Audio Data in Populist Political Speakers' Use of Emotion

Catherine Moez

August 18, 2023

Introduction

While the wave of populist victories and near-victories of the mid to late 2010s may have now subsided, interest in what drives support for unconventional political outsiders continues. Text, audio, and video image analysis tools help us to quantitatively and systematically (consistently; transparently) assess some of the theoretical claims that have been made about populist politicians in qualitative works. In a set of systematically sampled[1] clips taken from Canada's House of Commons between early 2015 and late 2017 (n = 1,557, Rheault and Borwein 2019; Cochrane et al. 2022), I find that populist left or right speakers are in fact *not* more likely to engage in emotionally activated speech in the Canadian parliament. Emotional activation, or emotional arousal, refers to a person being in a state of heightened emotion, with either a positive or negative valence. However, the left-populist New Democratic Party (NDP) is much more negative in sentiment valence than other parties or ideological blocs of parties, and the centre-left Liberal Party much more positive, even before coming to power in October 2015.

Some practical considerations and potential problems when measuring emotional activation from audio are discussed, as are potential solutions. As the main variable currently used to measure emotional arousal, a deviation in pitch above a speaker's typical pitch level ('pitch deviation' in this paper; similar to measures used in Dietrich et al. 2019a, 239; 2019b, 945), these include difficulties in capturing a speaker's 'true' pitch in an oppositional setting like House of Commons debates. Implications of speakers switching between the French and English languages, and other demographic variables (age; province or region; gender) are also considered. Finally, inductive analysis using supervised machine learning models (decision trees; random forests) are used to identify (a) potential other audio-based predictors of emotional activation, and (b) potential audio predictors of negative sentiment.

To briefly state findings, a high *standard deviation in pitch* in a given short clip is (also, in addition to deviation in pitch above mean) a strong predictor of emotional arousal, as are louder volume and higher variation in volume (amplitude) in an utterance. In the parliamentary setting at least, statements with negative sentiment tend to be more emotionally activated than positive ones, meaning many of the same audio variables are relevant. High deviation in pitch above a speaker's mean *with* high variation in pitch tends to signal anger and other more assertive or confrontational speech, whereas high variation in pitch in a short clip *without* pitch rising above a speaker's mean seems to signal nervousness or sadness (or, in some cases, simply background noise in the clip).

Working definition of populist speakers

I define populist left or populist right speakers ideologically, as those that reject at least some policy elements of either the conventional post-war centre-left 'bundle' (Gidron 2016) – moderate state intervention in the economy, gradual liberalization in social issues, relative openness to free trade immigration – or the centre-right bundle of moderately more free-market and socially conservative positions. This means that on the left, those who go further in opposing economic globalization or regional integration (NAFTA; the European Union) and advocate for a stronger role for the state in the economy (the 'old' or 'materialist' left) are counted as left-populist. On the right, due to more variation in principles than egalitarianism alone (Cochrane 2015),

[1] 'Ten time points from every third Question Period', Rheault and Borwein 2019, 7.

there may be either highly anti-immigrant but economically statist 'welfare chauvinists'; socially liberal, pro-market libertarians, and other unconventional combinations (Gidron 2016).[2]

In practice, given the relatively high number of political parties in contention in Canada, as well as extremely high party discipline in legislative voting, many of the models shown simply use party as an indicator of ideology: the NDP is left-populist; the Liberal Party, Bloc Quebecois, and Green Party centre-left (the latter two could be described as progressive-leaning 'new left' or postmaterialist parties; the Bloc Quebecois is also a regional party); and the Conservative Party of Canada (CPC), as a whole, is a mainstream right or centre-right party. In other models, 'right-populist' members of the CPC are split off from the mainstream of the party as a separate group, based on legislative vote clustering, former membership in the strongly right-wing Canadian Alliance or Reform Party of Canada, and/or expressing support for the anti-government and anti-COVID-measures 'Freedom Convoy' protests of early 2022).

Theoretical expectations

There has been no shortage of academic literature denouncing populists, from contemporaneous critics of the first self-described 'Populist' movement, the U.S. People's Party (Frank 2020), through to academic histories of early agrarian-populist parties in North America (Hofstadter's *The Age of Reform*, Jaeger 2023), and recent works written after a wave of new far-right parties emerged in the 1990s and after shocking electoral upsets for outsider figures such as Donald Trump in the 2010s (Mudde 2004; Inglehart and Norris 2017). For many of these authors, populist figures are especially xenophobic and discriminatory (at least on the right, Mols and Jetten 2016; Mudde 2004, 546, 550; Cinar et al. 2020, 242; Inglehart and Norris 2017, 446), anti-democratic (or 'anti-pluralist', Norris 2020, 699), negative, angry, and simplistically 'Manichean' (Cinar et al. 2020; Norris 2020, 698) in their rhetoric, moral beliefs, and policy proposals.

Many of these general claims remain curiously impressionistic, however, despite the many high-quality works studying specific populist movements. It seems amiss to attribute speaking for the mainstream or everyday 'people', and claiming to transcend traditional left or right ideologies, when many prototypical Third Way politicians did the same (Mudde 2004; Bickerton and Accetti 2021; Mair 2013). Bickerton and Accetti share the anecdote that both Emmanuel Macron and populist-right outsider Marine Le Pen began their 2017 presidential campaigns with the same slogan: 'Neither left nor right', before Macron changed his to 'both left and right' (2021, 60). A number of centre-left and centre-right parties also attempted to recover votes from the fringe right by taking stronger stances against the EU (David Cameron's Brexit referendum was itself an election promise from 2015) or refugees and immigrants (Also in the UK, New Labour's first successful election campaign featured a number of 'tough on crime' and 'controls on immigration' promises).[3] As such, focusing on populists in isolation fails to reveal when and how the mainstream left or right adopt similar messaging, imagery, and other tactics, such as presenting a casual or unconventional personal image, or communicating directly with followers and the public.

Furthermore, the political mainstream has been charged with having its own distinct negative emotional themes, at least in the United Kingdom. In both the Scottish independence referendum of 2014 and the 'Brexit' referendum of 2016, the half-joking phrase 'Project Fear' was coined to describe the dominant tenor of messaging from all major parties (excluding the pro-independence Scottish National Party in the first case; Dyson 2019). Party leaders across the spectrum warned of economic uncertainty and downturns should Scotland separate, or if the UK left the EU. In the latter case at least, perhaps the message that the UK was already experiencing heightened growth and prosperity, and was at risk of losing these, did not connect with many voters in 'left-behind' areas. Personally holding bleak views about the economic trajectory of the 'last 30 years' (Hall and Gidron 2017, S59), or about the future (Ward et al. 2021), strongly predicts populist voting.

Despite these selected examples, though, the conventional wisdom is that mainstream (establishment; broadly centrist, from centre-left to mainstream right) politicians are typically more subdued and emotionally neutral than politicians at the fringes (Major and Tomasevic 2023, 2). The mainstream of politics may also be

[2]A neither-left-nor-right 'syncretic populism', more of a random grab-bag of issue positions, is also seen in some countries and parties, such as Italy's Five-Star Movement, but is not a political force in Canada.

[3]'Accommodation' in this sense (Meguid 2005) may in fact backfire for the political mainstream by legitimating hateful views; Ranciere 2021; Krause et al. 2023.

generally more positive in emotional valence than the populist fringes (Consider 'America is already great' as Hillary Clinton's 2016 tagline, compared to the nostalgic and slightly more negative 'Make America Great Again' slogan used by Donald Trump). Moderate or mainstream politicians may seek to avoid expressing anger, an emotion associated with attributing blame (Bonansinga 2020, 95), perhaps as part of their carefully crafted outreach across 'old' left or right divisions (at least by more deliberately 'Third Way' centre-left politicians, Dyson 2019). Stylistically, populist politicians are generally agreed to be more emotionally intense (Moffitt 2016, 46, 167), displaying stronger highs and lows in the emotional themes that they engage in (anger and nostalgia, Bonansinga 2020, 95-96; expressions of empathy and 'caring', pride, and hope, in addition to negative themes, Albertazzi and Bonansinga 2023, 4), compared to the more restrained conventions of the broad political mainstream.

As such,

H1: Populist politicians will display emotional activation (emotional arousal) more frequently than their conventional politician counterparts. They may also exhibit more extreme signs of emotional activation when they do.

The analysis will also show inductive models indicating: potential alternatives to pitch increases above a speaker's mean ('pitch deviation') that could be used as an alternative signal of emotional arousal, as well as potential audio predictors of (negative) sentiment.

The dataset

The audiovisual data used in this analysis consists of 1,557 quasi-sentence-long clips (statements; utterances) delivered by members of parliament (MPs) in the Canadian House of Commons, between January 2015 and December 2017. A systematic sampling of 'ten time points from every third Question Period' was used for collection (Rheault and Borwein 2019, 7). English-language transcripts, the speaker's gender (0 = male; 1 = female), name, and binarized scores for activation (0 = unactivated; 1 = activated) and sentiment (0 = negative; 1 = positive) were attached. The activation and sentiment scores are first averaged and then binarized out of scores assigned on a 0 to 10 scale by (up to) three manual coders who watched a video clip of the statement (Rheault and Borwein 2019; Cochrane et al. 2022). 929 of the observations were coded by a set of graduate student coders (3 watching video; 3 reading the text transcript only; as disclosure, I was one of the text transcript coders). An additional 628 videos were coded in the same way, as well as for the emotion of anxiety, by workers on Amazon's MTurk service (Rheault and Borwein 2019). Please note that to use the main audio-based measurement of emotional activation, a speaker's mean pitch in a clip exceeding their mean pitch across all recorded time points for that speaker (pitch deviation; 'pitch_dev'), for many models the dataset must be subset to only speakers who have 2 or more observations in the dataset (n = 1,472). This avoids falsely attributing a pitch deviation score of 0, unactivated (the video's mean pitch is identical to the speaker's mean pitch), even if the speaker is visibly emotionally agitated in that one clip.

I have added variables capturing the language of the video (82.7% English; 17.2% French); a binary variable for whether the speaker appears to be heavily uncomfortable or unfamiliar with the language they are speaking in the clip ('langmismatch', binary variable, after preliminary data checks showed the most extreme observation with the highest pitch deviation above mean was a short clip of Andrew Scheer, a primarily English speaker, asking a question in French).[4] I have also obtained former party affiliation (since 1980), MP's current party, and some birth year information from OpenParliament.ca; I web-scraped additional missing birth dates from Wikipedia or Library of Parliament.

Variables

Emotional Activation

Emotional activation, also referred to as emotional arousal, or, in this paper, 'agitation', describes a state where a speaker feels some type of intense emotion, with either a positive or negative valence. Paul Ekman's theory of six basic emotions (sadness, happiness, surprise, disgust, anger, and fear, Rheault and Borwein 2019) considers anger, fear (anxiety), and surprise to be 'activated' emotions, while happiness, sadness, and disgust are toward a calmer pole (Gu et al. 2019; Cochrane et al. 2022, 99).

[4] It is important to note that simply having an accent was not coded as being 'mismatched' with the language, only those who appeared extremely unfamiliar with the words and had difficulty speaking the whole sentence somewhat fluently.

