



Technological University Dublin
ARROW@TU Dublin

Papers

6th International Workshop on Folk Music
Analysis, 15-17 June, 2016

2016-6

Formalising Cross-Cultural Vocal Production

Polina Proutskova

Goldsmiths College, University of London, proutskova@googlemail.com

Christophe Rhodes

Goldsmiths College, University of London, c.rhodes@gold.ac.uk

Tim Crawford

Goldsmiths College, University of London, t.crawford@gold.ac.uk

Geraint Wiggins

Queen Mary University of London, geraint.wiggins@gmul.ac.uk

Follow this and additional works at: <https://arrow.tudublin.ie/fema>

 Part of the [Musicology Commons](#)

Recommended Citation

Proutskova, P., Rhodes, C., Crawford, T., Wiggins, G. (2016). Formalising cross-cultural vocal production. *6th International Workshop on Folk Music Analysis*, Dublin, 15-17 June, 2016.

This Conference Paper is brought to you for free and open access by the 6th International Workshop on Folk Music Analysis, 15-17 June, 2016 at ARROW@TU Dublin. It has been accepted for inclusion in Papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#)



FORMALISING CROSS-CULTURAL VOCAL PRODUCTION

Polina Proutskova

Goldsmiths, University
of London

proutskova@googlemail.com

Christophe Rhodes

Goldsmiths, Univer-
sity of London

c.rhodes@gold.ac.uk

Tim Crawford

Goldsmiths, Univer-
sity of London

t.crawford@gold.ac.uk

Geraint Wiggins

Queen Mary Univer-
sity of London

geraint.wiggins@
qmul.ac.uk

1. INTRODUCTION

How do we speak about the timbre of a singer? How do we compare singers singing the same song? It wouldn't be particularly hard to distinguish a Chinese opera singer from a Western opera singer, but it would be much harder to verbalize how we distinguish them. And when a classical singer performs a rock song, we all hear it is stylistically wrong, but how do we explain to the singer what he needs to change?

All these questions are about vocal production and how it can be captured in words. As it currently stands there is no widely accepted vocabulary to talk about it, not even within a single culture or genre (Garnier, 2007; Mitchell, 2003). Publications in English analysing vocal production in other cultures are rare (Födermayr, 1971; Bartmann 1994). Singing teachers very often use idiosyncratic language based on their subjective perception or learnt from their own teachers, it is hard for teachers from different schools to agree about the terms (McGlashan, 2013). Medical professionals are mainly interested in vocal dysfunction (Little, 2009). Ethnomusicologists focus on the context of music making and rarely mention the sound itself; while for musicologists or music critics it is considered a virtue to use unique terms specific to the particular writer and objectivity of language is not a priority.

We became interested in the subject in the context of MIR, hoping to train a computational model to classify vocal production. Applications would include: differentiating recordings of singing from different cultures; singer recognition; distinguishing originals from covers and covers by different singers; genre classification, etc. All these tasks have been addressed by brute force computational algorithms and by more sophisticated approaches (Tsai 2006, Serra 2010, Holzapfel 2008). Yet there seems to be a glass ceiling of classification accuracy that can be achieved (Karydis 2010, Downie 2008). In MIR it is referred to as “semantic gap” (Wiggins, 2009). If a middle layer could be introduced of more objective categories where further human knowledge is incorporated in the model, that could help improve classification accuracy further.

2. MODELS OF VOCAL PRODUCTION

There is no theoretical model of vocal production which could provide the basis for predictions. There are no annotated datasets either. As we have seen above, there isn't even a vocabulary to talk about vocal production. We have found only three approaches to parametrising vocal production that have had a wider reach: one originating in ethnomusicology, another coming from vocal education and one formulated in singing voice science.

Ethnomusicological parametrization was introduced by Alan Lomax in his Cantometrics experiment in which over 5000 recordings from more than 500 cultures were analysed, performance practice was expressed via 36 parameters (Lomax 1976). 13 of these parameters were related to vocal production, including volume, rasp, vocal tension, glottal shake, nasality, vocal pitch, etc. Lomax took an auditory-perceptual approach: human listeners were trained to rate the value of each parameter after listening to an audio recording. Lomax tried to diversify the ratings by getting at least three people to rate each recording. But his raters were mainly US university students with similar life experiences and musical backgrounds. A proper diversification would include people of all ages and professions, from different cultures and with varying musical experience. It is a much bigger undertaking and would have been unworkable in Lomax's circumstances. Only if it were conducted this way though would we be able to say with certainty whether Cantometrics musical parameters are perceived similarly independently of cultural and musical background.

Johan Sundberg, the father of singing voice science, introduced phonation modes describing the voice source aspect of vocal production (Sundberg 1979). They are based on the relationship between subglottal pressure and transglottal airflow. Three of his phonation modes - breathy, neutral and pressed - are widely used by speech and language therapists and in other fields. Sundberg formalised the terms relating them to the aerodynamic processes from which each of the modes originates. He suggests ways to infer phonation mode from an audio recording of singing via inverse filtering. This model works on a milliseconds scale but becomes unmanageable on a seconds scale, which is necessary for humans to recognise music and to feel something about it or deduct its characteristics – the time scale on which the Cantometrics experiment was conducted. Sundberg's phonation model does not include the resonance body aspect, which is crucial for resulting timbre.

Jo Estill was an American singer, teacher and voice researcher, who suggested a physiology-based system for understanding and teaching vocal production. Her idea was to isolate physiological structures, learn to manage them independently and use these building blocks of vocal physiology to construct various kinds of vocal production, ultimately leading to the ability to build any singing style (Estill, 1979; Colton, 1981). While her scientific evidence was partial at best, her work has had a huge influence on contemporary singing education (Sadolin, 2000; Soto-Moretini, 2006; Kayes, 2004).

