Matthew Weisberg
Professor Devaney
Computational Musicology
Fall 2018

Timbral Salience and Attention: Spotlight on Donald Trump

I. Introduction

A twig snaps in the brush, and everyone's head turns to see what caused the sound. Despite the calls of birds, the rush of a nearby river, the chatter amongst the group, everyone's attention is grabbed by that sound. What makes that sound rise above the rest? Why does it capture our attention? These questions are the subject of the study of timbral salience, or the attributes of a sound that allow our ears to pick it out amidst a crowded auditory field. This question is interesting in both musical and cultural contexts. For composers, it is well known that orchestration and instrumentation greatly affect the experience of a piece. We are asked - why does a trumpet allow us to follow its melody cleanly above the orchestra, whereas a french horn would not be able to carry the same melody as clearly? As sound artists and experimental composers using increasingly abstract digital instruments, it is particularly useful to be able to consider a sound's timbral characteristics and its salience-potential. From a cultural perspective, timbral salience may be an unrecognized force in what is called an 'attention-based media ecology'. Considering the world of information propagation as an ecosystem, where 'success' is defined by propagation of information, the ability to be noticed among a saturated media field could offer a great advantage.

It has been shown that timbral characteristics can act on our preconscious perception of sound. Goydke et al. (2004) showed that slight timbral changes to an otherwise uniform series of single violin notes could induce a preconscious neurological phenomenon known as the Mismatch Negativity (MMN) effect. Jones et al. (1999) showed that changing the timbre of a repeated note sequence disrupts our ability to store memories of the pattern presented, more so than keeping the timbre constant and changing the note or changing both parameters simultaneously. Chuen et al. (2016) showed that changing the timbre of a sequence of notes caused changes in listeners heart rates and skin conductance, both measures of preconscious stress response. Bradley (2009) helps us understand these results. In both cases, our ears are responding to minute cues that deviate from our expectations, resulting first in the presence of MMN coupled with subsequent physiological changes in heart rate, breathing rate and skin conductance which are signs that our bodies feel threatened. Our bodies respond to expectation deviation with suspicion, and our attention turns towards the object - in this case small timbral cues - that our bodies are perceiving as a novel and possibly threatening.

A question then arises, is it possible to capture our attention by co-opting our bodies' threat-detection mechanisms through the use of abnormal timbral information? To explore this question, this project looks at the voice of Donald Trump, who in 2016 rose to the most powerful position in the country despite no history in governance. How did he achieve such success? Many cite his savvy navigation of our current media ecology to help spread his message and elicit strong emotional responses from supporters and opponents alike. Among the many explanations for his trajectory, this project seeks to explore the possibility that small timbral features of Donald Trump's voice are particularly effective at capturing our attention, which in a world of constant information exposure is a large advantage.

This hypothesis was based on an experiment conducted in December 2017 in partnership with Gabriela Baez. The experiment presented listeners with ten different clips of celebrities speaking, each 10 seconds long, and were asked to record the amount of time needed to identify the speakers. In 60 trials, Donald Trump was both the most frequently recognized voice and the voice identified most quickly, despite speaking on a topic unrelated to his rise to political notoriety. Interestingly, 'times identified' and 'speed of identification' were not linearly correlated, meaning his voice was not just widely recognized but also particularly easy to identify.

This experiment led us to wonder: is there something unique in the sonic content of Donald Trump's voice that gives him this recognizability advantage? Is there something in his timbral cues that is evoking some version of the MMN or other attention-grabbing stress responses? To explore this idea, this project focused on timbral cues taken from the field of timbral salience, that ability to stand out and capture our attention, and particularly on acoustic parameters identified from the literature as being especially salient. Odd-to-Even harmonic ratio and Inharmonicity were chosen based on the work of Chon (2013). In one experiment, they showed that Odd-to-Even harmonic ratio was the most significant factor in which voice a listener would follow when presented with two-part harmonies. In another experiment, they showed that inharmonicity significantly affected listeners' perceptions of the blendedness of unison dyads. Spectral Flux was chosen based on the work of Chuen et al. (2016), who suggest that varying Spectral Flux caused the greatest amount physiological response in their experiments. In addition, two parameters used in speech recognition - Spectral Flatness and Crest, are included (Peeters 2011). The hypothesis of this project is that Donald Trump's voice would have abnormal values for one or more of those spectral descriptors when compared to other people delivering the same speech material.