As described by Dietrich (2019a; 2019b), physiological changes experienced when a person is more emotionally aroused include the tightening of the vocal cords, which raises a speaker's pitch. These changes may be subtle and imperceptible to a listener. Commonly used phonetics software such as Praat (Boersma and Van Heuven 2001) and Parselmouth, a Python-based interface for working in Praat (Jadoul et al. 2018) can, however, record audio measurements for pitch, amplitude (volume or loudness, in everyday language), and other aspects of a speaker's voice, at many time points per second throughout an audio clip. Additional variables used in previous works on capturing distinctive aspects of a speaker's voice include formants (capturing where the tongue is placed when pronouncing vowels, which Neumann 2019; 2020 uses to study formality of speech); mel-frequency cepstrum coefficients (MFCCs, Teferra et al. 2022; the audio 'cepstrum' is a transformed version of a frequency spectrum that mimics more closely how the human ear hears sound), harmonics-to-noise ratio (higher when a voice is clearer and more distinct in tone; associated with ageing and some medical conditions), and more. Rheault and Borwein (2019) take an alternative approach of numerically encoding the (non-emotional) aspects of a speaker's voice as a voice embedding; this work instead reports the audio variables directly. Future extensions could also consider the use of pauses and the speed of speaking, also known as the articulation rate.

The main variable used to capture audio-based emotional activation is 'pitch_dev', a speaker's deviation, in Hertz (Hz), from their mean pitch across all time points in all of their speech clips. It is the audio clip's mean pitch, minus the speaker's mean. Alternative versions where this deviation in Hertz was divided by the speaker's standard deviation in pitch, either in a specific clip, or across all time points in all of their clips, are used by Dietrich et al. (2019a, 239; 2019b, 945) were also tested, and performed similarly to the main variable measured in Hz. The reason that 'pitch deviation' in Hz was not then transformed by dividing this number by the (video or speaker's) standard deviation in pitch, was twofold: first, as analysis showed, the baseline levels of emotional activation in the context studied (Question Period debates) were extremely high (44.4% of the original 1,557 clips per the manual labels; 40.8% of the 1,472 clips retained for audio analysis, per a categorical audio-based measurement).[5] This setting meant that most speaker's baseline mean pitch likely included many activated speech samples, and was therefore higher than a 'true' mean pitch measurement would reflect. Second, as inductive models showed, higher *standard deviation of pitch* within a given clip was positively related to emotional activation, measured as deviation above a speaker's mean pitch. This means that dividing the 'pitch deviation' itself by the standard deviation (of pitch) measurement may in fact *understate* the level of emotional arousal found in the clip.

To briefly give some descriptive figures, the typical MP in the dataset, after subsetting to those with 2 or more clips, has a median of 10 audio samples, the mean being 17. (25th percentile is 5; 75th is 20; the maximum is speaker Justin Trudeau, with 78). After subsetting, the Liberal Party has the most observations, at 677; the Conservatives have 496; the NDP 279; the Bloc Quebecois 13 (with 4 different MPs), and the Green Party 7 (all from 1 MP). The average 'pitch deviation' above or below a speaker's mean, after subsetting, is +0.087 Hz.

Populism

As mentioned in the introduction, political party labels are used in some models to capture ideology. The NDP is populist left, although, like many left parties internationally since the 1980s and 1990s, it has taken steps to distance itself from its more radical left past; the party removed the word 'socialism' from its party constitution in 2013 (Payton 2013). The Bloc Quebecois and Green Party, as well as the Liberals, can be categorized as centre-left. The Conservative Party, as a whole, can be considered mainstream right, or, equivalently in this work, centre-right.

Ideology: An Alternative Variable

As the CPC contains a mix of more conventional, moderate Conservatives with those seen as more unconventional or extreme, an additional variable, 'Ideology', is created to separate the two. Former party of affiliation, general public perceptions of party leaders and high-profile MPs, as seen in media coverage, and reported positions on the Freedom Convoy protest of 2022 (Hanes and Gray 2022; Levesque 2023; Nash 2022)

[5] The cutoff for where a pitch deviation in Hz score translated to a categorical 1, for activated, or 0, for not, was set slightly above 0 Hz, at 3.45 (Hz, above speaker's mean). The exact 3.45 Hz figure was chosen as it maximized accuracy between the manual binary labels for activation and the audio-based categorical labels, at approximately 62%.

are considered in attaching ideologically either mainstream right or 'populist right' labels to Conservative Party MPs. (See detailed note in Appendix). 64 Conservatives are left labelled as 'Conservative', meaning the mainstream of the party, and 17 as 'right-populist': Maxime Bernier (a CPC member at the time), Pierre Poilievre, James Bezan, Jason Kenney, Cheryl Gallant, Chris Warkentin, Gerry Ritz, Jeff Watson, John Barlow, Kerry-Lynne Findlay, Kevin Sorenson, Marilyn Gladu, Mark Strahl, Martin Shields, Michael Cooper, Ted Falk, Scott Reid, and Pierre Paul-Hus.

While some of the right-populist speakers have many emotionally activated speeches, according to the categorical audio measure (48% of James Bezan's 21 clips; 58% of Mark Strahl's 12; 60% of Kevin Sorenson's 10), some do not. Maxime Bernier, a Conservative until his narrow loss of the party leadership race in 2017, and subsequent founder of the right-populist People's Party of Canada (PPC), registers only one of his four speeches (25%) as activated. In contrast, current Conservative leader Pierre Poilievre, thought to be more associated with fringe movements than most predecessors, registers as emotionally activated in 40% of his 47 speech clips – a higher figure, but one comparable or lower to what many mainstream Conservatives score.

Demographic variables; Speaker's language

As noted above, gender is included in the original Rheault and Borwein dataset (2019); party and province of representation are added from OpenParliament.ca data by the author. Smaller provinces and territories are combined as 'Region': Nunavut, the Northwest Territories, and the Yukon as 'North'; New Brunswick, P.E.I, Nova Scotia, and Newfoundland as 'Atlantic'. The Western provinces, as well as Ontario and Quebec, are left as their own 'regions', due both to size and distinct political dynamics within each (the Prairies tend to favour either further left or right parties, but not the Liberal Party, for example).

Language of the video clip ('vidlang') is tagged manually by the author, as is a binary variable ('langmismatch') for speakers who seemed highly unfamiliar with the language of the clip that they were speaking in. (This variable was added after preliminary tests and data checks indicated that some English speakers making statements in French seemed to register much higher pitch in these utterances, compared to others). This variable captures when an MP is speaking very choppily and seems to be struggling in an unfamiliar language, not minor idiosyncracies such as having an accent in either official language. Given the heightened pitch associated with some of these clips, including the dataset's highest pitch deviation of all, it seems reasonable that attempting to speak in an unfamiliar language might trigger physiological activation – attempting to focus and not make a gaffe may induce some level of anxiety.

Why use audio?

The dataset at hand contains up to 6 manual coders' assessments of each audiovisual clip's emotional activation, and of sentiment. Some tests using the manual labels are shown below, as is the correspondence between manual tags and audio-signal-based scores. Cues for positive or negative emotional valence (sentiment) can also be readily found from text transcripts (either with automated text analysis, Rheault and Borwein 2019; or by transcript readers, Cochrane et al. 2022). On the other hand, emotional activation has been proven as more difficult to read from text alone (Cochrane et al. 2022; disclosure: I was one of the text coders).[6] One reason to use the audio-based variables is therefore to test their similarity to what was observed by multiple human coders, for the same dataset.

There are nonetheless other several reasons for audio-based measurements to be assessed. First, they provide a consistent and minimally resource-intensive way to tag short multimedia clips with measurements that could be used to detect emotional cues. Researchers may not have the resources to employ human coders; or may find this approach undesirable for other reasons (inconsistencies or various biases in reading emotions, for example). Second, training human coders to interpret emotions such as 'anxiety' in the same way, or to use the 0-10 point scale in the same way, may be less practicable than taking audio measurements, in some cases. For example, the video labellers also consistently label videos that are somber or sad in emotional tone as negative and activated (see the correspondence images, below), whereas these emotions are considered to be low-activation in Ekman's schema (Gu et al. 2019).

[6] As such, my scores are not part of the video coders' scores that are averaged, binarized, and used as activation and sentiment labels in the final dataset.

Third, Cochrane and colleagues (2022) show that manual coders who only score a textual transcript are closely aligned with video-viewing coders in judging sentiment, but are less accurate in assessing emotional activation (2022; the dataset is the same as the one used in this analysis, other than the subsequently added MTurk-coded videos). Using audiovisual inputs – videos consisting of only text (linguistic), audio, and image modalities themselves (Rheault and Borwein 2019, 5) – can help to avoid the pitfalls of relying solely on text. In the political domain, at least, speakers reading out nominally positive text in a sarcastic or angry tone of voice is far from rare. Audio cues may help to register the actual emotion that the speaker is projecting more accurately.[7] In addition to sarcasm or anger, a speaker coming across to an audience as nervous is unlikely to be captured in a text script.

For either budget, consistency (between coders, or within coders), accuracy (sensitivity to subtle emotional cues), or other reasons, such as studying specific emotions and phenomena that are harder to find in literal text, therefore, the quantitative analysis of audio signals could be a preferred choice for some research projects compared to either manual coding of a speaker's emotions, or automated text-based approaches.

Methodology

As such, the remaining analysis uses audio signals of emotional activation to explore (possible) ideological differences in the expression of emotional agitation, either in frequency or intensity. As such, a number of linear regressions, as well as logistic regressions, are shown. A subsequent section uses automatic feature selection in supervised machine learning models (decision trees and random forests, Breiman 2001) to inductively find potential indicators of sentiment, as well as potential alternative indicators of emotional activation, other than the pitch deviation measurement itself. (Finding alternatives could avoid the need for having multiple separate clips from each speaker in the dataset).

I first review a few findings from the manual emotional labels, and the correspondence between manual labels and audio-based measurements.

Results based on the manual labelling of speech clips

Proportion of speech clips labelled as 'activated', manual tags:

Ideology	ProportionActivated
Bloc	0.38
Conservative	0.37
Green	0.57
Liberal	0.44
NDP	0.57
RightPopulist	0.42

Based on manual emotion labels alone (binarized out of video coders' manual scores assigned on a 0 to 10 scale), the left-populist NDP and the Greens (with very few speeches) are the most likely to make emotionally activated statements. This finding holds, at the 99.9% significance level, in a chi-squared test of difference in proportions. In a logistic regression model with the Liberal Party as the reference category, the NDP are more frequently emotionally activated, at the 99.9% significance level, and the *mainstream* Conservatives less so, at the 95% level.[8]

Proportion of speech clips labelled as 'positive', manual tags:

[7] As a great deal of political psychology work emphasizes that people often form judgements of politicians based on emotional and intuitive cues, politicians' actual presence when speaking in public seems highly relevant to understanding these more character-based assessments of politicians.

[8] For comparability, even though all 1,557 original clips could be included as all have manual labels, only the same 1,472 observations that are used in audio data analysis are used here.

Ideology	ProportionPositive
Bloc	0.31
Conservative	0.36
Green	0.29
Liberal	0.68
NDP	0.16
RightPopulist	0.36

Based on manually labelled sentiment tags (1 = positive emotional valence; 0 = negative), the Liberal Party are much more positive in sentiment than all other parties. The difference is statistically significant, in a logistic regression model, with all other parties at at least 95%. The NDP are the most frequently negative in emotional tone, whereas other opposition parties (the more New Left than economically hard left Bloc Quebecois and Greens) are less so.

Notably, as a party in government is known to have more positive speech (Hirst et al. 2014; Rheault et al. 2016), the Liberals are more positive in sentiment than other parties *even before* they are elected to office on October 19, 2015. 48 percent of Liberal speeches prior to October 19th, 2015 are positive (5.5 or higher on the 0-10 scale average), compared to only 17 percent of the Conservatives' (and 44 percent for the Bloc; 25 percent for right-populist Conservatives, and only 7 percent for the NDP). The differences between the Liberals and all other parties or factions, other than the Bloc, are significant at the 99% level or higher. (The models are shown in the Appendix).