Since we could not verify the inter-personal and inter-cultural consistency of Cantometrics approach we concentrated on the physiology including phonation. We devised an ontology of vocal production based on Sundberg's and Estill's terminology with some minor additions (Table 1).

physiological dimensions	range	scale	metrics
subglottal pressure	low to high	5-point	interval
transglottal airflow	low to high	5-point	interval
phonation breathy	present/absent	2-point	nominal
phonation pressed	present/absent	2-point	nominal
phonation neutral	present/absent	2-point	nominal
phonation flow	present/absent	2-point	nominal
vocal folds modal vs. falsetto	modal/falsetto	2-point	nominal
vocal folds vibration mode thick to thin	thick/mixed thicker/mixed/mixed thinner/thin	9-point	interval
larynx height	low to high	9-point	interval
thyroid cartilage tilt	vertical/vertical slight tilt/tilted	5-point	interval
cricoid cartilage tilt	vertical/vertical slight tilt/tilted	5-point	interval
velum	low to high	5-point	interval
aryepiglottic sphincter (size of vocal tract)	wide to narrow	5-point	interval
tongue height	low to high	5-point	interval
tongue compression	present/absent	2-point	nominal
position within chest register	low to high	5-point	interval
position within head register	low to high	5-point	interval

Table 1. Our ontology of vocal production.

3. THE STUDY

The aim of our study is to assess the viability of the physiological approach to modelling vocal production as well as to verify applicability and usefulness of our preliminary ontology of vocal production (Table 1). The study is based on interviews with vocal physiology experts and combines a qualitative and a quantitative approach (Bryman, 2006).

We chose eleven tracks from the Cantometrics dataset (see Chapter on vocal width in Lomax, 1977), all from different musical cultures. Nineteen physiologically stable fragments were extracted from the tracks, which were then used as entities of analysis in the interviews. We recruited 13 participants: otolaryngologists, speech language therapists, singing teachers. Participants' professional involvement with vocal physiology ranged from 10 to over 40 years. Three of them had a non-Western cultural background.

Interviews were structured and lasted from 90 minutes to several hours. Participants were asked to rate physiological dimensions from the preliminary ontology with which they were familiar; they were encouraged to explain their ratings, to point out complexities, to suggest better terms and approaches.

4. RESULTS

Participants showed confidence in the majority of terms introduced in the preliminary ontology: only 20% of physiological dimensions were rated by less than 80% of participants. While experts generally supported the ontology, the inter-participant agreement on the ratings was low. Only for two descriptors – position of the larynx and AES – was there a tendency to agreement.

In this talk we shall present the results of the qualitative analysis of the interviews, the analysis of inter-participant (dis)agreement including problem cases and searching for possible causes. We shall demonstrate using the words of our participants how some common themes have emerged from the interviews and how these findings could explain the disagreement. The advantages and disadvantages of physiological vs perceptual approaches to vocal production as well as their possible combinations will be discussed. We shall outline future research directions for this largely understudied area and explain the significance of our findings for academic and applied fields outside MIR.

5. REFERENCES

- Bartmann, M. (1994). Rauhigkeiten in der Volksmusik in der Kanarischen Insel El Hierro. In *Böckner, M., editor, Berichte aus dem ITCM- Nationalkomitee Deutschland*, volume 3, Bamberg.
- Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative research*, 6(1):97–113.
- Colton, R. H. and Estill, J. A. (1981). Elements of Voice Quality: Perceptual, Acoustic, and Physiologic Aspects. In *Speech and language: advances in basic research and practice*, Lass, Norman J. (ed.), volume 5, pages 311–403. Academic Press, New York.
- Downie, J. S. (2008). The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. *Acoustical Science and Technology*, 4(29):247–255.
- Estill, J. and Colton, R. (1979). The identification of some voice qualities. *The Journal of the Acoustical Society of America*, 65(S1).
- Födermayr, F. (1971). Zu gesanglichen Stimmgebung in der außereuropäischen Musik. Ein Beitrag zur Methodik der vergleichenden Musikwissenschaft. Stieglmayr, Wien.
- 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 descriptions. *Journal of interdisciplinary music studies*, 1(2):62–91.
- Holzappel, A. and Stylianou, Y. (2008). Musical genre classification using nonnegative matrix factorization-based features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2):424–434.
- Karydis, I., Radovanovic, M., Nanopoulos, A., and Ivanovic, M. (2010). Looking through the “glass ceiling”: A conceptual framework for the problems of spectral similarity. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*.
- Kayes, G. (2004). *Singing and the Actor*. Routledge.

- Little, M., et. al. (2009). Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. *Journal of Voice*.
- Lomax, A. (1968). *Folk Song Style and Culture*. Transaction Books, New Brunswick, New Jersey.
- 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 examples? Presented at *PEVOC 10 (Pan-European Voice Conference)* Prague, Czech 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.
- Sadolin, C. (2000). *Complete vocal technique*. Shout Publishing Copenhagen, Denmark.
- Serra, J., Gomez, E., and Herrera, P. (2010). Audio cover song identification and similarity: background, approaches, evaluation, and beyond. In *Advances in Music Information Retrieval*, pages 307–332. Springer.
- Soto-Moretini, D. (2006). *Popular Singing: A Practical Guide To: Pop, Jazz, Blues, Rock, Country and Gospel*. A&C Black.
- Sundberg, J. (1987). *The science of the singing voice*. Illinois University Press.
- Tsai, W.-H. and Wang, H.-M. (2006). Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(1):330–341.
- Wiggins, G. (2009). Semantic gap?? schemantic schmap!! methodological considerations in the scientific study of music. *Proceedings of IEEE AdMIRe*.