Towards Singing Perception Universals

Polina Proutskova
Queen Mary, University of London
proutskova@googlemail.com

1. INTRODUCTION

How do we talk about singing? When describing a difference between two performers singing the same song, we may mention the mood they create, their skill, list musical choices they make (tempo, dynamics, etc.). Yet when we try to characterise what they sound like, we are left to invoking metaphors and comparisons to other areas of life. It seems that there is no widely understood vocabulary about vocal production in singing, not even within an established and theorised tradition such as Western classical music (Proutskova, 2018; Garnier et al., 2007; McGlashan, 2013). Or is there? If we learn that someone's singing is somewhat husky or really twangy, wouldn't we understand similarly? Are there any terms we would interpret in the same way?

2. DEFINITION AND IMPACT

I would like to introduce a notion of a *singing perception universal* – a descriptor of vocal production that would be understood similarly by a large group of people. In contrast to musical universals (Brown and Jordania, 2013; Harwood, 1976), *singing perception universals* are not about being present or absent in a large number of cultures; rather they are about being recognised and interpreted similarly e.g. by people coming from different cultures. The advantage of this approach is that it does not rely on a "correct" description of a vocalisation; it merely measures the agreement between listeners' judgements.

Singing is the most widespread artistic activity (e.g. 37 million choral singers in Europe, [Bartel and Cooper, 2015]) yet it is one that is less well understood. The lack of vocabulary impedes further research. In MIR, research on singing recordings lags far behind other types of data, due to lack of annotations. But which characteristics should be annotated in absence of commonly understood vocabulary? With the advance of deep learning, there is a possibility that automatic classification and computer generation of vocals would soon emerge from the leading digital music content holders. Yet their models will be biased in the same way as their data: if they possess pop music alone, only pop music will be generated. In order to represent the whole variety of human vocalisation in AI models, we should be looking for ways to capture that variety. Understanding how we understand singing would help us to grasp how a consensus is formed about matters which are considered difficult to verbalise, having a "mystery" about them. Also, in previous studies vocal production has been linked to societal traits such as status of women in a society. Such hypotheses cannot be independently validated without a more rigorous approach to semantic formalisation.

3. PREVIOUS WORK, PLAUSIBILITY

The motivation to this strand of investigation comes from my PhD in which I intended to formalise the language about singing on the basis of objective physiological traits (Proutskova, 2018). In my investigative study 13 experts annotated singing samples from a wide range of cultures using my ontology of vocal production: there was a tendency to agreement for only 2 out of 19 dimensions they were rating. This result indicates that experts cannot generally hear physiology underlying vocal sounds. Yet there were two descriptors about which they agreed – *larynx height* and *narrowness of the vocal tract* (often referred to by its main vocal function – the *twang*). These two descriptors could possibly be easier to agree about for vocal production experts and therefore constitute *singing perception universals* for this group, though further studies with more participants would have to confirm that.

While there is little research on cross-cultural vocal production, one experiment stands out – the Cantometrics project. Alan Lomax and Victor Grauer designed a parametrisation system for cross-cultural singing performance practice, consisting of 36 parameters, some of which reflect vocal production (Lomax, 1977). They chose characteristics for a rough subjective description of a singing sample, which were self-explanatory or easy to teach to non-experts: volume, rasp, nasality, accent, enunciation, vocal blend, glottal shake, etc. In their experiment they had 5000 singing recordings rated by three raters each, and the agreement between raters was good.

Why was the Cantometrics approach more successful in finding consensus between human listeners than my study? Cantometrics used subjective, perceptual descriptors rather than physiological ones. That allowed non-specialists, not familiar with vocal physiology, to become raters. In fact research in voice pathology confirms that experts tend to disagree more than amateurs when analysing voices (Kreiman et al., 1993).

For example, in Cantometrics, *vocal width* is the dimension stretching between a wide, open, relaxed and reso-

nant sound on the one hand, and on the other hand a narrow, tense squeezed singing. I demonstrated that this definition from the physiological point of view was flawed, lumping together three distinct dimensions of the vocal sound which are ambiguous and not well correlated: narrow/wide, constricted/relaxed and more resonant/less resonant (Proutskova, 2018). Then how was it possible that Lomax found his raters agreeing about *vocal width* while I observed no agreement on its dimensions in my study? Was it due to the fact that a general look at a wider phenomenon, without going into much detail, allows for a better consensus? Or did it reflect the extensive training procedure that all Cantometrics raters were subject to?