As such, based on manual tags alone, the NDP are distinctly more negative than other parties, and the Liberals much more positive in tone. The mainstream Conservatives appear to be somewhat less emotionally activated (in terms of frequency of activated statements) than either their party's right wing, or the left parties. The NDP also are more emotionally activated than the Liberals, based on these tags. This provides tentative support for **H1**.

Correspondence between Manual Tags and Audio-Based Measures

With a cutoff point for 'activated' videos set to a speaker's mean pitch in a video exceeding their mean pitch (across all time points in all clips) by 3.45 Hz or more, 40.8 percent of audio clips are found to be activated, while 59.2 percent are not. The 3.45 Hz cutoff point was chosen to maximize the correspondence between video scorers' assessments of activation and the audio-based categorical labels. Deviations from a speaker's mean pitch, in Hertz, ranged from -78.4 at the minimum to +137.6 Hz at the maximum, with a median of -0.95, mean or 0.09, and standard deviation of 22.40. As mentioned above, the confrontational and oppositional nature of parliamentary debates (Question Period perhaps especially, compared to bill debates or other proceedings) means that emotional activation is highly common, and as such, speakers' pitch means observed in the dataset may be somewhat higher than what a 'true' measure based on clips of the speaker in a less emotionally conflictual setting might be.

The standard deviation of pitch measurements in each video ('stdevF0Hz', or 'vid_pitchsd' in models and graphs below) is depicted in images below as it was found inductively (through machine learning models, shown later) to relate most strongly to the speaker's 'pitch deviation' (their mean pitch in the video, in Hz above their pitch mean across all videos). The potential use of this variable as an alternative audio variable capturing emotional activation is considered.[9]

The relatively strong relationship between standard deviation in pitch and heightened pitch above a speaker's mean (indicating emotional arousal) makes the standard deviation measurement potentially promising in

[9] The video clip's standard deviation in pitch is correlated with pitch deviation at Pearson's r = 0.257 for 'vid_pitchsd', measured from one audio-processing package in Python; Pearson's r is 0.322 for 'stdevF0Hz', measured from another. The two measures come from different Python packages and functions, and are not completely identical. Caution may be needed regarding how different audio processing software may differ.

measuring emotional activation in speakers who may only have one audio clip in a dataset. A word of caution is that women appear to have higher variation in pitch, in general (3.24 Hz higher when using ‘vid_pitchsd’; 12.01 Hz higher when using ‘stdevF0Hz’; some audio measurements differ when different packages or functions are used), and so an appropriate control may be needed.[10] Variation in pitch did not appear to correlate with speaker’s age, at least after controlling for gender.

Visualizing the Audio-Based Activation Variable; Comparing Manual Emotion Tags to Audio-Based Measures

The relationship between the primary audio-based measurement of emotional arousal, ‘pitch deviation’ above a speaker’s mean, is shown plotted against standard deviation of pitch in the audio clip, below.

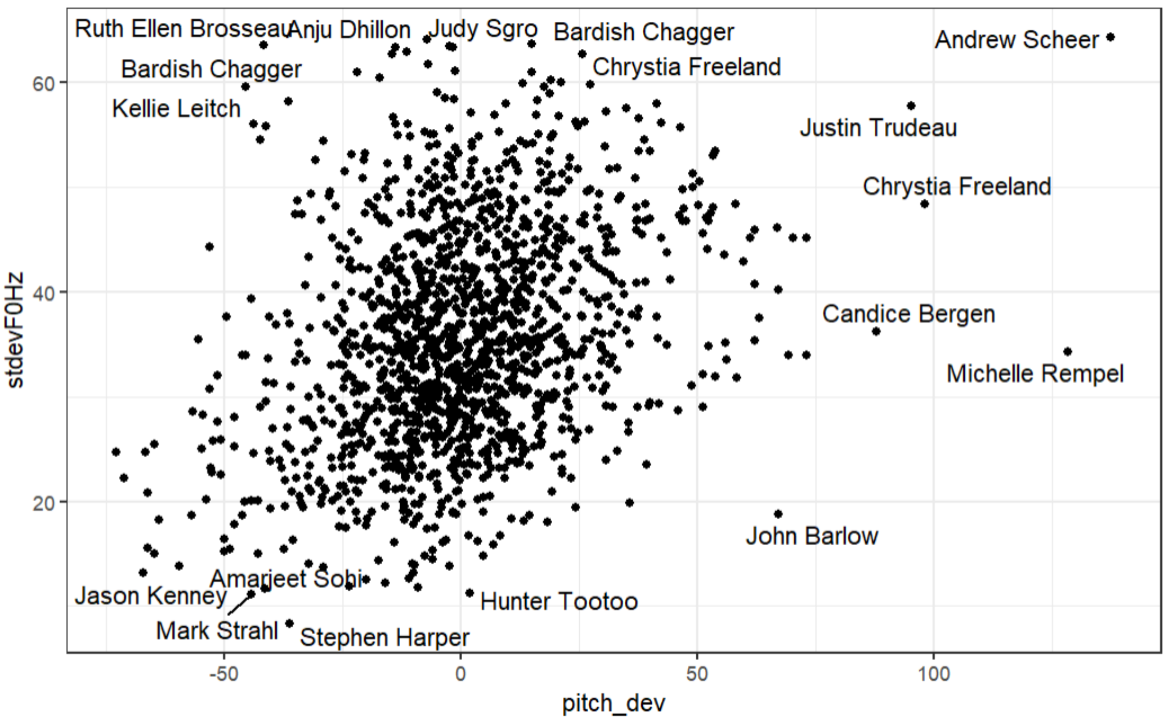


Figure 1: The relationship between the primary audio-based measurement of emotional activation, speaker’s deviation in pitch above the typical mean, and standard deviation of pitch in the video clip.

As seen in the image above, female speakers tend to have higher pitch variation within each audio clip then men, in general. Notwithstanding this gap, a weak but positive relationship between a speaker’s deviation from their usual pitch mean (‘pitch_dev’) and the standard deviation of pitch values in a given audio clip (‘stdevF0Hz’) can be seen.

Next, compared to the manual activation tags (scores on a 0-10 scale given by up to 3 video coders, binarized to (5.5+) 1, activated, or (<5.5) 0, unactivated), we see that higher pitch in a given audio clip does tend to correlate with manual recognition of activation, with some notable exceptions, including some of the speeches with the highest pitch deviations – the furthest to the right on the x-axis.

Replacing the binarized scores with the averaged video coder scores out of 10, the relationship between manual and audio-based assessments seems weak at best. Several of the most activated clips (statements from Andrew Scheer, rated 4; Chrystia Freeland, rated 3; and Candice Bergen, rated 2) based on an audio

[10] The difference was not due to women having more statements that were activated in general; manual tags found 47% of women’s speeches to be activated and 43% of men’s; the audio-based categorical variable found 41% of women’s speeches and 41% of men’s speeches to be activated.

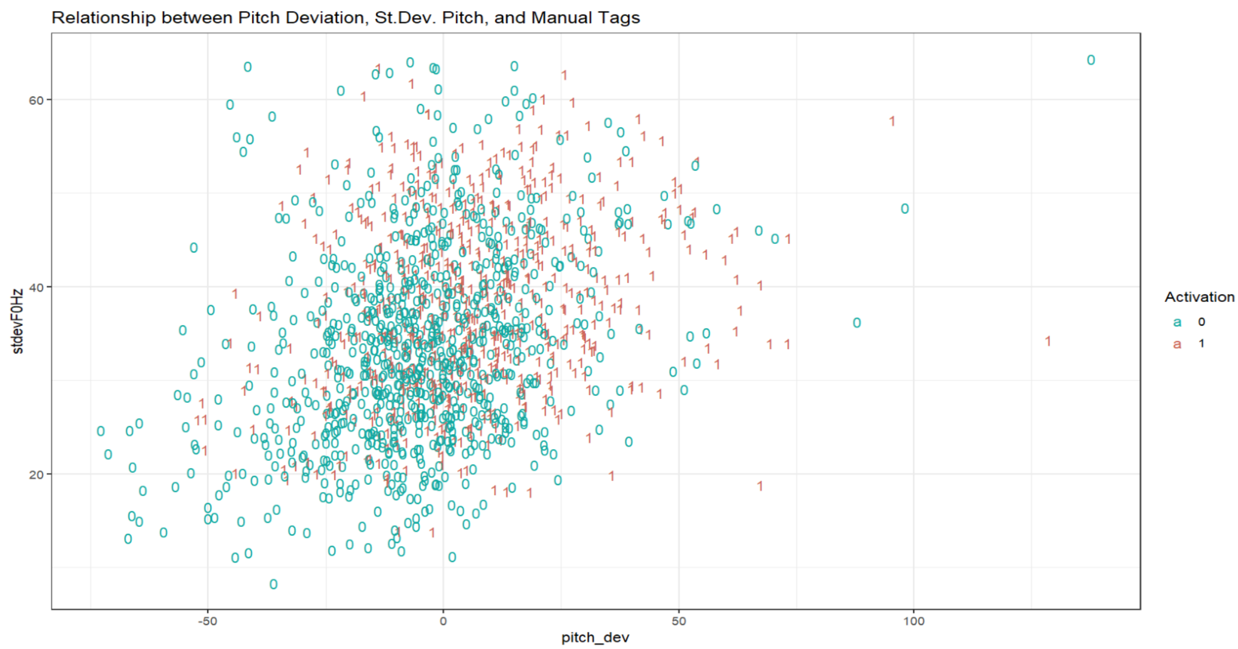


Figure 2: The relationship between speaker's deviation in pitch above the typical mean, standard deviation of pitch in the video clip, and manual tags for activation.

measurement are rated lower than some of those in the lower left quadrant, which should be the least emotionally activated.

Closer inspection of the audio and text for these clips at the extremes revealed that many in the lower left quadrant consist of speakers expressing the emotion of sadness – commenting on a public tragedy or death, for example. This makes sense in some way – while sadness may not have physical activation in Ekman's system, it seems reasonably that a viewer would see the strong emotion expressed and rate this as more 'activated' than a more dry statement about procedural issues, for example. At the top end (on the x-axis), there may simply be a problem with the quality of manual labels given.[11]

Inspections of the 'off-line' cases that were high in standard deviation of pitch, but not in 'pitch deviation', also revealed instances of background noise or the brief appearance of multiple speakers in the clip (which, additionally, may be more common for female MP speakers), or, in other cases, speakers who seemed emotionally activated in either a nervous or sad register, as opposed to the observations found toward the right of the plot, where the most common emotionally activated emotion seen was anger.

The texts of the most highly 'activated' speeches, with pitch deviations of 55 Hz or more above the speaker's usual pitch, in descending order, are as follows:

audiofile	name	text
sample_HOC_0827.wav	Andrew Scheer	Once again, who are they?
sample_HOC_0489.wav	Michelle Rempel	Why can it not do the same thing for genocide victims?