II. Methods

A. Procuring Recorded Material

The dataset of Donald Trump speaking consists of 12 clips taken from Youtube (Appendix B). Clips were chosen to represent a diversity of emotional contexts, with the aim of presenting a full picture of DJT's speaking voice. Each clip was transcribed and 8 additional speakers were then recorded reading the transcripts. Some speakers recorded themselves at home and submitted the results to the author, while others were recorded by the author directly. The clips recorded by the author were edited and exported using Ableton Live 9 as 44.1 kHz sample rate, 16-bit, .wav files. Speakers were compensated with pizza or extorted by their emotional commitment to the author to contribute their recordings to the experiment. The resulting dataset had 108 clips total, 12 clips from 9 different speakers. In order to simulate the conditions of the initial experiment, each clip was trimmed to 3-seconds long which was the average time needed to identify a generic speaker.

B. Definitions

Below are the definitions for the parameters used to analyze the audio files, all of which are taken from Peeters (2011) except Average Onset Frequency, which was defined in the documentation of the BBC Vamp Plugin Package. Equations for these parameters can be found in Appendix A.

Odd-to-Even Harmonic Ratio is defined as the relative strength of the odd harmonics of the estimated fundamental frequency to the even harmonics of the estimated fundamental frequency.

Inharmonicity is defined as the relative strength of the all the non-harmonic partials compared to the harmonic partials in a given signal.

Spectral Flux is defined as amount of variation of the spectrum over time, calculated by the total distance of spectral powers from one window of the file to the next.

Spectral Flatness is defined by the comparison of the geometrical mean and the arithmetical mean of the spectrum.

Spectral Crest is defined by comparing the maximum and the arithmetic mean of the spectrum.

Average Onset Frequency is defined by the average number of onsets per minute.

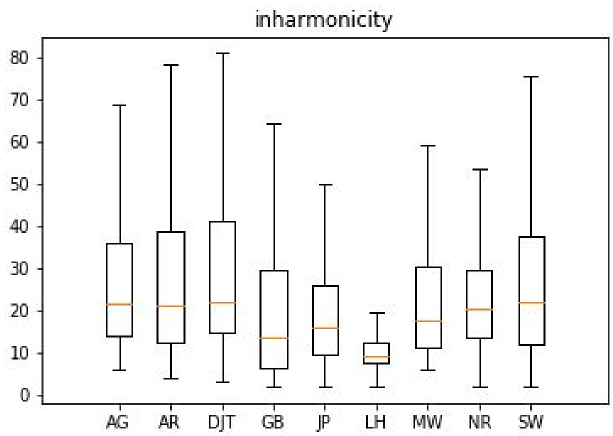
C. Batch Analysis using Vamp Plugins

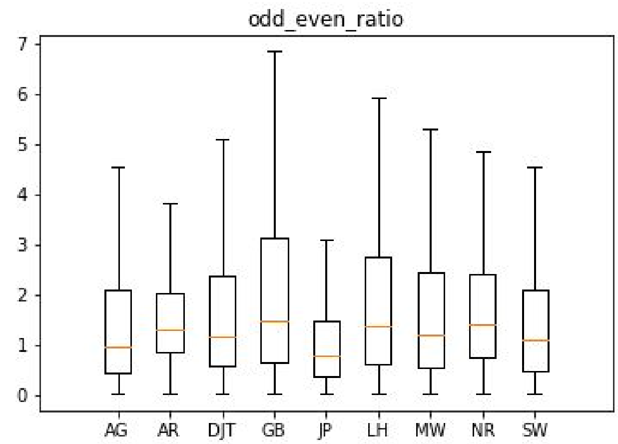
All speaking clips were using Vamp Plugins and batch analyzed using Sonic Annotator with the following settings:

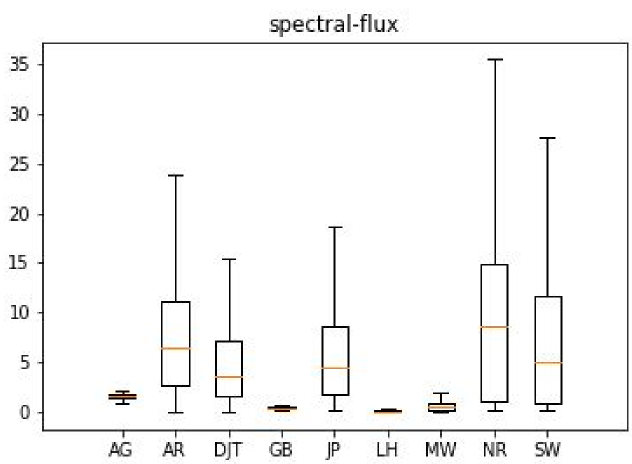
Attribute	Creator	Window Size (frames)	Hop Size (frames)	Window Type
Odd-to-Even Harmonics Ratio	Libxtract	1024	256	Hanning
Inharmonicity	Libxtract	1024	256	Hanning
Spectral Flux	BBC	1024	32	Hanning
Spectral Flatness	Libxtract	1024	256	Hanning
Spectral Crest	Libxtract	1024	256	Hanning
Average Onset Frequency	BBC	1024	256	Hanning

