UNDERSTANDING DESIGN FEATURES OF MUSIC AND LANGUAGE: THE CHORIC/DIALOGIC DISTINCTION

FELIX HAIDUK[1], and W. TECUMSEH FITCH[1,2]

[*]Corresponding Author: felix.haiduk@univie.ac.at
[1]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria
[2]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

Music and spoken language share certain characteristics: both consist of sequences of acoustic elements that are combinatorically combined, and these elements partition the same continuous acoustic dimensions (frequency, formant space and duration). However, the resulting categories differ sharply: scale tones and note durations of small integer ratios appear in music, while speech uses phonemes, lexical tone, and non-isochronous durations. Why did music and language evolve to become the two systems we have today, differing in these specific features? Based on a reverse-engineering perspective on design features of language and music proposed by Hockett (1960) and Fitch (2006), we developed a framework suggesting that design features of prototypical music and language are a response to their differential deployment along three different continuous dimensions, which we call the goal dimension, the novelty dimension and the interactivity dimension. Also, we suggest that the hybrid nature of intermediate systems like poetry, chant, or solo lament follows from their deployment in the less typical context. At the core of all three dimensions are predictive processes about linguistic or musical sequences unfolding in time (see Koelsch et al., 2019; Kuperberg & Jaeger, 2016). Predictions involve uncertainty and surprise/ information which can be quantified using information theory (Shannon, 1948). Trajectories of information and uncertainty at multiple levels of the unfolding sequence thus form the basis of our theorizing.

(1) The goal dimension concerns the broader purpose of linguistic or musical sequence productions, whether to convey semantic messages, or to elicit and modulate aesthetic responses in a broad sense (see Huron, 2016). Both language and music can be deployed in propositional and aesthetic contexts, and similar responses follow: with more propositional goals, the multiple levels of the speech or musical sequence are more interdependent, and vary their information density to support successful inference of propositional content. For aesthetic goals, independent variation across levels enables more unconstrained variation in uncertainty and surprise, effectively exploiting the human reward system.

(2) The novelty-repetition dimension involves the repeatability or novelty of (groups of) elements and their relations at different scales and at multiple levels of musical or linguistic sequences. While language usually allocates repeatability to the phonological level and novelty to the morphosyntactic and semantic levels (related to propositionality), music typically allows both novelty and repeatability across all levels of the musical sequence, and meter seems to be especially crucial as a predictive layer throughout, enabling both prediction and surprise.

(3) We argue that the familiar goal and novelty dimensions alone are not sufficient to explain differences in design features between music and language: the interactivity between individuals is crucial. The interactivity dimension, the poles of which we term 'choric' and 'dialogic', concerns the temporal coordination of linguistic or musical productions of multiple participants. For dialogic contexts the only coordinative constraint concerns the timing of the turn-taking between individuals. Thus, we predict a lower information density and thus higher predictability towards the end of phrases, across all levels of the sonic stream, for both music and language. This aligns well with neural markers like the N400 (e.g. Grisoni et al., 2017), changes in speech rate at turn completion points in conversations (Walker, 2010; Wightman et al., 1992) or the notion of musical closure in harmony and melody (Huron, 2006). In contrast, choric performance requires tight temporal coordination of all contributing individuals, enabled by high predictability in timing and frequency of sonic events (Keller et al., 2014). Isochronous meter and discrete pitches in scales are design solutions that enable a group of participants to join in making a coherent sound sequence, allowing both novelty and repeatability. Simultaneous speaking of multiple talkers requires uttering of precisely the same words at the same time, as happens for example in religious chanting. Attention should then be much more focussed on coordination than in dialogic speech acts, an isochronic and/or metrical scaffolding should develop (cf. Bowling et al., 2013), and body motion, facial expressions or prosodic intonation should be more pronounced in a spoken choric context.

Our framework avoids an overly simplistic dichotomy between language and music by also encompassing non-canonical forms of music and language like chant, poetry, or exchange of musical solos. It supports comparisons of different forms of communication across distinct modalities and can help to generate new hypotheses about optimal design of signals satisfying multiple different requirements. We hope that it will also be fruitfully employed in animal communication research, especially for species engaging in chorusing or duetting, such as various bird species, dolphins, bats, or gibbons, broadening the scope of comparisons with music and/or language.

Bowling, D. L., Herbst, C. T., & Fitch, W. T. (2013). Social origins of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS ONE*, *8*(11), e80402. https://doi.org/10.1371/journal.pone.0080402

Fitch, W. T. (2006). The biology and evolution of music : A comparative perspective. *Cognition*, *100*, 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *The Journal of Neuroscience*, *37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Huron, D. (2006). *Sweet Anticipation*. Cambridge, Massachusetts: MIT Press.

Huron, D. (2016). Aesthetics. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (2 ed., pp. 233–245). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.19

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130394. https://doi.org/10.1098/rstb.2013.0394

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Walker, G. (2010). The phonetic constitution of a turn-holding practice: rush-throughs in English talk-in-interaction. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), *Prosody in Interaction. Studies in discourse and grammar (23).* (pp. 51–72). Amsterdam, Philadelphia: John Benjamins Publishing Company. https://doi.org/https://doi.org/10.1075/sidag.23

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717.