[11] Sentiment tags are similarly patchy in quality; statements rated as 'positive' include: 'The tiny room I had in Lebanon was not safe.' (John McCallum), as well as Freeland and Bergen's statements shown below in text transcripts. Clear instruction given to manual coders may help to reduce this problem. While not all video clips were scored by all three video coders, number of coders does not appear to be responsible for some unexpected scores; see Appendix for number of coders per clip.

audiofile	name	text
sample_HOC_0314.wav	Chrystia Freeland	Here is a news story: The Conservative government has hired two former White House communications strategists as part of a 'sustained' effort to raise Canada's profile in
sample_HOC_0919.wav	Justin Trudeau	We did not create the Phoenix problem, but we are going to fix it.
sample_HOC_0622.wav	Candice Bergen	I know there are some hard-working people with integrity on that side. Will any of them stand up, show some independence, and say no to the Prime Minister and no to the House leader, who has completely botched this for all of you?
sample_HOC_1209.wav	Scott Brison	meanwhile they've slashed the Building Canada Fund for the next two years
sample_HOC_0091.wav	Stephen Harper	We are still trying to find 40 million of those dollars.
sample_HOC_1006.wav	Marc Garneau	the Aerospace industry, everybody should be very happy about it, thank you.
sample_HOC_0094.wav	Ralph Goodale	Mr. Speaker, here is the government's record: the economy is shrinking, unemployment is up, TD Bank says it is getting worse, CIBC says job quality is at the lowest ebb in 25 years, BNN is reporting the most pessimistic business outlook since the last recession, and the Bank of Canada says only substantial monetary stimulus is keeping Canada from falling back into recession.
sample_HOC_0781.wav	John Barlow	Will the finance minister make those calls?
sample_HOC_0052.wav	Charlie Angus	Let us stick on the issue here, because not only are Conservative insiders being subpoenaed, but last week we learned that a number of key Conservative MPs have been called to testify.
sample_HOC_0569.wav	Rona Ambrose	My question is simple.
sample_HOC_0286.wav	Chris Warkentin	Will the minister extend these provisions that the farm families desperately need?
sample_HOC_0917.wav	Justin Trudeau	We told Canadians that we would run deficits, so we could grow the economy and put more money in the pockets of Canadians who needed it.
sample_HOC_0875.wav	Justin Trudeau	The fact is we always focus on the security of Canadians, and we always will.
sample_HOC_0411.wav	Sean Casey	It is a process that includes candidates from Atlantic Canada and respects regional representation.
sample_HOC_0068.wav	Scott Brison	Yet, Alberta's finance minister is set to deliver a budget on March 26.
sample_HOC_0906.wav	Pierre Poilievre	Now small businesses are expected to have faith in the government's idea of reasonable.
sample_HOC_0002.wav	Thomas Mulcair	It is a simple question.

audiofile	name	text
sample_HOC_1464.wav	Jason Kenney	by Twitter we want the decision to be made by the people mr. or politician
sample_HOC_0901.wav	Marilyn Gladu	How can the minister continue to mislead Canadians?
sample_HOC_0381.wav	Mark Strahl	Why are the Liberals so intent on threatening British Columbia families with their made-in-Ottawa, job-killing carbon tax?
sample_HOC_1016.wav	Chrystia Freeland	will the Conservatives quick parroting their talking points, face this grim reality, and admit that they have no effective plan for jobs and

We can observe that while the text of some speeches suggests strong emotion, others are more neutral in content, but are delivered in an agitated tone of voice. Scheer's statement is delivered in French, perhaps explaining a subconscious level of activation that manual coders did not notice.

In contrast, several of the speeches with the lowest pitch deviations (65 Hz below speaker's mean, or lower), are sad and somber in tone – or, alternatively, seem neutral to positive in content but, despite the use of the word 'urgent' in one, more subdued emotionally. The text segments are longer, and perhaps seem more carefully prepared, than in the most activated clips. Speakers in many of the most activated audio clips are asking short, confrontational questions, whereas some wordings in the least activated appear more rote ("again, I appreciate him bringing it to our attention").

audiofile	name	text
sample_HOC_0691.wav	Chrystia Freeland	I would like to recognize the work of the member for Selkirk-Interlake-Eastman on this bill, as well as my colleague the Member for Etobicoke Centre and the great Irwin Cotler.
sample_HOC_0911.wav	Todd Doherty	Mr. Speaker, since day one, we have been challenging the government to make securing a new softwood agreement its number one trade priority.
sample_HOC_1007.wav	Jason Kenney	and has offered a transition plan to this individual as well. Mr. Speaker, we will continue to support leading seaman Young
sample_HOC_0924.wav	Michelle Rempel	Mr. Speaker, women who have undergone female genital mutilation suffer infections, difficult urination and childbirth, pain during intercourse, and more.
sample_HOC_0212.wav	Justin Trudeau	Mr. Speaker, I would like to thank the member opposite for her heartfelt words and indeed add to them personally that the entire government and indeed the country stands with the community of La Loche.
sample_HOC_1313.wav	Mark Holland	urgent we will take a look at this situation, I will be happy to get back to the member and again I appreciate him bringing it to our attention

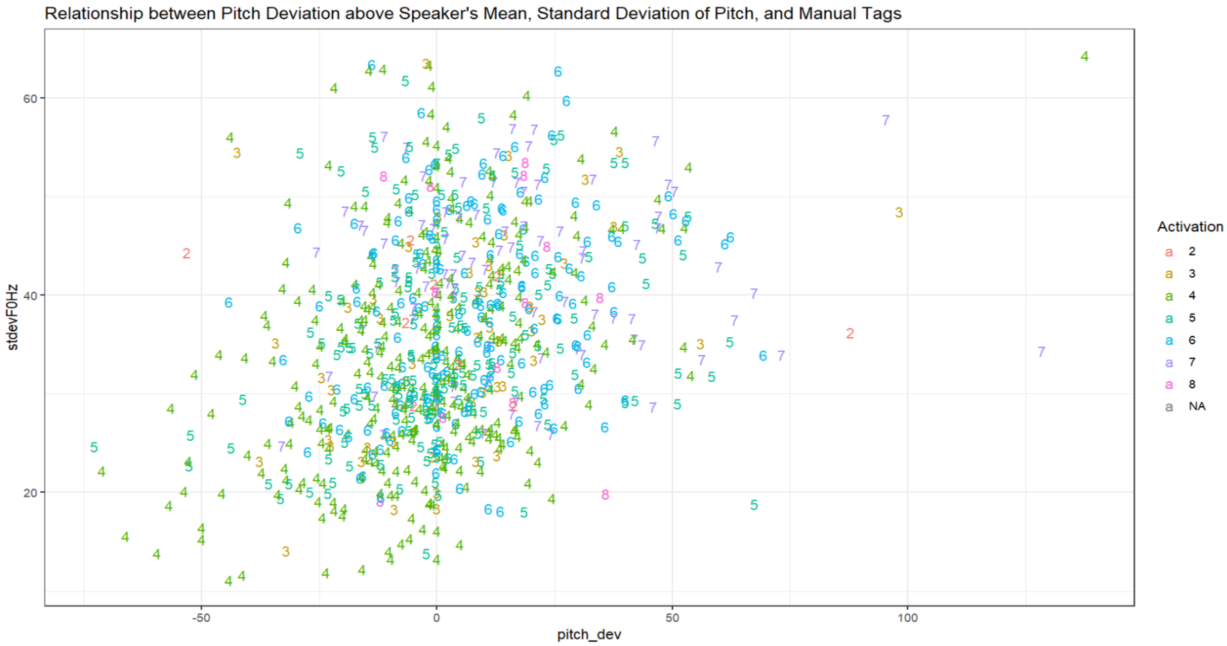


Figure 3: The relationship between speaker's deviation in pitch above the typical mean, standard deviation of pitch in the video clip, and manual tags for activation.

Main Results

The first two models shown use the basic 'pitch deviation' variable, in Hz, as the dependent variable. Model 1 compares all parties' (or ideological groups') averages on this metric. The next two models show average pitch deviation scores *among activated speeches only*, tagged by either the manual tag for activation (Model 2) or the audio cutoff variable (i.e., only for speeches with at least a 3.45 Hz deviation above mean, Model 3). All are linear regressions. The final model uses the categorical audio activation variable as the dependent variable, in a logistic regression. It shows how frequently the party's speeches fell into the activated category, as opposed to the first three, which show extremity of pitch deviation.

Additional models with demographic variables for age, province or region, and gender are shown in the Appendix. (A model with only negative *and* activated speeches included returned essentially the same effects as the activated-only speeches; not shown). The reference category for Ideology is Liberal.

In a simple comparison of what proportion of video clips were scored as activated (categorical; cutoff at 3.45 Hz above speaker's mean) by party, no substantial differences are seen between the three main parties (Conservatives, Liberals, NDP). Once the Conservatives are split into the centre-right through mainstream bloc of the party ("Conservative") and its more unconventional rightward fringe ("RightPopulist"), a slightly higher proportion of speeches made by the right-populist group are found to be activated. This difference in proportions, however, is not found to be statistically significant in a Chi-squared test ($p = 0.746$), or in the Model 4 regression shown below.

In short, no clear support was found for **H1** based on audio indicators. If anything, the NDP appears to be *less* emotionally aroused, at least among the subset of speeches deemed activated (Models 2 and 3). Additional models containing age, province, and gender variables similarly did not show clear patterns, other than speakers from Alberta being more emotionally activated than others. (In further tests, the NDP and right-populist Conservatives appear to drive this trend, but differences by ideology or party within Alberta were not statistically significant).

The null results are perhaps due to the generally emotional atmosphere of Question Period, which is designed to be confrontational and oppositional. Including all speeches, rather than subsetting to particular topics a party or ideological group may have stronger positions on, may also explain the relative lack of systematic differences in emotional arousal.

Table 5:

	Dependent variable:			
	pitch_dev			pitchdev_cat
	OLS			*logistic*
	(1)	(2)	(3)	(4)
ideologyBloc	4.766	0.575	−0.749	−0.084
	(6.277)	(9.681)	(7.374)	(0.575)
ideologyConservative	0.606	1.344	0.563	−0.066
	(1.487)	(2.275)	(1.727)	(0.136)
ideologyGreen	3.304	0.243	−6.372	0.673
	(8.517)	(10.805)	(8.229)	(0.768)
ideologyNDP	−0.363	-3.612^{*}	-4.022^{**}	0.001
	(1.595)	(2.104)	(1.827)	(0.145)
ideologyRightPopulist	2.143	3.035	0.441	0.256
	(2.001)	(2.958)	(2.164)	(0.179)
Constant	−0.267	6.512^{***}	20.735^{***}	-0.386^{***}
	(0.862)	(1.244)	(0.987)	(0.078)
Observations	1,472	658	601	1,472
R^2	0.001	0.009	0.011	
Adjusted R^2	−0.002	0.002	0.003	
Log Likelihood				−993.621
Akaike Inf. Crit.				1,999.242
Residual Std. Error	22.418 (df = 1466)	21.467 (df = 652)	16.340 (df = 595)	
Note:				*p<0.1; **p<0.05; ***p<0.01

(Gender reference = 0, male; Date of Birth (DOB) decade reference = 1940s; Ideological group reference = Liberal; Region or province reference = Ontario).

Language of the speech clip itself, as well as a speaker appearing to struggle with the language of the clip, were not statistically significant in additional regressions (shown in Appendix).

Determining Activation Level from Text?

For those who may not have manual labels available, but want an alternative check on activation levels measured from audio, text-based analysis may help to sort observations into relevant categories. It appears that activation level is less clearly read from text alone than sentiment is, based on Cochrane et al. (2022)'s findings. Nonetheless, there may be systematic patterns in words used when speakers are emotionally agitated, rather than subdued.

The following images shows distinctive words belonging to either the 'unactivated' or 'activated' speech statements, using the categorical pitch-based measure of activation. The relatively simple and transparent term frequency-inverse document frequency (TF-IDF) classifier is used.