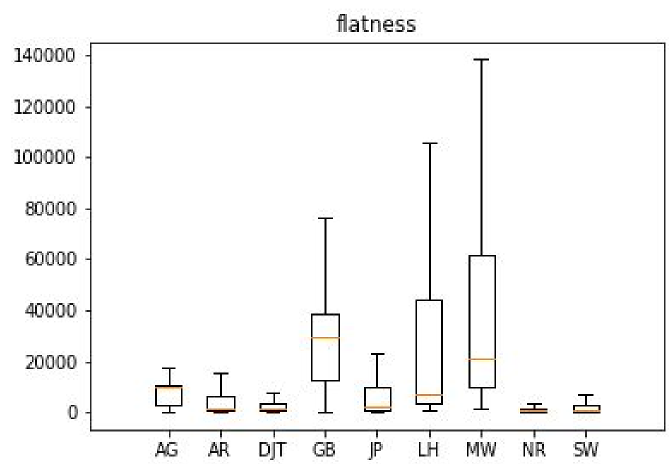
IV. Results

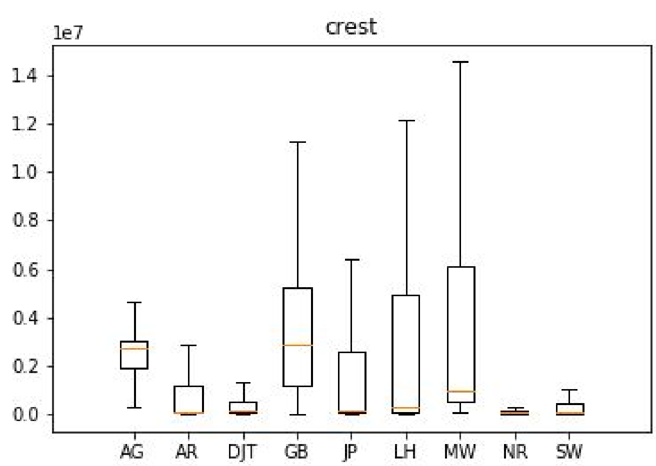
CSV files containing the audio analyses were brought into python using the Pandas library and plotted using MatPlotLib. Comparisons of spectral characteristics were made using boxplots of Interquartile Ranges (IQR). In the below diagrams, **'DJT'** represents Donald J Trump:





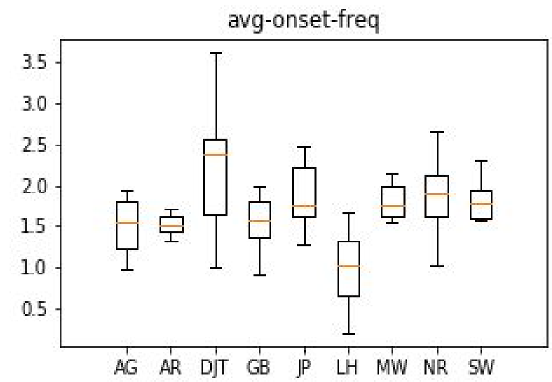






V. Discussion

In order to deem a distribution significant, the IQRs of that distribution would need to be distinct from other distributions to which it is being compared. We do not observe any such isolation from our analyses, and therefore conclude Donald Trump's voice having exceptional qualities for any of the parameters we searched for. Further experiments interested in DJT's speech patterns of DJT might try analyzing different auditory parameters, such an intonation patterns, rhythmic timing, and phoneme onset frequency. One analysis showed that there is something noticeable about DJT's onset rate, so that may be a fruitful avenue for further research. In general the parameters highlighted in this study are related to the behavior of onsets, so focusing on smaller bits of audio, say a word or syllable, may be more informative than examining whole phrases. Of note, this experiment did not control for the effects that various recording, compression, and export settings might have on the audio files. Further experimental designs should consider this issue more carefully.



