VOCALS IN MUSIC MATTER: THE RELEVANCE OF VOCALS IN THE MINDS OF LISTENERS

Andrew Demetriou[1,2] **Andreas Jansson**[2] **Aparna Kumar**[2] **Rachel M. Bittner**[2]
[1] Multimedia Computing Group, TU Delft, The Netherlands
[2] Spotify Inc., New York City, USA

ABSTRACT

In music information retrieval, we often make assertions about what features of music are important to study, one of which is vocals. While the importance of vocals in music preference is both intuitive and anticipated by psychological theory, we have not found any survey studies that confirm this commonly held assertion. We address two questions: (1) what components of music are most salient to people's musical taste, and (2) how do vocals rank relative to other components of music, in regards to whether people like or dislike a song. Lastly, we explore the aspects of the voice that listeners find important. Two surveys of Spotify users were conducted. The first gathered open-format responses that were then card-sorted into semantic categories by the team of researchers. The second asked respondents to rank the semantic categories derived from the first survey. Responses indicate that vocals were a salient component in the minds of listeners. Further, vocals ranked high as a self-reported factor for a listener liking or disliking a track, among a statistically significant ranking of musical attributes. In addition, we open several new interesting problem areas that have yet to be explored in MIR.

1. INTRODUCTION

The Music Information Retrieval (MIR) community has historically focused on content-based understanding of music. The type of content-based analysis studied over time is typically driven by the data available to the task, or the interests of the specific researchers. An alternative motivator could be to study topics that are salient in the minds of listeners, especially with respect to listener's musical preference. Specifically, understanding which attributes of music contribute the most to music preference, and their relative weight, could help guide research efforts. One attribute of music we would expect to be salient in the minds of listeners is the singing voice.

Psychology research anticipates the importance of the human voice as a salient stimulus, and as a component of music in particular. The human ability to communicate exceeds that of any other species studied thus far, with both speech and singing being cultural universals reliant on vocal production. It is theorized that the advanced human ability to communicate, discriminate, and to experience emotional responses in vocalizations has allowed for the emergence of music [8]. Our emotions are often accompanied by involuntary changes in our physiology and nonverbal expressions, such as facial expressions and vocalizations [15]. Our reactions to the emotional content expressed in the vocals in music may have similar effects. As such, much psychological research has focused on the singing voice even more than speech, due to the precision required to execute and process musical vocalizations [5]. This makes musical vocals a well-anticipated candidate for study as a feature of music, as we would expect people to have a sophisticated ability to deliver, empathize with, and process vocal communications.

We would therefore expect that the vocals in music would be an especially salient component, if not the most salient. While a complete review is beyond the scope of this paper, some research is particularly worth noting. For example, it has been shown that both adults [18] and children [17] recall melodies more correctly when sung with the voice than when played with instruments. Hutchins and Moreno [5] review literature that shows relatively precise perception of pitch in the human voice, yet fewer noticeable pitch errors in the voice relative to musical instruments or synthesized voices [6]. Neuroscience studies show specific areas of the brain involved in processing human voices [2]. Although similar regions of the brain are involved in processing both music and voices, there is differential processing of the human voice relative to music [1]. As such, the human voice may be processed as a uniquely significant sound.

However, while prior research suggests that vocals would be especially relevant to music preference, no study to our knowledge has assessed the importance of the voice in music, relative to other musical components. To address this gap, we test the hypothesis that the voice is as or more important than other musical components across implicit and explicit datasets, using traditional social science techniques, as well as data mining techniques. First, we mine data available from Spotify, including playlist titles, search data and artist biographies, to test whether terms related to vocals are prevalent. However, we show that the results of the data mining are inconclusive as to whether or not

vocals are salient in the minds of listeners. Specifically, it is not clear whether the vocals can be disentangled from other factors in playlist titles and search queries, such as genre. For more conclusive results, we gather data from users explicitly. To this aim we conduct two online survey studies: the first gathered subjective data on the salient components of music directly from listener reports, which were separated into semantic categories using card-sorting. The second asked participants to rank the semantic categories from the first study in terms of importance to their musical preference. We conclude that two aspects related to the voice are especially salient, namely the voice itself, and the lyrics of the song. Furthermore, we highlight the importance of gathering explicit data to complement implicit techniques, in situations where factors may not be easily disentangled.

