production mechanisms in Human Beatboxing. *Proc. Mtgs. Acoust 29 November 2021* 45(1). 060005. https://doi.org/10.1121/2.0001543.

Deutsch, Diana, Trevor Henthorn & Rachael Lapidis. 2011. Illusory transformation from speech to song. *The Journal of the Acoustical Society of America* 129(4). 2245–2252. https://doi.org/10.1121/1.3562174.

Diehl, Randy L. 1991. The role of phonetics within the study of language. *Phonetica* 48(2–4). 120–134. https://doi.org/10.1159/000261880.

Diehl, Randy L. & Keith R. Kluender. 1989. On the Objects of Speech Perception. Ecological Psychology. Routledge 1(2). 121–144. https://doi.org/10.1207/s15326969eco0102_2.

Dresher, B Elan. 2011. The phoneme. In *The Blackwell companion to phonology*, 241–266.

Errington, J. 2007. *Linguistics in a colonial world: A story of language, meaning, and power*.Wiley-Blackwell.

Evain, Solène, Adrien Contesse, Antoine Pinchaud, Didier Schwab, Benjamin Lecouteux & Nathalie Henrich Bernardoni. 2019. Beatbox sounds recognition using a speech-dedicated HMM-GMM based system. *Models and Analysis of Vocal Emission for Biomedical Applications* 11(1). 2704-5846. https://doi.org/10.36253/978-88-6453-961-4

Feld, Steven & Aaron A Fox. 1994. Music and Language. *Annual Review of Anthropology* 23. 25–53.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge: MIT Press.

Fowler, Carol A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8(1). 113–133. https://doi.org/10.1016/S0095-4470(19)31446-9.

Fowler, Carol A. & Lawrence D. Rosenblum. 1990. Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance* 16(4). 742-754. https://doi.org/10.1037/0096-1523.16.4.742.

Fowler, Carol A., Donald Shankweiler & Michael Studdert-Kennedy. 2016. "Perception of the speech code" revisited: Speech is alphabetic after all. *Psychological Review* 123(2). 125–150. https://doi.org/10.1037/rev0000013.

Fukuda, Masaki, Kimura, Kosei, Blaylock, Reed, & Lee, Seunghun. 2022. Scope of beatrhyming: segments or words. In Proceedings of the AJL 6 (Asian Junior Linguists), 59–63. Sogang University, Seoul, Korea.

Greenwald, Jeff. 2002. Hip-hop drumming: The rhyme may define, but the groove makes you move. *Black Music Research Journal* 22(2). 259–271. https://doi.org/10.2307/1519959.

Guinn, Devon & Aleksei Nazarov. 2018. Evidence for features and phonotactics in beatboxing vocal percussion. Presented at the 15th Old World Conference on Phonology, University College London, United Kingdom.

Hale, Mark & Charles Reiss. 2000. Phonology as Cognition. In Noel Burton-Roberts, Philip Carr & Gerard Docherty (eds.), *Phonological knowledge: Conceptual and empirical issues*, 161–184. Oxford: Oxford University Press.

Hauser, Marc D, Noam Chomsky & W Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? 298. 1569–1579.

Hayes, Bruce & Abigail Kaun. 1996. The role of phonological phrasing in sung and chanted verse. The Linguistic Review 13(3–4). https://doi.org/10.1515/tlir.1996.13.3-4.243.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. Phonetically Based Phonology. Cambridge University Press.

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. Most people are not WEIRD. Nature. Nature Publishing Group 466(7302). 29–29. https://doi.org/10.1038/466029a.

Himonides, Evangelos, Thomas Moors, Donatella Maraschin & Marv Radio. 2018. Is there potential for using beatboxing in supporting laryngectomees? findings from a public engagement project.

Hockett, Charles F. 1955. A manual of phonology (Memoir of the International Journal of American Linguistics). Vol. 21. Indiana University Publications in Anthropology and Linguistics.

Hockett, Charles F. 1959. Animal "Languages" and Human Language. Human Biology 31(1). 32–39.

Hockett, Charles F. 1960. The Origin of Speech. Scientific American 203(3). 88–97.

Icht, Michal. 2018. Introducing the Beatalk technique: using beatbox sounds and rhythms to improve speech characteristics of adults with intellectual disability: Using beatbox sounds and rhythms to improve speech. International Journal of Language & Communication Disorders 54. https://doi.org/10.1111/1460-6984.12445.

Icht, Michal. 2021. Improving speech characteristics of young adults with congenital dysarthria: An exploratory study comparing articulation training and the Beatalk method. Journal of Communication Disorders 93. 106147. https://doi.org/10.1016/j.jcomdis.2021.106147.

Icht, Michal & Micalle Carl. 2022. Points of view: positive effects of the Beatalk technique on speech characteristics of young adults with intellectual disability. International Journal of Developmental Disabilities 1–5. https://doi.org/10.1080/20473869.2022.2065449.

Jakobson, R., C. G. Fant & M. Halle. 1951. Preliminaries to speech analysis: The distinctive features and their correlates. MIT Press.