Related to Cantometrics *vocal width* is the notion of *open throat*, which is widely referred to in Western countries, associating open throat vocalisation with a big, relaxed and aesthetically pleasing sound, whereas the absence of open throat is often considered as incorrect singing. Mitchell et al. (2003) showed that singing teachers project different meanings on the term. Yet the term persists – could that be an indication of a general consensus about a more rough definition of an *open throat*?

4. METHOD

I suggest to approach the choice of possible candidates for *singing perception universals* from two different angles. Firstly, the terms which in previous studies were shown to lead to agreement among raters should be independently investigated: the Cantometrics descriptors (Lomax, 1977), *larynx height* and *twang* from my study (Proutskova, 2018), widely used terms from established vocal schools (see e.g. Granier, 2007, for an analysis of Western classical vocabulary). For these terms an experiment in the form of an online game should be developed, collecting participants' ratings alongside their cultural background and singing proficiency. The agreement can then be examined for a variety of demographic groups.

Secondly, a similar online game approach can be used for an agnostic study, in which semantic dimensions are not pre-defined. In contrast to absolute ratings on a scale collected in the first case, here participants will be asked to provide relative similarity estimations of singing samples through odd-one-out choice (triadic comparisons, see Farrugia et al., 2016). Multidimensional scaling (Weller and Romney, 1988) can then be used to extract the most salient factors accounting for dissimilarity of singing samples. These factors would be the perfect candidates for *singing perception universals*. The advantage of this approach is that no prior knowledge of semantic categories is assumed and no training of participants is required. The disadvantage lies in the fact that many more ratings will be required.

Some study design questions will have to be examined beforehand to achieve transparency and replicability: e.g. whether raters should be pre-trained for the semantic categories they are supposed to rate. The training can anchor them to the same scale (zero and extremes) and thus boost agreement. Yet over-training can lead to "overfitting".

5. CONCLUSION

Should *singing perception universals* be found, they might prove to be the key to analysing vocal style and its change in time and space, the geographic spread may provide insights into human migration and cultural evolution (Grauer 2006); they will also inform our singing perception models. If, on the other hand, no such characteristics emerge, vocal production would make a perfect test case for automated methods to surpass human classification ability; and the singing voice would retain the aura of mystery continuing to enthrall us with its versatility and expressivity, finding its way directly to our hearts.

6. REFERENCES

Bartel, R. and Cooper, C. F. (2015). Singing Europe. Technical report, European Choral Association.

Brown, S. and Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, 41(2):229–248.

Farrugia, N., Allan, H., Müllensiefen, D., and Avron, A. (2016). Does it sound like progressive rock? a perceptual approach to a complex genre. In Gonin, P., editor, *Prog Rock in Europe: Overview of a persistent musical style*, pages 197–212. Editions Universitaire, Dijon.

Garnier, M., Henrich, N., Castellengo, M., Sotiropoulos, D., and Dubois, D. (2007). Characterisation of voice quality in western lyrical singing: from teachers' judgements to acoustic descrip- tions. *Journal of interdisciplinary music studies*, 1(2):62–91.

Grauer, V. A. (2006). Echoes of our forgotten ancestors. The World Of Music, 48(2).

Harwood, D. L. (1976). Universals in music: A perspective from cognitive psychology. *Ethnomusicology*, 20(3):521–533.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech, Language, and Hearing Research*, 36(1):21–40.

Lomax, A. (1977). Cantometrics: A Method of Musical Anthropology (audio-cassettes and handbook). Berkeley: University of California Media Extension Center.

McGlashan, J. (2013). What descriptors do singing teachers use to describe sound ex- amples? *Presented at PEVOC 10 (Pan-European Voice Conference)* Prague, Czeck Republic.

Mitchell, H. F., T. Kenny, D., Ryan, M., and Davis, P. J. (2003). Defining 'open throat' through content analysis of experts' pedagogical practices. *Logopedics Phoniatrics Vocology*, 28(4):167–180.

Proutskova, P. (2018). Investigating the Singing Voice: Quantitative and Qualitative Approaches to Studying Cross-Cultural Vocal Production. PhD thesis, Goldsmiths University of London.

Weller, S. C. and Romney, A. K. (1988). Systematic data collection, volume 10. Sage publications.