2. VOCALS IN SEMANTIC DATA

Prior research has shown that semantic descriptors of music may be an appropriate means for users to query music databases [12]. Given the large amount of semantic data available to Spotify such as playlist titles, search results, and artist biographies, one might hypothesize that terms describing the vocals would commonly appear in this implicit data.

2.1 Playlist Tags and Search Queries

Non-common words or groups of words and emojis appearing in the titles of a large number of Spotify's user-generated playlists were aggregated to create a list of the 1000 most frequently occurring *tags*. Each of these 1000 tags was assigned a category by a professional curator based on the tag itself and information from the tracks most frequently associated with the tag. The categories, determined by the curator, were Genre (e.g. "K-Pop"), Mood (e.g. "sad"), Activity (e.g. "gym"), Popularity (e.g. "Today's hits"), Artist (e.g. "Justin Timberlake"), Era (e.g. "70's"), Culture (e.g. "Latin"), Lyrics (e.g. "clean"), Rhythm (e.g. "groove"), Instrument (e.g. "guitar"), Tempo (e.g. "slow"), Voice (e.g. "female singers"), or Other (e.g. "favorites", "Jenna", "hi"). The percentage of playlists containing each of these tag categories is displayed in Figure 1, top.

Surprisingly, we see that tags explicitly related to vocals are not at all common compared to other types of tags, with the most common tags being related to genre, mood, or activity. Playlist titles can be viewed as labels for groups of music, and this analysis suggests that people do not often label groups of music based on explicit characteristics of the vocals. However, specific vocal characteristics (as well as many other musical attributes) may be implicit in many of the other tag categories, particularly for genre, mood, and artist. As vocal delivery style and genre are closely related, emotions communicated by the voice and the mood of the collection of songs may be related, and as each artist has a unique voice, we conclude that the relative weight of vocals may not have been disentangled from other factors.

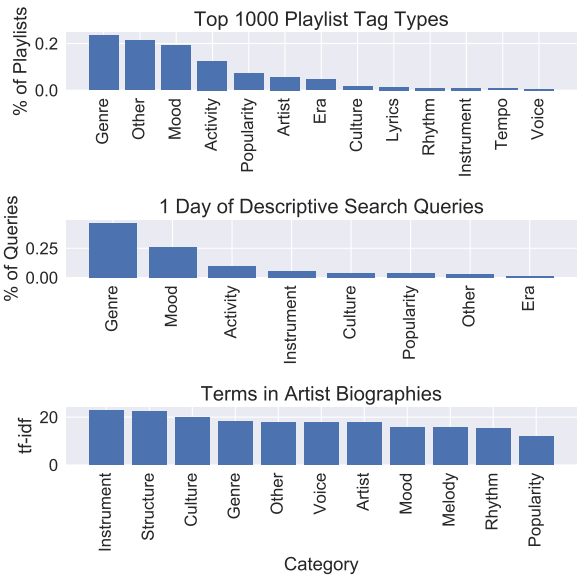

Figure 1: (Top) Percentage of Spotify playlists containing one of the top 1000 tags corresponding to each category. (Middle) Percentage of descriptive search queries corresponding to each tag category, sampled from one day of search data. (Bottom) tf-idf for each term category in artist biographies compared with Wikipedia term frequencies.

We perform a similar analysis on descriptive terms from one day's worth of Spotify search queries, and obtained similarly inconclusive results, shown in Figure 1, middle.

2.2 Artist Biographies

Finally, we analyze descriptive terms that occur in 100,000 professionally authored artist biographies on Spotify. We use TF-IDF [16] to retrieve terms that are distinctive to music writers, by comparing the frequency of terms in artist biographies to the frequency of the same terms in Wikipedia. The 100 most distinctive terms, grouped into semantic categories, are displayed in Figure 1, bottom. While many terms are much more frequent in music text (e.g. "bassist", "jazz", "songwriter"), vocals specifically were not more frequently mentioned than other musical aspects. One can hypothesize that the TF-IDF method is insufficient for this particular task, due to vocals being commonly discussed outside the context of music, and thus a relatively more common word in Wikipedia.