Further, this project was motivated by a desire to engage with an observation driven by past experimental work and personal observation: that Donald Trump's voice has a powerful impact on our psyches and bodies. Throughout both experiments conducted by author, the author observed that listeners to DJT's speeches tend to have noticeable reactions with their facial expressions and body language. This observation has implications for further research. As noted before, Chuen et al. 2016 showed that small timbral changes can induce physiological changes such as heart rate, breathing rate and skin conductivity. It would be interesting to assess the relative physiological changes induced in listeners when listening to DJT speak.

Lastly, this project was not solely a scientific exploration of timbre and its effect on attention, but also an artistic project intended to critically engage with Donald Trump as a character, voice, and representative of our current political and media climates. It was the artistic intent of the author to confront their hatred and disgust for Donald Trump head-on by engaging with hours of speech material and in order to curate for broader exposure sound bytes that demonstrated both the dangerousness, stupidity and sliminess of the 45th president of the United States of America.

Works Cited

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, *46*(1), 1-11.

Chon, S. H. (2013). *Timbre saliency, the attention-capturing quality of timbre* (Doctoral dissertation, McGill University Libraries).

Chuen, L., Sears, D., & McAdams, S. (2016). Psychophysiological responses to auditory change. *Psychophysiology*, *53*(6), 891-904.

Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, *21*(3), 351-359.0.

Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 464.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902-2916.

Appendix A

$$\mathrm{OER}(t_m) = \frac{\sum_{h=1}^{H/2} a_{2h-1}^2(t_m)}{\sum_{h=1}^{H/2} a_{2h}^2(t_m)}.$$

$$\mathrm{inharmo}(t_m) = \frac{2}{f_0(t_m)} \frac{\sum_{h=1}^{H} (f_h(t_m) - hf_0(t_m)) a_h^2(t_m)}{\sum_{h=1}^{H} a_h^2(t_m)}.$$

$$\mathrm{SFM}(t_m) = \frac{\left(\prod_{k=1}^{K} a_k(t_m)\right)^{1/K}}{\frac{1}{K}\sum_{k=1}^{K} a_k(t_m)}.$$

$$\mathrm{SCM}(t_m) = \frac{\overset{a}{\underset{k}{\max}}\, k(t_m)}{\frac{1}{K}\sum_{k=1}^{K} a_k(t_m)}.$$

$$= 1 - \frac{\sum_{k=1}^{K} a_k(t_{m-1}) a_k(t_m)}{\sqrt{\sum_{k=1}^{K} a_k(t_{m-1})^2}\sqrt{\sum_{k=1}^{K} a_k(t_m)^2}}.$$

Spectral Flux

Appendix B

Clip #1: It's Called Weather (1:09-1:31)
I think that there'll be little change here it'll go up it''ll get a little cooler it'll get a little warmer like it alway has for millions of years it'll get cooler it'll get warmer it's called weather

Clip #2: Bad Ones **(1:36)**
You think they're gonna put they're good ones they don't put they're good ones they put they're bad ones

Clip #3: Sudden Screeching and Beautiful Halt (2:10-217)
The never ending growth of red tape in America has come to a sudden screeching and beautiful halt

Clip #4: Brexit is Brexit (1:10)
You know Brexit is Brexit

Clip #5: I know a lot About Wind (2:40-2:48)
I know a lot about wind wind is destroying a lot of problems number one it's killing the birds the birds and the eagles and you see what's happening in california

Clip #6: They Were Laughing With Me (1:40-143)
They weren't laughing at me they were laughing with me

Clip #7: A Man's Life is Shattered (1:15-1:25)
A man's life is shattered his wife is shattered his daughters who are beautiful incredible young kids they destroy people they want to destroy people these are really evil people

Clip #8: Maybe it's True **(**2:00:220)
Number one in his class at yale perfect human being great father great husband this is a great person and people are saying well maybe it's true and because of the fact that maybe it's true he should not become a united states supreme court justice

Clip #9: Victims of Immigrant Crime Engagement (**00**:24-00:40)
And it's because of families like yours that my administration created the new office of DHS the victims of immigrant crime engagement which has been doing i hear a fantastic job we call it VOICE so that your voices can be heard

Clip #10: What do I do Mom? (2:27-3:11)
Mom i'm so pleased to tell you i just got a fantastic job with IBM I just got a fantastic job with general motors i just got i'm so proud mom a terrible thing just happened a person who i never met said that i did things that are horrible and theyre firing me from my job mom i dont know what to do mom what do i do what do i do mom

Clip #11: Your Organization is Terrible (1:36-1:56)
Not you Not you your organization is terrible

Clip #12: So Gross (:44-:52)
I don't think Ivanka would do that inside the magazine although she does have a very nice figure i've said that if ivanka weren't my daughter perhaps I'd be dating her