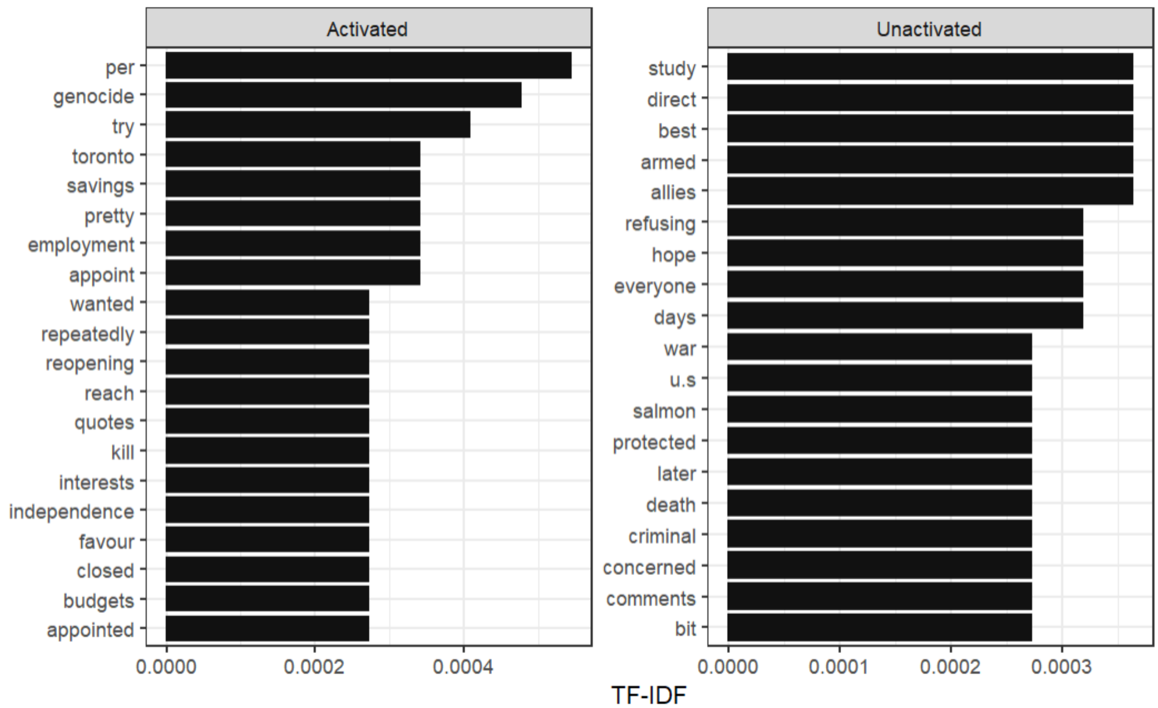


Figure 4: Distinctive words and bigrams of emotionally unactivated or activated statements.

The results are somewhat suggestive of outrage or strong feeling being present in the activated speeches (‘kill’, ‘genocide’, economic words such as ‘employment’), although many similarly war- or death-related words appear in the unactivated speeches as well (‘armed’, ‘allies’, ‘criminal’, ‘death’).[12] Alternatively, a stemmed version of words could be input into a classifier, or numeric representations of word meaning. The original word forms have been used here, to minimize changes made to the text and to possibly derive additional meaning from information such as part-of-speech or tense.

Alternatively, with speeches first split by the categorical pitch-based measure of activation, and then by video coders’ sentiment labels, distinctive words for all four categories are shown in Image 5.

These results are suggestive and illustrative at best, but pursuing text-based measurements of activation cues may be worthwhile. Notably, the emotion of sadness appears to predominate in the negative and unactivated category, whereas outrage or anger seem more present in the negative-activated combination. While Ekman considers anxiety to be emotionally activated as an emotion, the anxiety-suggesting word ‘concerned’ appears to be delivered primarily when the speaker is calm.

In short, custom dictionary development for detecting distinctive words used when speakers are emotionally activated may help in scoring texts. As with sentiment dictionaries, however, the words that signal particular moods may differ by the context or domain of the speech (Rheault et al. 2016).

[12] ‘Pretty’ is always used as an adverb (‘pretty difficult’; ‘pretty sure’; etc.); the word ‘quotes’ only around allegations of genocide; ‘kill’ in the activated column refers only to ‘jobs’.

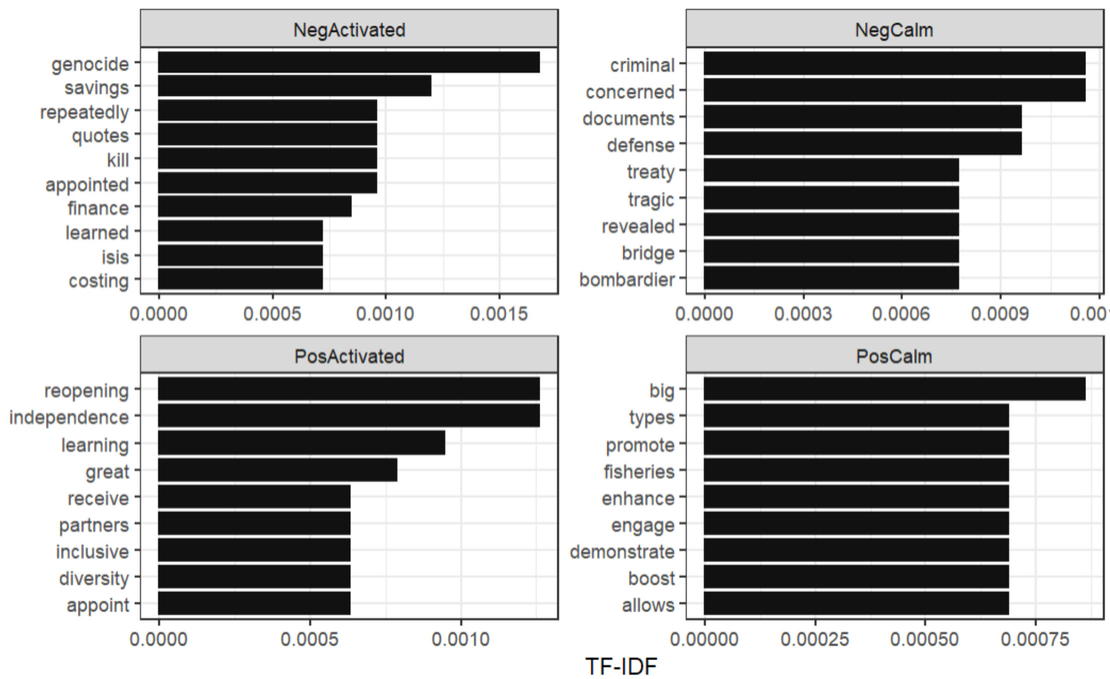


Figure 5: Distinctive words and bigrams of all four sentiment and activation categories.

Additional Audio Cues of Emotional Activation: Automatic Feature Selection

Using the speaker's increase in pitch over speaker's mean pitch variable ('pitch deviation', 'pitch_dev') as the dependent variable (DV) below, a random forest model is run to identify features (variables) in the dataset that best predict (high) scores on emotional activation.

The random forest model consists of 1000 decision trees (Breiman 2001), each split in each tree being limited to considering a randomly selected subset of 30 of the total number of variables. This limitation forces the models to use a wider range of features in attempting to make predictions, as opposed to relying heavily on (what potentially might be) a very small number of variables that strongly predict outcome scores.

The trees are defined as regression and not as classification trees (categorical DV). (One tree, to illustrate, is shown below.) 80 percent of the dataset is used as training data for the model, with 20 percent withheld for subsequent goodness-of-fit testing. (These are standard figures; any of these parameters could be automatically finetuned as well. As the purpose is largely to identify audio variables of interest, rather than to maximize prediction accuracy, manually chosen settings in line with commonly chosen values were retained).

A feature importance plot demonstrates which variables, over the 1000 trees, consistently improve the quality of predictions best. The plot does not show the direction of the relationship each has with the dependent variable, but simple tests such as exploring individual trees or comparing group means can be used to identify the direction of the relationship.

An illustrative tree demonstrates that low variation in pitch throughout the clip ('stdevF0Hz' lower; 'vid_pitchsd' lower) is associated with lower emotional activation. A lower Mel-frequency cepstrum coefficient 2 ('M2') appears associated with heightened activation, and lower local absolute jitter also associates with higher emotional activation.

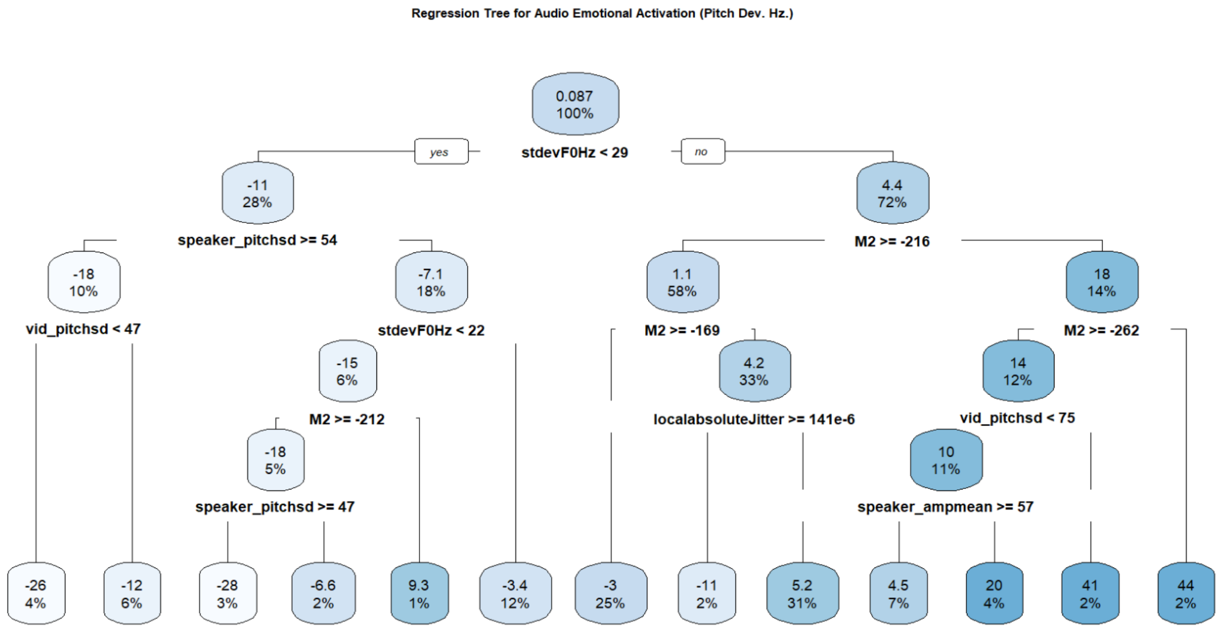
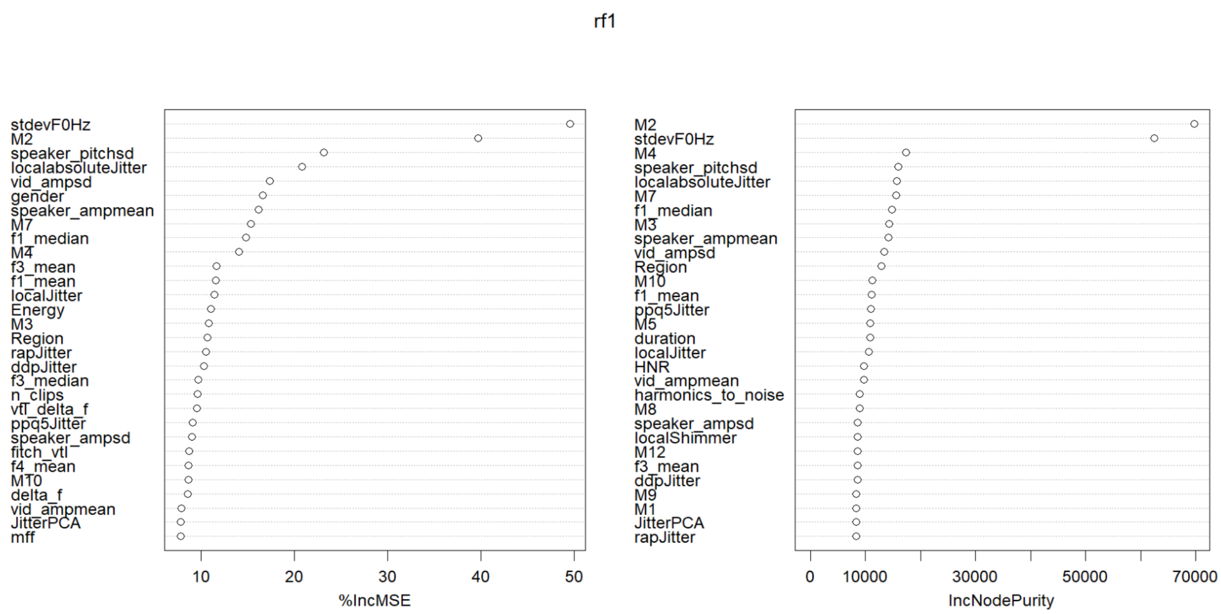


Figure 6: Tree with pitch deviation in Hz, continuous, as the outcome variable.