2.3 Conclusions

Our results thus far do not show support for our general hypothesis. It may be the case that the intuitive notion of the relevance of vocals to user preference is misleading. On the other hand, it may also be the case that the importance of vocals is implicit in this data, as certain vocal styles are indicative of genre or mood. As such, the overlap between the voice and a number of the tags and descriptors analyzed prevents us from disentangling the unique effect of the voice from other musical components.

3. VOCALS IN SURVEY DATA

In order to disentangle the unique effect of the voice among other components, we gathered explicit data from users. Specifically, we conducted two online survey studies in order to collect self-reported data on 1) the salient components of music, and 2) their relative ranking. Unlike prior surveys, such as [12] that presented users with short musical excerpts and groupings of adjectives to rate, we allowed the users to freely enter their responses to the question "When you listen to music, what things about the music do you notice?". This allowed us to assess whether vocals would emerge as a salient component of music. In addition, we explored what aspects of the voice users report as being important to their musical taste.

3.1 Survey 1: Semantic Components of Music

The aim of our first survey was to establish an unranked set of self-reported salient components of music. While our hypothesis was that the vocals would be prominent, it was crucial to avoid biasing respondents as the data collected were explicit. As such, our first survey asked participants what they notice when listening to music that might make them like or dislike a song. We deliberately did not specify anything further, such as the type of music, or that we were interested in components of music, nor were participants asked to listen to musical excerpts so as not to bias responses. As an exploratory measure, we then asked participants to describe what about vocals specifically might make them like or dislike a song *after* the previous open ended questions, so as not to bias responses. Responses to these two open-response questions were manually sorted into semantic categories by the researchers.

3.1.1 Recruitment

A random sample of 50,000 people was drawn from the database of Spotify's Monthly Active Users (MUAs), divided approximately equally between the United States and Canada. 860 individuals responded to the survey, however 224 did not respond to any questions beyond the consent form, and 9 were removed for giving nonsensical responses. 626 individuals — 338 women (average age 33.6 years with a standard deviation of 16.1); 288 men (average age 30.6 years with a standard deviation of 15.5) — completed the survey in its entirety.

3.1.2 Survey

An online consent form was first presented to respondents. We then asked:

> *Q1:* When you listen to music, what things about the music do you notice? Please list as many as you can think of here:

The respondents were shown a screen with open-response format fields to complete, in which they could complete up to seven fields. On the following screen, respondents were presented with a list of their responses in random order, and asked:

emotions	When I can either relate or empathize with them and when the song projects the emotions onto me.
emotions	If it doesn't feel like there's emotion behind it, or somehow lacking.

Figure 2: Survey 1 sample answers for *Q3*. (Top) Card for an answer to *Q3a*. (Bottom) Card for an answer to *Q3b*.

> *Q2:* Please rank how important the aspects you listed are to your musical preference, where 1 is the most important.

They were then asked the following two questions about the items they ranked from 1 to 3:

> *Q3:* (a) What about ____ would make you like a song? (b) What about ____ would make you dislike a song?

Lastly, to explore what aspects of vocals may be relevant, participants responded to the following:

> *Q4:* (Please ignore these questions if you've already mentioned the vocals, the voice, the singer/rapper etc.) (a) When would vocals make you like a song? (b) When would vocals make you dislike a song?

They were then given the opportunity to comment on the survey, and were shown a final debriefing screen.

3.1.3 Semantic Categorization

A number of partially completed surveys contained responses sufficiently complete for card sorting. 317 sufficient responses — 262 from the completed surveys as well as 55 sufficiently complete partial – were then card-sorted by a team of researchers. Card-sorting is a common technique used in social sciences and elsewhere to discover clusters of related concepts [14]. Traditionally, individuals are presented with physical paper "cards" that have terms and/or descriptions printed on them, printed pictures, or a group of objects. They are then asked to group items in a way that makes sense, given the research question. Here, we apply card-sorting to derive semantically meaningful groupings of musical components from the freely entered words and phrases that participants entered in each field.