Keating, Patricia A. 1996. The Phonetics-Phonology Interface. In UCLA Working Papers in Phonetics, vol. 92, 45–60.

Kelso, J. A. Scott, Kenneth G. Holt, Philip Rubin & Peter N. Kugler. 1981. Patterns of Human Interlimb Coordination Emerge from the Properties of Non-Linear, Limit Cycle Oscillatory Processes. Journal of Motor Behavior. Routledge 13(4). 226–261. https://doi.org/10.1080/00222895.1981.10735251.

Kelso, J. A. Scott & Betty Tuller. 1984. A Dynamical Basis for Action Systems. In Michael S. Gazzaniga (ed.), Handbook of Cognitive Neuroscience, 321–356. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-2177-2_16.

Kelso, J. A. & B. Tuller. 1984. Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 246(6). R928–R935. https://doi.org/10.1152/ajpregu.1984.246.6.R928.

Kelso, J. Scott, Betty Tuller, E. Vatikiotis-Bateson & Carol A. Fowler. 1984. Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance. US: American Psychological Association 10(6). 812–832. https://doi.org/10.1037/0096-1523.10.6.812.

Kugler, Peter N., J. A. Scott Kelso & M. T. Turvey. 1980. On the Concept of Coordinative Structures as Dissipative Structures: I. Theoretical Lines of Convergence. In George E. Stelmach & Jean Requin (eds.), Advances in Psychology (Tutorials in Motor Behavior), vol. 1, 3–47. North-Holland. https://doi.org/10.1016/S0166-4115(08)61936-6.

Kuhl, Patricia K. & James D. Miller. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. The Journal of the Acoustical Society of America. Acoustical Society of America 63(3). 905–917. https://doi.org/10.1121/1.381770.

Ladefoged, Peter. 1989. Representing Phonetic Structure. Phonetics Laboratory, Department of Linguistics, UCLA.

Lederer, Karen. 2005. The Phonetics of Beatboxing. Introduction (The Phonetics of Beatboxing). https://www.humanbeatbox.com/articles/the-phonetics-of-beatboxing-abstract/. (27 April, 2022).

Leonard, Wesley Y. 2017. Producing language reclamation by decolonising 'language.' Aperio 14. https://doi.org/10.25894/ldd146.

Lerdahl, Fred & Ray Jackendoff. 1983. A Generative Theory of Tonal Music. MIT press.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler & M. Studdert-Kennedy. 1967. Perception of the speech code. Psychological Review. Washington, US: American Psychological Association 74(6). 431–461. https://doi.org/10.1037/h0020279.

Liberman, Alvin M., David Isenberg & Brad Rakerd. 1981. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics 30(2). 133–143. https://doi.org/10.3758/BF03204471.

Liberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. Cognition 21(1). 1–36. https://doi.org/10.1016/0010-0277(85)90021-6.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. Language 48(4). 839. https://doi.org/10.2307/411991.

Lindblom, Björn. 1983. Economy of Speech Gestures. In Peter F. MacNeilage (ed.), The Production of Speech, 217–245. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4613-8202-7_10.

Lindblom, Björn. 1990. On the notion of "possible speech sound." Journal of Phonetics (Linguistic Approaches to Phonetics Papers Presented in Honor of J.C. Catford) 18(2). 135–152. https://doi.org/10.1016/S0095-4470(19)30398-5.

Lindblom, Björn, James Lubker & Thomas Gay. 1979. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics 7(2). 147–161. https://doi.org/10.1016/S0095-4470(19)31046-0.

MacNeilage, Peter F. 1998. The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21(4). 499–511. https://doi.org/10.1017/S0140525X98001265.

Maess, Burkhard, Stefan Koelsch, Thomas C. Gunter & Angela D. Friederici. 2001. Musical syntax is processed in Broca's area: an MEG study. Nature Neuroscience. Nature Publishing Group 4(5). 540–545. https://doi.org/10.1038/87502.

Mann, Virginia A. & Alvin M. Liberman. 1983. Some differences between phonetic and auditory modes of perception. Cognition 14(2). 211–235. https://doi.org/10.1016/0010-0277(83)90030-6.

Martin, Mark & Kaila Mullady. Education. Lightship Beatbox. https://www.lightshipbeatbox.com/education. (6 June, 2022).

McPherson, Laura. 2018. The Talking Balafon of the Sambla: Grammatical Principles and Documentary Implications. Anthropological Linguistics 60(3). 255–294. https://doi.org/10.1353/anl.2019.0006.

McPherson, Laura & Kevin M. Ryan. 2018. Tone-tune association in Tommo So (Dogon) folk songs. Language 94(1). 119–156. https://doi.org/10.1353/lan.2018.0003.

Mielke, Jeff. 2011. Distinctive Features. In The Blackwell Companion to Phonology, 1–25. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444335262.wbctp0017.

miles-hercules, deandre. a. (2020). Scoping the Scope: Revisiting the Aims and Paradigms of Linguistics. This IS Linguistics: Scope, Positionality, and Graduate Apprenticeship when Diversifying the Linguistics Curriculum, Linguistics Society of America webinar, September 11, 2020.