Across 1000 tree models, with each split constrained to consider only 30 of the roughly 60 audio variables in the dataset, the most consistently important variables in improving predictions were, again, the audio clip's standard deviation in pitch; M2; others relating to pitch variation (speaker's standard deviation in pitch across all videos, 'speaker_pitchsd'), and local jitter.



Metadata variables including speaker's gender and region (province) also emerged as moderately influential. (Men having a higher pitch deviation, on average, across the dataset, at 0.26 versus -0.27 for women). Two variables indicating the clip's duration in seconds, 'duration', and the speaker's number of clips in the dataset, 'n_clips', were relatively unimportant, confirming that these two variables do not heavily influence the effects measured.

Pitch Standard Deviation as a Possible Alternative Measure of Activation?

One downside of the main audio variable used to detect emotional activation, a speaker rising above their typical speaking pitch, is that it requires multiple sound clips from each speaker to determine the correct baseline. This forces the deleting of some audio clips, and does not work as well in datasets with many speakers with few observations each than in tracking a few prominent speakers with extensive footage each (e.g., Dietrich et al.'s work on the US Supreme Court, 2019a). In some contexts, such as parliamentary debates, speakers may be in a state of heightened emotions relatively often – nearly half of the full dataset (44.4%) was labelled by manual coders as activated.

The predominance of a video's standard deviation in pitch (stdevF0Hz; vid_pitchsd) in inductive models used to find additional signals of activation means that standard deviation in a short audio clip's pitch may work as an alternative signal of emotional arousal. Previous literature does not (in my reading; to my knowledge) flag pitch varying more widely as a sign of agitation, but it may also reflect physiological arousal – faster heartbeat, tenser vocal cords, and so on. A speaker may potentially be jumping between their regular pitch and a heightened pitch when activated, or they may not have a consistent higher pitch when stressed or otherwise activated.

Of course, while the relationship between deviation in pitch above a speaker's mean (pitch deviation) and the standard deviation of pitch *within* a clip is positive for both genders in the dataset, women's pitch deviation tends to be higher by 3-12 Hz, depending on the specific audio software package used). As such, prior testing in other datasets and an appropriate control variable will likely be needed. No relationship was found between a clip's standard deviation in pitch and the speaker's age, or geographic region.

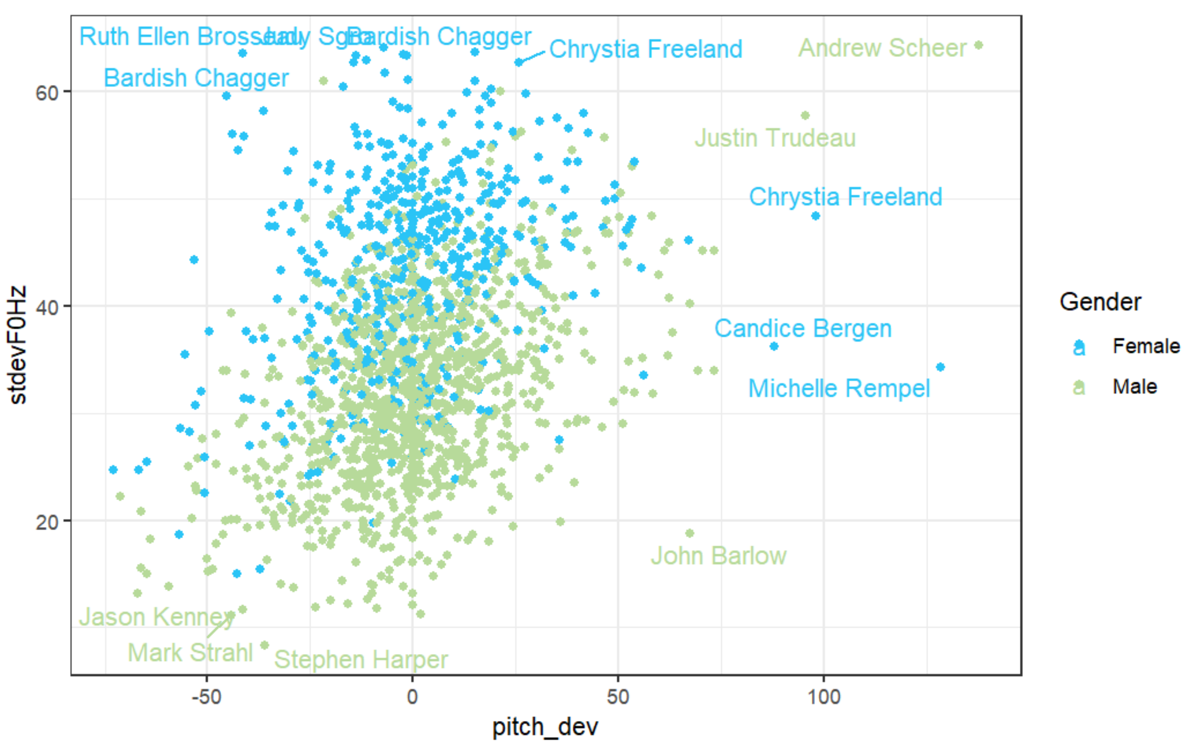


Figure 7: Gender differences in audio clip's standard deviation in pitch.

High variation in pitch within a clip may also indicate heckling, or multiple speakers.

Determining Sentiment from Audio Signals?

Taking advantage of the manually tagged positive and negative sentiment of each audio clip, the categorical outcomes of positive or negative sentiment were predicted using automatically ascertained features.

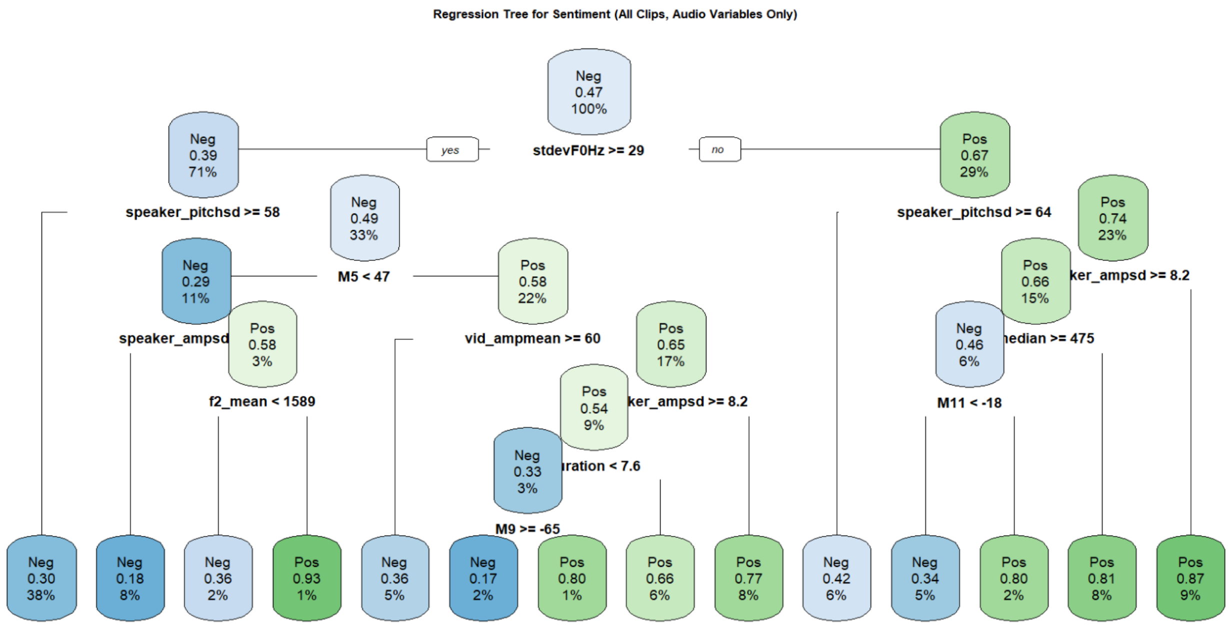


Figure 8: Automatic feature selection of predictors of positive (1) or negative (0) sentiment.

Within activated videos only, predictors of the negative sentiment category included higher variation in pitch (suggesting that negative videos are perhaps more activated), louder speech (vid_ampmean), and a more varied volume of speech (speaker_ampsd). A lower F2 mean (a higher harmonic of the 'fundamental frequency', heard as pitch, F0) was associated with negative sentiment, as was lower M5 and M11, both mel-frequency cepstrum coefficients. This finding may simply suggest that negative emotions are simply felt and expressed more strongly in the parliamentary setting than positive emotions are – in this data, at least, speakers simply seem to be more emotionally activated when they are negative.

A forest of 1000 trees showed that standard deviation in pitch, amplitude mean, and amplitude standard deviation were the top variables predicting sentiment. Number of clips was also relatively high in the mix of important variables, which may be due to Liberal speakers being much more frequently positive than others, and the presence of many frequently-speaking Liberal ministers in the dataset). Subsetting observations to *only* those that were emotionally activated, either based on manual tags or by audio-based cutoffs, consistently identified these variables as important, with the last specification mentioned also flagging low f1 median, another higher harmonic of pitch, as associated with negative sentiment. (See Appendix).

A cautionary note about amplitude-related measures is that they may not be usable in all datasets. The parliamentary clips were all recorded on the same microphones, with speakers all present in person in the chamber. Datasets where the recording conditions are more varied should perhaps not include amplitude-related variables. Nonetheless, as louder speech may indicate either strong negative emotions like anger, or positive excitement, amplitude may be valuable when recording conditions are consistent.

The mean squared residual within the training data, 80 percent of the relevant videos, was 359.10 (Hz), and the mean squared residual within unseen testing data was 329.99. Additional variations of sentiment-predicting tree and forest models are shown in the Appendix.

Discussion and Conclusions

While the findings regarding ideological grouping and the expression of emotional arousal in Parliament are limited, and at times against expectations, some models based on manual labels do show that the left-populist NDP, and right-populists among the Conservatives, engage in emotionally activated speech more than their centre-left (Liberals), New Left (Bloc; Green Party), or mainstream right counterparts. The centre-left Liberals also appear to be much more positive in emotional register than other parties, even before taking office in October 2015. However, audio-based models do not show the same ideological differences in activation. (This may be due to artifactual problems with the high baseline of emotional arousal in Question Period, or could reflect manual coders perhaps recognizing higher levels of emotion in populist parties's MPs).