Participant responses to *Q1* (i.e. "When you listen to music, what things do you notice?") were printed twice, once next to their response to *Q3a* ("What about _ would make you like a song?"), and again next to the response to *Q3b* ("What about _ would make you dislike a song?"). As such, researchers had respondents' top 3 terms printed out twice, once next to the positive descriptive aspects of the term, and once next to the negative descriptive aspects. A term (e.g. "the lyrics") and its descriptor (e.g. "when they have meaning") comprised a card. Figure 2 shows examples of positive and negative cards that were used in card sorting.

As some responses were unclear (e.g. "the melody" was mentioned, but the descriptor clearly focused on the quality of the singer's voice), the research team was instructed

to look at both the term and its descriptor when determining its semantic category. The researchers then reviewed the cards a second time, and defined sub-categories where necessary.

3.1.4 Results

The output of this study was two sets of semantic categories: broad semantic categories of music, and vocal-specific semantic categories. Statistical testing was not possible, given the intentionally imprecise nature of the responses. However, out of the 626 responses to the first question, 186 (29.7%) mentioned the vocals, the voice, or the singer, 348 (55.6%) mentioned the lyrics, or the words, and 101 (16.1%) mentioned both. While this is no indication of relative importance, it does demonstrate that the voice and the lyrics were salient musical components to our respondents.

The broad semantic categories determined by the researchers are presented in the left column of Table 1 (note that the other results in Table 1 are from Study 2). The category of *Emotion/mood* referred to the ability of a song to evoke emotion, whether the emotion was a match or a mismatch to the current or desired mood or current activity, whether the emotion was desirable or undesirable, and nostalgia. *Voice* included genre related terms (e.g. mumble rap, metal, auto-tune, speechiness/rapping), descriptions of how the voice is used (e.g. unique/novel, screaming, pitch/pitch range, presence or absence of effects, intensity/effort/power, emotionality, authenticity, whininess/nasality, melodic-ness), skill, the innate qualities of the voice, liking/disliking, and the mix/blend. The *Lyrics* category represented items that indicated whether or not lyrics were present, their intelligibility, the presence of profanity, how "well" crafted they were, the "message", the meaning behind them or general lyrical content and how relatable they are. *Beat/Rhythm* referred to whether it was liked/disliked, whether it "fit" the song, danceability, and uniqueness. The *Structure/complexity* of songs included liking or disliking the hook or chorus, and the song length. Instrumentation referred to drums, bass, and guitar. *Sound* referred to audio quality and related concerns. Self-explanatory categories included *Tempo/BPM*, the mention of a *Specific Artist*, *Genre*, *Harmony*, *Chords*, *Musicianship*, *Melody*, and *Popularity/Novelty*.

3.2 Survey 2: Component Ranking

While the first study aimed at determining what attributes of music were salient in the minds of listeners, the aim of the second survey was to determine the relative importance of each of the components. Specifically, we explored whether the voice would be ranked highest among a list of musical attributes. To accomplish this, participants were asked to rank a list of attributes derived from the results of our first survey, thus allowing an assessment of whether or not vocals rank above other components.

3.2.1 Recruitment

A randomized sampling method was employed among the database of Spotify's Monthly Active Users (MAUs) that had not opted-out of email correspondence. An email with a link to an online survey was sent to 50,000 potential respondents, approximately equally divided among the United States and Canada.

A total of 531 respondents — 263 of which were women (average age 31.8 years, with a standard deviation of 16.5); 268 were men (average age 34.2 years, with a standard deviation of 14.8) — completed the survey in its entirety. 429 participants completed the first half of the survey (broad semantic categories), whereas 360 participants completed the second half (vocal semantic categories).

