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.dit.ie/fema

Part of the Musicology Commons

Recommended Citation
FORMALISING CROSS-CULTURAL VOCAL PRODUCTION

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 even 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 parametrizing 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 300 cultures were analysed, performance practice was expressed via 36 parameters (Lomax 1976). Of these parameters were related to vocal production, including volume, rasps, 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 deduce 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-Morettini, 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).

<table>
<thead>
<tr>
<th>Physiological dimensions</th>
<th>Range</th>
<th>Scale</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subglottal pressure</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Transglottal airflow</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Phonation breathy</td>
<td>Present/absent</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Phonation pressed</td>
<td>Present/absent</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Phonation neutral</td>
<td>Present/absent</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Phonation flow</td>
<td>Present/absent</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Vocal folds modal vs. falsetto</td>
<td>Modal/falsetto</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Vocal folds vibration</td>
<td>Thicker/thinner/mixed</td>
<td>9-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Mode thick to thin</td>
<td>Xed/thinner/thin</td>
<td>9-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Larynx height</td>
<td>Low to high</td>
<td>9-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Thyroid cartilage tilt</td>
<td>Vertical/slight tilt/tilted</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Cricoid cartilage tilt</td>
<td>Vertical/slight tilt/tilted</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Velum aryepiglottic s.</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Tissue (size of vocal tract)</td>
<td>Wide to narrow</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Tongue height</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Tongue compression</td>
<td>Present/absent</td>
<td>2-point</td>
<td>Nominal</td>
</tr>
<tr>
<td>Position within chest register</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
<tr>
<td>Position within head register</td>
<td>Low to high</td>
<td>5-point</td>
<td>Interval</td>
</tr>
</tbody>
</table>

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 (Bryan, 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


