Music Cognition: Theory Testing and Model Selection

Henkjan Honing (honing@uva.nl)
Music Cognition Group (www.hum.uva.nl/mmm)
Institute for Logic, Language and Computation (ILLC), University of Amsterdam
Nieuwe Doelenstraat 16, 1012 CP Amsterdam, NL

Introduction

How should we select among computational models of cognition? This is a question that has attracted quite some discussion recently. While the most common way of evaluating a computational model is showing a good fit with the empirical data, the discussion addresses the problems that might arise with the assumption that this is actually strong evidence for a model. Some authors consider a good fit between a theory and the empirical observations a good and necessary starting point but clearly not the end point of model selection or verification (e.g., Rodgers & Rowe, 2002). Others suggest alternatives to a goodness-of-fit measure, such as preferring the simplest model, both in terms of its functional form and number of free parameters (e.g., Pitt & Myung, 2002). Yet others propose to prefer a theory that predicts an empirical phenomenon that was least expected, considering a good fit of less relevance or even misleading (e.g., Roberts & Pashler, 2000).

Case Study in Model Selection

However, the aim of this paper is not to add to this lively debate in a philosophical or methodological sense. Instead, it will focus on a specific problem from music cognition, i.e., modeling the temporal aspects of music (Longuet-Higgins, 1987; Desain, Honing et al., 1998). It presents a *case study* on how one can select between one and another computational model, informed by the methodological discussion mentioned in the introduction.

Two families of computational models will be compared. The first takes a kinematic approach (*K-model*; Honing, 2003) to the modeling of expressive timing in music performance: what timing patterns are commonly found in music performance and how do they conform to the laws of physical motion. This approach will be contrasted with a perceptual approach (*P-model*; Honing, 2005) that predicts the amount of expressive freedom a performer has in the interpretation of a rhythmic fragment (cf. Desain & Honing, 2003). The two approaches will be compared using three different model selection criteria: goodness-of-fit, model's simplicity, and the amount of surprise in the predictions.

Conclusion

While both models fit the empirical data equally well, in the light of what accounts as strong evidence for a model, i.e. making precise (constrained), non-smooth, and relatively surprising predictions (cf. Roberts & Pashler, 2000), the perception-based model is preferred over the kinematic model, however simpler and natural the latter model might seem. (For a full paper on this topic, see Honing, 2004).

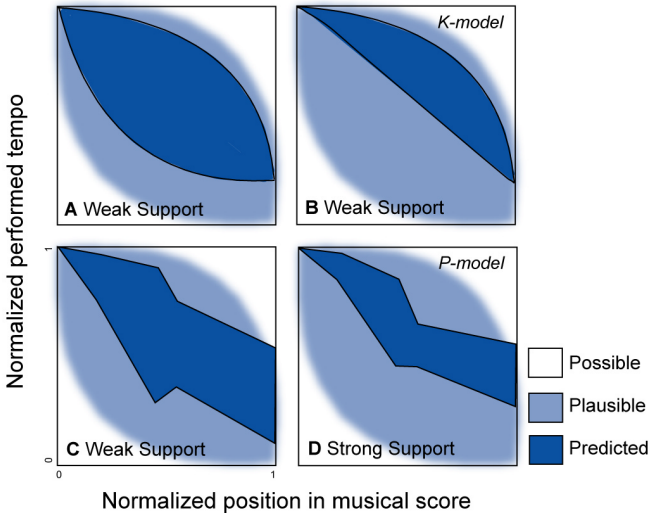

Figure 1: Schematic diagram of strong and weak support for a model of *ritardandi* in music performance (adapted from Honing, 2004).

References

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341-365.

Desain, P., Honing, H., Thienen, H. van, & Windsor, W. L. (1998). Computational modeling of music cognition: Problem or solution? *Music Perception*, 16(1), 151-166.

Honing, H. (2003). The final ritard: on music, motion, and kinematic models. *Computer Music Journal*, 27(3), 66-72.

Honing, H. (2004). Computational modeling of music cognition: a case study in model selection. *ILLC Prepublication* PP-2004-14. [http://dare.uva.nl/en/record/146627]

Honing, H. (2005, in press). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception*. [http://dare.uva.nl/en/record/120423]

Longuet-Higgins, H.C. (1987). *Mental Processes*. Cambridge, Mass.: MIT Press.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Science*, 6(10), 421- 425.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367.

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000). *Psychological Review*, 109(3), 599–604.