3.2.2 Survey

An online consent form was first presented to respondents. The derived semantic categories were rephrased to be more easily understood (see Table 1, Description). Participants were presented with the new list of descriptions in random order, and asked to "Please click all the items below that would make you like or dislike a song." They were then presented with a list of all the items they had clicked, also in random order, and asked to rank them.

As a continuation of our exploratory study of vocal characteristics, a second list was then presented, comprised of terms derived from the vocal and lyrics semantic categories. For clarity, the terms were rephrased as they appear in Table 2.

3.2.3 Analytic Strategy

Responses were subjected to Borda counting [3] and Robust Rank Aggregation [9]. Borda counting is a simple procedure for aggregating votes by summing ranks. The Borda score B_i for an item i is computed as $B_i = \sum_{p=0}^{N} (|r_p| - r_{p,i})$ where N is the number of participants, $r_{p,i}$ is participant p's rank of item i, starting at zero, and $|r_p|$ is the number of items ranked by p. The Borda method does not naturally extend to partial lists [4] — we have chosen to award higher scores to preferred items in long lists.

To verify the statistical significance of our findings we supplement the Borda count with Robust Rank Aggregation (RRA), in which we compare our survey results to a null hypothesis. Each item receives a score based on its observed position, compared to an expected random ordering. Upper bounds to p-values are computed using Bonferroni correction, with values of 1.0 indicating null findings. In this work we used the implementation provided by the RobustRankAggreg package [1].

3.2.4 Results and Conclusion

Results can be found in Tables 1 and 2, with categories ordered by descending Borda count. We are able to show statistical significance of both the most salient broad and vocal semantic categories. Importantly, our results show that the Vocals and Lyrics ranked second and third among

[1] cran.r-project.org/web/packages/RobustRankAggreg

Broad Semantic Category	Description	Borda score	p-value
Emotion/mood	How it makes you feel - the emotions/mood	4641	**<0.001**
Voice	Voice/vocals	3688	**<0.001**
Lyrics	Lyrics	3656	**<0.001**
Beat/rhythm	Beat/rhythm	3460	**<0.001**
Structure/Complexity	How it's composed, the hook, the structure	2677	1.000
Musicianship	Skill of the musicians, musicianship	2583	1.000
Melody	The main melody	2577	1.000
Sound	The "sound", or the recording quality	2406	1.000
Specific Artist	The specific artist	2349	1.000
Genre	The specific genre	2293	1.000
Instrumentation	The musical instruments (e.g. drums, bass, guitar)	2084	1.000
Tempo/BPM	How fast or slow the song is	1828	1.000
Harmony	Harmony	1763	1.000
Chords	The chords	1086	1.000
Popularity/Novelty	How popular or unique it is	777	1.000

Table 1: Broad semantic categories and their clarifying descriptions created during Study 1, ordered by rankings from Study 2 (see Study 1 results for attribute descriptions). The Borda scores and p-values from Study 2 are reported in columns 3 and 4. Statistically significant p-values are shown in bold. p-values of 1.000 indicate that the ranking is no different from random.

the list of components (Borda scores and RRA agree on the order of the first four broad categories). This indicates that, relative to other musical components, respondents overall indicated the importance of the vocals and lyrics.

4. NEW AVENUES FOR RESEARCH

While the musical attributes related to the broad musical categories (Table 1) are well studied in MIR, the attributes related to vocals (Table 2) present a number of exciting and unexplored research directions. A limiting factor to studying some of these problems, as is often the case, is the availability of data, and we encourage researchers to focus data collection efforts in these areas as well. A further limiting factor is that users of online musical platforms may come from a specific demographic, e.g. regular internet users typically younger than 35, who engage in music related activities in about one third of the online time, have had at least some musical education, and have a preference for pop, rock and classical music [12]. In addition, our sample was derived from the U.S. and Canada. As such, a cross-cultural sample may differ in their relative preference for vocals.

Our exploratory data suggest that there is a vast space of research in tagging and measuring different qualities of the singing voice, such as whether a singing voice is authentic, powerful, natural, melodic, nasal, or emotional (Table 2, rows E, H, I, K, M and G). In addition to these categories, determined by untrained listeners, there are a number of other more specific categories such as modes of phonation that could be explored. Further, in addition to vocal qualities, there are genre-centric vocal styles, such as identifying rap or screaming (Table 2, rows S and O).

