



HAL
open science

Reply to Knight et al.: The complexity of inferences from speech prosody should be addressed using data-driven approaches

Emmanuel Ponsot, Juan Jose Burred, Pascal Belin, Jean-Julien Aucouturier

► To cite this version:

Emmanuel Ponsot, Juan Jose Burred, Pascal Belin, Jean-Julien Aucouturier. Reply to Knight et al.: The complexity of inferences from speech prosody should be addressed using data-driven approaches. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115 (27), pp.E6104-E6105. 10.1073/pnas.1806857115 . hal-02481125

HAL Id: hal-02481125

<https://hal.science/hal-02481125v1>

Submitted on 17 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REPLY TO KNIGHT ET AL.: The complexity of inferences from speech prosody should be addressed using data-driven approaches

Emmanuel Ponsot^{a,b,c,1}, Juan Jose Burred^d, Pascal Belin^{e,f,g}, and Jean-Julien Aucouturier^{a,h}

^aSciences et Technologies de la Musique et du Son, UMR 9912, Institut de Recherche et Coordination Acoustique/Musique, CNRS and Sorbonne Université, 75004 Paris, France;

^bLaboratoire des Systèmes Perceptifs, UMR 8248, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France;

^cDépartement d'Études Cognitives, Ecole Normale Supérieure, Paris Sciences et Lettres Research University, 75005 Paris, France;

^dIndependent Researcher, 75013 Paris, France;

^eInstitut de Neurosciences de la Timone, UMR 7289, Centre National de la Recherche Scientifique and Aix-Marseille Université, 13007 Marseille, France;

^fInstitute of Neuroscience and Psychology, University of Glasgow, Glasgow G128QQ, United Kingdom;

^gDépartement de Psychologie, Université de Montréal, Montréal, QC H3T 1J4, Canada; and

^hOkanoya Emotional Information Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Wako, Saitama 332-0012, Japan

We are glad our proposed methodological approach (1) raises interest in the community. Knight et al. (2) make two important theoretical considerations that we would like to further develop here.

The first point they raise concerns the specificity of the pitch prototype of dominance/trustworthiness: They argue one should demonstrate that these prototypes are specific (i.e., not shared by other emotional or linguistic traits). Dominance and trustworthiness, while difficult to explicitly define, are not vague concepts: As is the case for social faces in vision (3), they constitute the two principal dimensions of the social space experimentally derived for speech with two-syllable utterances (4). These traits are therefore robust and similar enough between subjects to emerge as main dimensions. The one-dimensional intonation prototypes we derived support this, with striking similarities across participants. We agree that an important goal of further experimental studies is to understand the exact relationship between social traits and the multidimensional space of speech prosody in its full complexity (pitch contour, but also variations of, for example, intensity, speech rate, and timbre). However, we believe that there is no theoretical need that the prosodic map of social dimensions be fully specific. This question can be addressed experimentally thanks to data-driven methodologies such as the one we proposed: Cracking the code of speech prosody underlying the emotional or linguistic spaces will provide a unique possibility for quantitatively investigating the similarities between emotional, social, and linguistic representations, shedding further light on the mechanisms of human communication.

Knight et al.'s (2) second concern relates to the generalizability of the prototypes across multiple utterances, speakers, and contexts. Our point of view is that targeting a “generic code” does not imply single, invariant, and linear prototypes for each trait. Considering the cognitive processes underlying the formation of social impressions from speech prosody as a linear operation (template matching) is clearly a coarse approximation. However, deriving prototypes in this framework allows us to examine, in a second step, the limits of this model. Looking for the common roots of speech prosody does not imply that a unique code is implemented across all speech signals.

Speech is by essence a highly complex and variable signal, such that certain social traits may interact with other, for example, emotional, traits. Therefore, we encourage future studies—notably by sharing our tool in open source (forumnet.ircam.fr/product/cleese/)—to investigate these next questions: Which social traits are robust, which depend more on the underlying content, and how so? Our first results indicate that the pitch prototype of dominance is robust across other utterances and speakers, but that the trustworthiness prototypes derived on “hello” do not generalize as well to other utterances and speakers. Whether a single or multidimensional prototype is at play for these and other prosodic dimensions remains to be established.

Reverse-correlation studies involving large datasets (many different utterances and large sample size) will in particular provide a powerful means to make explicit the causal link between prosodic variations in speech and the concept of, for example, trustworthiness in its global ecological complexity.

Acknowledgement: Work funded by ERC project CREAM 335536 (to JJA.)

- (1) Ponsot E, Burred JJ, Belin P, Aucouturier J-J (2018) Cracking the social code of speech prosody using reverse correlation. *Proc Natl Acad Sci USA* 115:3972–3977.
- (2) Knight S, Lavan N, Kanber E, McGettigan C (2018) The social code of speech prosody must be specific and generalizable. *Proc Natl Acad Sci USA* 115:E6103.
- (3) Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. *Proc Natl Acad Sci USA* 105:11087–11092.
- (4) McAleer P, Todorov A, Belin P (2014) How do you say ‘hello’? Personality impressions from brief novel voices. *PLoS One* 9:e90779