Muwwakkil, Jamaal. (2020). Positionality in Curriculum Design, Pedagogy, and Research. This IS Linguistics: Scope, Positionality, and Graduate Apprenticeship when Diversifying the Linguistics Curriculum, Linguistics Society of America webinar, September 11, 2020.

Moors, Thomas, Sanjeev Silva, Donatella Maraschin, David Young, John Quinn V, John Carpentier, Johan Allouche & Evangelos Himonides. 2020. Using Beatboxing for Creative Rehabilitation After Laryngectomy: Experiences From a Public Engagement Project. Frontiers in Psychology 10. 2854. https://doi.org/10.3389/fpsyg.2019.02854.

Ohala, John J. 1990. There is no interface between phonology and phonetics: a personal view. Journal of Phonetics 18(2). 153–171. https://doi.org/10.1016/S0095-4470(19)30399-7.

Paroni, Annalisa, Nathalie Henrich Bernardoni, Christophe Savariaux, Hélène Lœvenbruck, Pascale Calabrese, Thomas Pellegrini, Sandrine Mouysset & Silvain Gerber. 2021. Vocal drum sounds in human beatboxing: An acoustic and articulatory exploration using electromagnetic articulography. The Journal of the Acoustical Society of America 149(1). 191–206. https://doi.org/10.1121/10.0002921.

Pillot-Loiseau, Claire, Lucie Garrigues, Didier Demolin, Thibaut Fux, Angélique Amelot & Lise Crevier-Buchman. 2020. Le human beatbox entre musique et parole : quelques indices acoustiques et physiologiques. Volume ! (16 : 2 / 17 : 1). 125–143. https://doi.org/10.4000/volume.8121.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. John Wiley & Sons.

Ravignani, Andrea, Henkjan Honing & Sonja A. Kotz. 2017. Editorial: The Evolution of Rhythm Cognition: Timing in Music and Speech. Frontiers in Human Neuroscience 11. https://www.frontiersin.org/article/10.3389/fnhum.2017.00303. (6

June, 2022).
Rialland, Annie. 2005. Phonological and phonetic aspects of whistled languages. Phonology. Cambridge University Press 22(2). 237–271. https://doi.org/10.1017/S0952675705000552.
Saltzman, Elliot L. & Kevin G. Munhall. 1989. A Dynamical Approach to Gestural Patterning in Speech Production. Ecological Psychology. Routledge 1(4). 333–382. https://doi.org/10.1207/s15326969eco0104_2.
Santana, Carlos. 2016. What Is Language? Ergo, an Open Access Journal of Philosophy 3(20201214). https://doi.org/10.3998/ergo.12405314.0003.019.
Sapir, Edward. 1925. Sound Patterns in Language. Language 1(2). 37–51.
Sedarous, Yourdanis & Savithry Namboodiripad. 2020. Using audio stimuli in acceptability judgment experiments. Language and Linguistics Compass 14(8). https://doi.org/10.1111/lnc3.12377.
Schellenberg, Murray Henry. 2013. The Realization of Tone in Singing in Cantonese and Mandarin. Vancouver: The University of British Columbia.
Shaw, Patricia A. 2008. Scat syllables and markedness theory. Toronto Working Papers in Linguistics 27. 145–191.
Shih, Shu-hao & Paul De Lacy. 2019. Evidence for Sonority-Driven Stress. Catalan Journal of Linguistics 18. 9. https://doi.org/10.5565/rev/catjl.256.
Stevens, Kenneth N. 1989. On the quantal nature of speech. Journal of Phonetics 17(1–2). 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7.
Stevens, Kenneth Noble & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. Journal of Phonetics 38(1). 10–19. https://doi.org/10.1016/j.wocn.2008.10.004.
Stowell, Dan. 2003. The Beatbox Alphabet. The Beatbox Alphabet. http://www.mcld.co.uk/beatboxalphabet/. (27 April, 2022).
Stowell, Dan & Mark D. Plumbley. 2008. Characteristics of the beatboxing vocal style. Queen Mary, University of London.
Swissbeatbox. 2019. We Speak Music | Episode 4 | When Art Meets Therapy. https://www.youtube.com/watch?v=iS4LsXmZpHE. (23 January, 2023).
TyTe, Gavin & Mark SPLINTER. 2002/2014. Standard Beatbox Notation (SBN). HUMAN BEATBOX. https://www.humanbeatbox.com/articles/standard-beatbox-notation-sbn/. (10 September, 2021).
Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7(1). 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3.
Woods, K J. 2012. (Post)Human Beatbox Performance and the Vocalisation of Electronic and Mechanically (Re)Produced Sounds.
Wyttenbach, Robert A., Michael L. May & Ronald R. Hoy. 1996. Categorical Perception of Sound Frequency by Crickets. Science. American Association for the Advancement of Science 273(5281). 1542–1544.
Ziegler, Wolfram. 2003a. Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech. Aphasiology 17(1). 3–36. https://doi.org/10.1080/729254892.
Ziegler, Wolfram. 2003b. To speak or not to speak: Distinctions between speech and nonspeech motor control. Aphasiology 17(2). 99–105. https://doi.org/10.1080/729255218.