Another interesting and (as far as we are aware) unexplored research area is to measure whether a voice fits or blends well with the background music (Table 2, row B). This is somewhat related to the problem of determining "mashability" in automatic-mashup generation. This is a broad problem that is likely based on many factors, such as the style of the vocalist compared to the background, the way the song is mixed, and the overall expectations of the musical genre. We suspect this could be most easily studied when isolated vocals/backgrounds are available in order to automatically generate examples of vocals that do not match the background by blending random combinations.

The problem of identifying whether a voice is "unique" is likely challenging (Table 2, row F), as it is not necessarily a quality that can be determined in isolation, but rather relative to many other voices. One possible approach to this problem would be to treat the problem as one of outlier detection.

Production effects applied to the singing voice are increasingly common, especially different types of distortion or the infamous auto-tune (Table 2, row Q). Automatic identification of these production effects presents an interesting challenge, and one where data could be automatically generated with the help of plugins for generating effects and databases with isolated vocals with corresponding backgrounds.

Measuring the relatability (Table 2, row J) of a singer is a quality that is relative to the listener, rather than absolute. Factors that could affect a singer's relatability could include the age, gender, culture or language of the singer relative to the listener, which might require automatic identification of each of these attributes of the singer.

Lyric intelligibility (Table 2, row L) has not been well studied, and also presents a novel challenge [7]. This problem does not necessarily directly require lyric transcription, and may be able to be determined from qualities of

	Vocal Semantic Categories	Borda score	p-value
A	Singing skill	3423	**<0.001**
B	How well the voice fits or matches the rest of the music	3380	**<0.001**
C	Lyrical skill / cleverness / wit	3145	**<0.001**
D	The meaning, or the "message" of the words	3038	**0.048**
E	Authenticity / "realness"	2884	**<0.001**
F	Uniqueness	2780	**<0.001**
G	If the voice is emotional	2771	**0.006**
H	Voice strength / intensity / effort	2721	1.000
I	If the voice sounds natural	2480	1.000
J	Being able to relate	2256	1.000
K	If the voice is melodic	2202	1.000
L	Whether or not you can understand the lyrics	2056	1.000
M	If it's whiny or nasal	1801	1.000
N	Whether or not there's screaming	1771	1.000
O	The overall pitch, or the range of the pitch	1400	1.000
P	Whether or not there are lyrics	1250	1.000
Q	Whether it has production effects on it, like autotune	1230	1.000
R	Profanity, explicit lyrics	1086	1.000
S	Whether or not there is rapping	909	1.000

Table 2: Vocal-specific semantic categories from Study 1, ordered by rankings from Study 2. Columns 2 and 3 show the Borda scores and p-values. Statistically significant p-values are shown in bold. p-values of 1.000 indicate that the ranking is no different from random.

the audio. Similarly, determining whether a singing voice contains lyrics or is wordless has not been studied (Table 2, row P).

Automatic lyric transcription has been studied [11, 13] but is not yet solved, and would power the automatic estimation of many of these vocal attributes. For lyric-related terms, given textual lyrics, while some attributes would be relatively simple to estimate (e.g. whether or not there is profanity), others present interesting NLP challenges, such as estimating whether the lyrics are "clever" or are "meaningful" (Table 2, rows R, C, and D).

5. DISCUSSION AND CONCLUSIONS

While our analyses of playlist titles and search queries were inconclusive, we show evidence that English-speaking respondents from the U.S. and Canada clearly indicated that the voice is a salient component of music. Specifically, Spotify users were asked what they notice about music while listening. Despite the unassuming nature of the question, our results showed that the voice was indeed salient among the group of reported musical attributes. Furthermore, users ranked the voice as the second most important component to their musical preference, after emotions.