Future work could perhaps benefit from obtaining more clips of speakers in more emotionally subdued settings, such as interviews, as House of Commons debate is frequently (44.4 percent of observations) confrontational and emotionally agitated. Further variables relating to speaker's articulation rate (speed of talking; pauses; De Jong et al. 2021) or zero-crossing rate (implicated in some, but not all, anxiety-in-speech studies, Teferra et al. 2022) may help in distinguishing between different negative emotions (anxiety, sadness, anger), or activation level. The speaker-specific 'voice encoding' approach pursued by Rheault and Borwein (2019) could form another alternative to looking at distinct audio variables, as done here. Text transcripts may also be promising in helping to identify specific emotions expressed, or, perhaps, specific subjects that different politicians associate with anger, for example.

The audio variables indicating emotional arousal – pitch above a speaker's typical level, and possibly some others identified, such as high variation in pitch within an audio clip, or louder speech – seem to be particularly useful where text transcripts alone fail. Literature on political disaffection (Rhodes-Purdy et al. 2023; Mair 2013) points to a common belief among the disenchanted that politicians are often lying, putting on excessively false performances, and otherwise have untrustworthy character traits. The subtle physiological cues involved in emotional arousal could help to test which politicians seem to be more emotionally engaged in the sad, outraged, or worried statements they are making. The uses of specific emotions do appear to differ somewhat by party or ideological bloc, and so more investigation into audio and multimodal data, perhaps from different settings, could help to illuminate patterns further.

Acknowledgements

Many thanks to Ludovic Rheault for recommending the Parselmouth package in Python, and for the idea to run regressions with control variables for some of the main audio features.

(References, libraries)

The Python-based Parselmouth interface (Jadoul et al. 2018) to phonetics program Praat (Boersma and Van Heuven 2001; Boersma and Weenink 2021) provided the majority of audio measurements, as did subsequent analyses in R. FFMPEG was used to convert video files into audio-format .wavs.

References

Albertazzi, Daniele, and Donatella Bonansinga. 2023. "Beyond anger: the populist radical right on TikTok." Journal of Contemporary European Studies: 1-17.

Bickerton, Christopher J., and Carlo Invernizzi Accetti. 2021. Technopopulism: The new logic of democratic politics. Oxford University Press.

Boersma, Paul, and David Weenink. 2021. "Praat: doing phonetics by computer computer program." Version 6.1.38, retrieved 2 January 2021 from http://www.praat.org/.

Boersma, Paul, and Vincent Van Heuven. 2001. "Speak and unSpeak with PRAAT." Glot International 5, no. 9/10: 341-347.

Bonansinga, Donatella. 2020. "Who thinks, feels. The relationship between emotions, politics and populism." Partecipazione e conflitto 13, no. 1: 83-106.

Breiman, Leo. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." Statistical Science 16, no. 3: 199-231.

Cinar, Ipek, Susan Stokes, and Andres Uribe. 2020. "Presidential rhetoric and populism." Presidential Studies Quarterly 50, no. 2: 240-263.

Cochrane, Christopher. 2015. Left and Right: The Small World of Political Ideas. McGill-Queen's University Press-MQUP.

Cochrane, Christopher, Ludovic Rheault, Jean-François Godbout, Tanya Whyte, Michael W-C. Wong, and Sophie Borwein. 2022. "The automatic analysis of emotion in political speech based on transcripts." Political Communication 39, no. 1: 98-121.

De Jong, N.H., Pacilly, J., & Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, Assessment in Education: Principles, Policy & Practice, 28:4, 456-476, DOI: 10.1080/0969594X.2021.1951162

Dietrich, Bryce J., Ryan D. Enos, and Maya Sen.2019. ("2019a"). "Emotional arousal predicts voting on the US Supreme Court." Political Analysis 27, no. 2: 237-243.

Dietrich, Bryce J., Matthew Hayes, and Diana Z. O'brien. 2019. ("2019b"). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech." American Political Science Review 113, no. 4: 941-962.

Dietrich, Bryce J., Jeffery J. Mondak, and Tarah Williams. 2019. ("2019c"). Using the Audio from Telephone Surveys for Political Science Research." Midwest Political Science Association, Chicago, IL (2019).

Dyson, Stephen Benedict. 2019. Imagining Politics: Interpretations in Political Science and Political Television. University of Michigan Press.

Frank, Thomas. 2020. The People, No: A Brief History of Anti-Populism. Metropolitan Books.

Gidron, Noam. 2016. "Many Ways to Be Right: The Unbundling of European Mass Attitudes and Partisan Asymmetries across the Ideological Divide." PhD diss., Harvard University.

Gidron, Noam, and Peter A. Hall. 2017. "The politics of social status: Economic and cultural roots of the populist right." The British Journal of Sociology 68: S57-S84.

Gu, Simeng, Fushun Wang, Nitesh P. Patel, James A. Bourgeois, and Jason H. Huang. 2019. "A model for basic emotions using observations of behavior in Drosophila." Frontiers in Psychology 10: 781.

Hanes, Rachel, and Mackenzie Gray. 2022. "Which Conservative MPs attended anti-vaccine presentation from convoy figures?" CTV News, 23 June, 2022 (26 June, 2022 version; accessed August 14, 2023). https://www.ctvnews.ca/politics/which-conservative-mps-attended-anti-vaccine-presentation-from-convoy-figures-1.5959764

Hirst, Graeme, Yaroslav Riabinin, Jory Graham, Magali Boizot-Roche, and Colin Morris. 2014. "Text to Ideology or Text to Party Status?" From Text to political positions: Text analysis across disciplines 55: 93-115.

Inglehart, Ronald, and Pippa Norris. 2017. "Trump and the populist authoritarian parties: The silent revolution in reverse." Perspectives on Politics 15, no. 2: 443-454.

Jadoul, Yannick, Bill Thompson, and Bart de Boer. (2018). Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71, 1-15. https://doi.org/10.1016/j.wocn.2018.07.001

Jaeger, Anton. 2023. "Populism and the Historians: Richard Hofstadter and the Birth of a Global Populism Debate." History of Political Thought 44, no. 1: 153-194.

Krause, Werner, Denis Cohen, and Tarik Abou-Chadi. 2023. "Does accommodation work? Mainstream party strategies and the success of radical right parties." Political Science Research and Methods 11, no. 1: 172-179.

Levesque, Catherine. 2023. "A year later, have the Conservatives embraced Freedom Convoy politics?" The National Post, 29 January 2023 (Accessed 14 August 2023). https://nationalpost.com/news/politics/conservatives-detente-over-freedom-convoy-endures

Mair, Peter. 2013. Ruling the void: The hollowing of Western Democracy. London: Verso Books.

Major, Sara, and Aleksandar Tomašević. 2023. "The Face of Populism: Examining Differences in Facial Emotional Expressions of Political Leaders Using Machine Learning." arXiv preprint arXiv:2304.09914 (2023). https://arxiv.org/pdf/2304.09914.pdf

Meguid, Bonnie M. 2005. "Competition between unequals: The role of mainstream party strategy in niche party success." American Political Science Review 99, no. 3: 347-359.

Moffitt, Benjamin. 2016. The global rise of populism: Performance, political style, and representation. Stanford University Press.

Mols, Frank, and Jolanda Jetten. 2016. "Explaining the appeal of populist right-wing parties in times of economic prosperity." Political Psychology 37, no. 2 (2016): 275-292.

Mudde, Cas. 2004. "The populist zeitgeist." Government and Opposition 39, no. 4: 541-563.

Nash, Chelsea. 2022. "Conspiracy theories and disinformation in Canadian politics amped up after 'Freedom Convoy' ". The Hill Times. 28 February. Accessed 14 August 2023. https://www.hilltimes.com/story/2022/02/28/conspiracy-theories-and-disinformation-in-canadian-politics-amped-up-after-freedom-convoy/230012/

Neumann, Markus. 2019. "Hooked with phonetics: The strategic use of style-shifting in political rhetoric." In Annual Meeting of the American Political Science Association. Washington, DC. 2019.

Neumann, Markus. 2020. "Nonverbal Political Communication in the US." PhD Diss. https://etda.libraries.psu.edu/files/final_submissions/21734

Norris, Pippa. 2020. "Measuring populism worldwide." Party Politics 26, no. 6: 697-717.

Payton, Laura. 2013. "NDP votes to take 'socialism' out of party constitution". CBC News, 14 April. Accessed 17 August 2023. https://www.cbc.ca/news/politics/ndp-votes-to-take-socialism-out-of-party-constitution-1.1385171

Pedersen, Kylee. 2022. "Michelle Rempel Garner is giving 'serious consideration' to UCP bid." CBC News, 15 June 2022, accessed 14 August 2023. https://www.cbc.ca/news/canada/calgary/michelle-rempel-garner-ucp-leadership-race-1.6490534

Ranciere, Jacques. 2022. Les trente inglorieuses: scenes politiques 1991-2021. Paris: La Fabrique editions.

Rheault, Ludovic, and Sophie Borwein. 2019. Multimodal techniques for the study of affect in political videos. Working Paper, Preliminary version prepared for the 2019 PolMeth Conference, MIT, Cambridge, MA, July 18-20. https://polmeth.mit.edu/sites/default/files/documents/RheaultBorwein_PolMeth2019.pdf

Rheault, Ludovic, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. "Measuring emotion in parliamentary debates with automated textual analysis." PloS one 11, no. 12: e0168843.

Rhodes-Purdy, Matthew, Rachel Navarre, and Stephen Utych. 2023. The Age of Discontent. Cambridge University Press.

Teferra, Bazen Gashaw, Sophie Borwein, Danielle D. DeSouza, William Simpson, Ludovic Rheault, and Jonathan Rose. 2022. "Acoustic and linguistic features of impromptu speech and their association with anxiety: validation study." JMIR Mental Health 9, no. 7: e36828.

Ward, George, Jan-Emmanuel De Neve, Lyle H. Ungar, and Johannes C. Eichstaedt. 2021. "(Un) happiness and voting in US presidential elections." Journal of Personality and Social Psychology 120, no. 2: 370.

Appendix: Additional Images, Tables, and Tests

Detailed Note on Conservative Party (CPC) 'Ideology' Labels

Ten MPs in the audio dataset had either former Progressive Conservative (PC) or Reform affiliations: Bernard Valcourt and Rob Nicholson were sorted to the 'centre-right' category due to their former Progressive Conservative affiliation (One Liberal MP, Scott Brison, and one Bloc member, Louis Plamondon, also formerly belonged to the PCs). Stephen Harper, Jason Kenney, Deepak Obhrai (also a former Progressive Conservative), and Gerry Ritz came from the neoliberal-populist Reform Party, as did Scott Reid and Kevin Sorenson from Reform's successor party, the Canadian Alliance. The Reform Party presented itself publicly as non-ideological, but party positions hewed closely to a free-market, socially conservative, and somewhat anti-establishment profile (Moez 2023, CPSA). (Stephen Harper and Deepak Obhrai, however, are not classed with the rest of the former Reformers as populist-right, Obhrai due to his PC affiliation and later moderate stances; Harper due to his largely conventional and uneventful time in office).

Out of the remaining Conservative MPs in the data (total number of 81), vote-based clustering identified one potential small cluster of further-right or unconventional right MPs: Pierre Poilievre, Rick Norlock, Kevin Sorenson, Jeff Watson, James Bezan, and Gerry Ritz. (Dean Allison, Daryl Kramp and Ron Cannan belong to the same vote-based group, but are not present in the final dataset used as they have only 1 speech clip each). [Dean Allison is among the 24 MPs who reportedly met with Freedom Convoy leadership, Hanes and Gray 2022; Poilievre has also described himself as 'anti-establishment' (Levesque 2023) and has sent out party messaging decrying the "globalist Davos elites" at the World Economic Forum (Canadian Press 2023)).]