Our results have a number of implications. With regards to MIR research specifically, our results suggest that the voice and lyrics are indeed relevant attributes that warrant further study. While individuals may not necessarily want or know how to describe vocals themselves, i.e. in their playlists or search queries, surveying listeners directly does indicate that they find vocals to be important. As such, clarifying how the voice relates to music preference is an important topic for future research.

Secondly, users indicated that the ability of a song to evoke emotions was the most important factor. This confirms findings in prior research of the relevance of emotional content in music, and how it is linked to musical preference, e.g. [10]. Therefore, examining how music affects the emotions of listeners remains an important theme. Interestingly, while genre was the most frequent term used to label playlists or search for music, respondents did not rank the specific genre as important relative to the other attributes. Understanding why this is the case warrants further study.

More relevant to our hypothesis, is that the vocals and the lyrics of a song were ranked second and third by respondents who were directly asked what components of music are important to their preferences. Therefore the link between emotions perceived in the voice and lyrics, and the emotions felt in listeners, is very relevant to questions of music preference. Clarification of these links was out of scope in these studies, and could be addressed in future research.

Lastly, we show the relevance of explicitly collected data that might guide future research. While we showed inconclusive findings regarding the prevalence of vocals in implicit data, we did show that the unique effect of vocals on music preference may be observed using survey data. As such, explicit data-gathering techniques often found in the social sciences, as well as collaborations with social scientists, may be of great use to MIR researchers.

6. REFERENCES

[1] Jorge L Armony, William Aubé, Arafat Angulo-Perkins, Isabelle Peretz, and Luis Concha. The specificity of neural responses to music and their relation to voice processing: An fmri-adaptation study. *Neuroscience letters*, 593:35–39, 2015.

[2] Pascal Belin, Robert J Zatorre, Philippe Lafaille, Pierre Ahad, and Bruce Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309, 2000.

[3] Jean C de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Academie Royale des Sciences*, 1781.

[4] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[5] Sean Hutchins and Sylvain Moreno. The linked dual representation model of vocal perception and production. *Frontiers in psychology*, 4:825, 2013.

[6] Sean Michael Hutchins and Isabelle Peretz. A frog in your throat or in your ear? searching for the causes of poor singing. *Journal of Experimental Psychology: General*, 141(1):76, 2012.

[7] Karim M Ibrahim, David Grunberg, Kat Agres, Chitralekha Gupta, and Ye Wang. Intelligibility of sung lyrics: A pilot study. International Society for Music Information Retrieval Conference, 2017.

[8] Patrik N. Juslin and Petri Laukka. Communication of emotions in vocal expression and musical performance: Different channels, same code? *Psychological Bulletin*, 129:770–814, 2003.

[9] Raivo Kolde, Sven Laur, Priit Adler, and Jaak Vilo. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, 28(4):573–580, 2012.

[10] Carol Lynne Krumhansl. Listening niches across a century of popular music. *Frontiers in psychology*, 8:431, 2017.

[11] Anna M Kruspe and IDMT Fraunhofer. Retrieval of textual song lyrics from sung inputs. In *INTERSPEECH*, pages 2140–2144, 2016.

[12] Micheline Lesaffre, Liesbeth De Voogdt, Marc Leman, Bernard De Baets, Hans De Meyer, and Jean-Pierre Martens. How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the Association for Information Science and Technology*, 59(5):695–707, 2008.

[13] Matt McVicar, Daniel PW Ellis, and Masataka Goto. Leveraging repetition for improved automatic lyric transcription in popular music. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3117–3121. IEEE, 2014.

[14] George A Miller. A psychological method to investigate verbal concepts. *Journal of mathematical psychology*, 6(2):169–191, 1969.

[15] Stephen W Porges. The polyvagal theory: phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, 42(2):123–146, 2001.

[16] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

[17] Michael W Weiss, E Glenn Schellenberg, Sandra E Trehub, and Emily J Dawber. Enhanced processing of vocal melodies in childhood. *Developmental Psychology*, 51(3):370, 2015.

[18] Michael W Weiss, Sandra E Trehub, and E Glenn Schellenberg. Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, 23(10):1074–1078, 2012.