Subsequently, high-profile Conservative MPs, such as top leadership candidates and party leaders, were sorted based on general knowledge (confirmed by web searches for media coverage in unclear cases) into a default, centre-right through mainstream right 'Conservative' group (Rona Ambrose, Lisa Raitt, Chris Alexander, Steven Blaney) and a more unconventional right or further right 'right-populist' group. MPs sorted to the right-populist group in this manner were Maxime Bernier, who split from the party to create his own right-populist People's Party of Canada, after narrowly losing a leadership election in 2017; Michelle Rempel (see, e.g., Pedersen 2022 for her unconventional mix of policy views), and Bradley Trost, a strong social conservative.

Based on Hanes and Gray's reported list of the 24 Conservative MPs (20 percent of the total) who attended the Freedom Convoy leadership's visit to Parliament (2022), Cheryl Gallant, Chris Warkentin, James Bezan, John Barlow, Marilyn Gladu, Martin Shields, and Ted Falk were also added to the 'right-populist' group. Mark Strahl, incidentally the son of prominent Reform MP Chuck Strahl, and Pierre Paul-Hus, were also vocal critics of the Trudeau government's actions on vaccine mandates and the trucker convoy (Nash 2022).[13]

[13] Additional Conservatives included in the right-populist list, who do not appear in some of the mail models as they have only one audio statement each, are: Ron Cannan, Rick Norlock, Dean Allison, and Daryl Kramp, based on the vote-based clustering (Dean Allison was also reportedly the sponsor of the Freedom Convoy leadership's visit to Parliament); and Melissa Lantsman, Dan Muys, Leslyn Lewis, Alex Ruff, Jeremy Patzer, Jamie Schmale, Ryan Williams, Warren Steinley, Demien Kurek, Gerald Soroka, Scott Davidson, Corey Tochor, and Tako van Popta, based on reported attendance at the Freedom Convoy parliamentary event (Hanes and Gray 2022).

Regression Models Using Manual Labels

Model 1 shows binarized activation, manual tag, as the dependent variable; Model 2, binarized sentiment; Model 3, binarized sentiment for only January to October 19th, 2015, prior to the Liberals governing. All three are binary logistic regression models.

Table 6:

	Dependent variable:		
	activation	sentiment	
	(1)	(2)	(3)
ideologyBloc	−0.230	-1.583^{***}	−0.124
	(0.575)	(0.607)	(0.684)
ideologyConservative	-0.286^{**}	-1.349^{***}	-1.483^{***}
	(0.136)	(0.140)	(0.270)
ideologyGreen	0.528	-1.688^{**}	
	(0.768)	(0.841)	
ideologyNDP	0.536^{***}	-2.420^{***}	-2.540^{***}
	(0.144)	(0.183)	(0.414)
ideologyRightPopulist	−0.100	-1.360^{***}	-1.025^{***}
	(0.181)	(0.187)	(0.346)
Constant	-0.240^{***}	0.772^{***}	−0.099
	(0.077)	(0.083)	(0.134)
Observations	1,472	1,472	519
Log Likelihood	−998.356	−881.623	−274.647
Akaike Inf. Crit.	2,008.711	1,775.245	559.294
Note:	*p<0.1; **p<0.05; ***p<0.01		

Systematic Differences in Audio-Manual Codes' Alignment?

Accuracy scores for manual codes for 'activation' compared to an audio-based cutoff of 3.45Hz deviation above speaker's mean ('Acc_AudioManual'), by speaker:

As seen, there are some systematic differences in emotion-scoring accuracy by group. (In this context, the accuracy metric refers to the proportion of total cases where manual coders' categorical scores concur with audio-based categorical scores. Bloc Quebecois MPs and women have systematically less agreement between their audio-based scores and manual scores, while younger MPs (born 1980s) have higher alignment between the two measures, while holding all else constant. Region or province did not appear to matter.

Manual labels shaded by number of coders

To test whether some potential mis-labellings of activation or sentiment were due to fewer video coders watching the clip, the scores below are shaded by number of coders:

Main results

Main models, with demographic variables added (Region reference category = Ontario; Party reference category = Liberal) show only a slight tendency toward higher (Models 1-3) or more frequent (logistic regression, Model 4) emotional activation in Alberta. The ideological factions with the most frequently

Table 7:

	Dependent variable:	
	Acc_AudioManual	
	(1)	(2)
gender	-0.025^{**}	-0.023^{**}
	(0.010)	(0.011)
dobdecade1950s	-0.020	-0.020
	(0.020)	(0.020)
dobdecade1960s	-0.019	-0.020
	(0.018)	(0.018)
dobdecade1970s	-0.036^{*}	-0.030
	(0.019)	(0.019)
dobdecade1980s	0.064^{***}	0.063^{***}
	(0.023)	(0.023)
partyBloc	-0.106^{**}	-0.098^{*}
	(0.050)	(0.051)
partyConservative	0.025^{**}	0.026^{**}
	(0.011)	(0.011)
partyGreen	0.122^{*}	0.134^{*}
	(0.069)	(0.070)
partyNDP	-0.009	-0.002
	(0.013)	(0.014)
regionAB		0.016
		(0.016)
regionAtlantic		0.038^{**}
		(0.019)
regionBC		-0.003
		(0.017)
regionMB		0.038^{*}
		(0.022)
regionNorth		0.030
		(0.039)
regionQC		0.004
		(0.013)
regionSK		-0.008
		(0.025)
Constant	0.638^{***}	0.626^{***}
	(0.017)	(0.019)
Observations	1,462	1,462
R^2	0.028	0.033
Adjusted R^2	0.022	0.022
Residual Std. Error	0.178 (df = 1452)	0.178 (df = 1445)
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

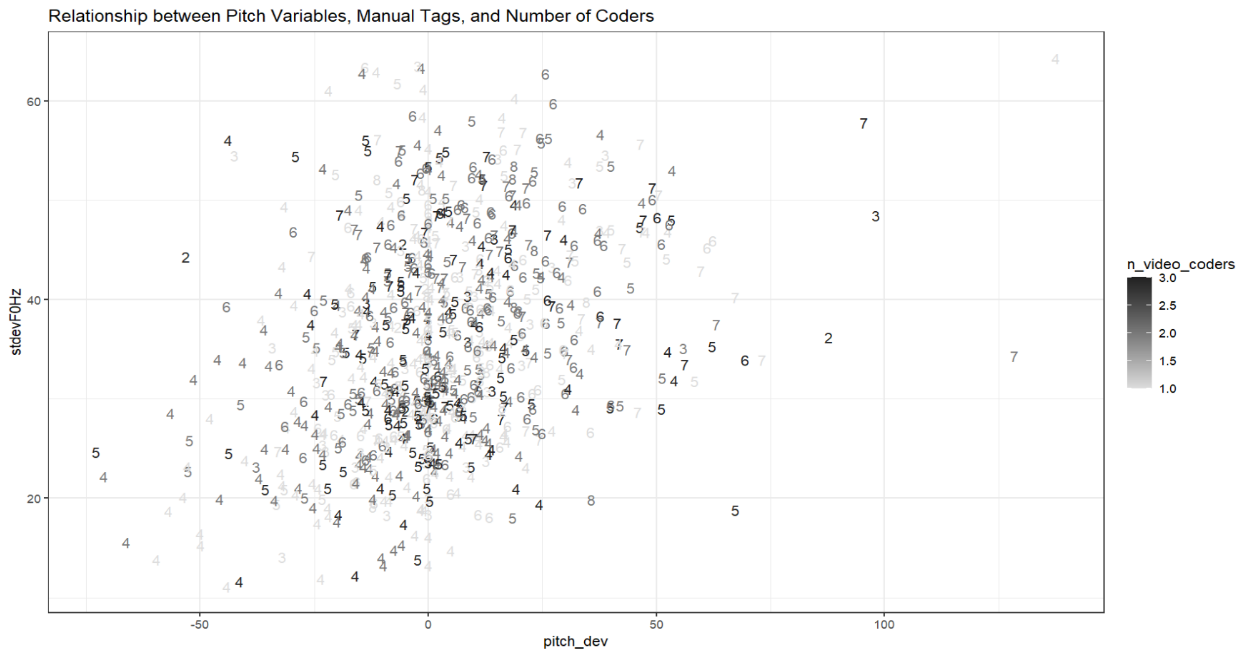


Figure 9: The relationship between speaker's deviation in pitch above the typical mean, standard deviation of pitch in the video clip, and manual tags for activation.

activated speeches, in Alberta, are the left-populist NDP and individual right-populist Conservatives, but a model with either an interaction between party and region, or one subset to Alberta only, does not find this ideological difference to be statistically significant. Nonetheless, the direction of the effect is suggestive.

(Models are shown on the following page).

English-French language effects? Unfamiliar language effects?

While language spoken has not (to my knowledge) come up in previous political science literature using audio data, which primarily has focused on US politics, it seems at least possible that: (a) a bilingual speaker may tend to shift pitch in a second language, or a language may tend to have higher or lower pitch in general, compared to English. It also seems plausible that a speaker tyring to make a statement in a highly unfamiliar language (risking ridicule or other negative reactions if the performance is not done well) might experience physiological activation when making the statement – high focus may be required, and nerves may be higher.

The reference category is 'ENEN', English language videos by speakers comfortable in English. 'FREN' means a French-language clip spoken by an unfamiliar-with-French speaker. The effect direction is suggestive, but not conclusive.

Table 8:

	Dependent variable:			
	pitch_dev			pitchdev_cat
	OLS			*logistic*
	(1)	(2)	(3)	(4)
ideologyBloc	5.806	0.407	−2.210	−0.017
	(6.455)	(9.879)	(7.645)	(0.589)
ideologyConservative	0.376	0.258	0.223	−0.121
	(1.603)	(2.510)	(1.871)	(0.146)
ideologyGreen	4.291	3.438	−4.636	0.582
	(8.897)	(11.341)	(8.736)	(0.798)
ideologyNDP	0.039	−3.039	-3.627^{*}	−0.026
	(1.760)	(2.407)	(2.042)	(0.160)
ideologyRightPopulist	2.182	1.960	0.630	0.202
	(2.190)	(3.307)	(2.396)	(0.196)
gender	0.055	−1.175	0.471	0.096
	(1.387)	(1.938)	(1.618)	(0.126)
dobdecade1950s	1.090	−1.484	−1.392	−0.022
	(2.521)	(3.708)	(2.957)	(0.231)
dobdecade1960s	1.887	0.085	−2.680	0.152
	(2.332)	(3.537)	(2.722)	(0.213)
dobdecade1970s	1.359	−0.986	−2.109	0.154
	(2.489)	(3.726)	(2.894)	(0.227)
dobdecade1980s	0.033	3.261	−0.157	−0.024
	(2.948)	(4.337)	(3.453)	(0.270)
regionAB	1.343	6.609^{*}	1.370	0.385^{**}
	(2.071)	(3.450)	(2.318)	(0.187)
regionAtlantic	2.017	2.115	1.230	0.326
	(2.404)	(3.342)	(2.706)	(0.216)
regionBC	−0.576	−0.376	−2.292	0.278
	(2.185)	(3.005)	(2.468)	(0.197)
regionMB	−1.737	1.364	2.235	0.271
	(2.790)	(3.847)	(3.123)	(0.251)
regionNorth	1.227	−3.224	−5.227	−0.046
	(4.996)	(15.405)	(6.022)	(0.461)
regionQC	−0.050	0.703	0.231	0.159
	(1.701)	(2.374)	(2.019)	(0.155)
regionSK	1.279	−0.440	3.429	0.075
	(3.150)	(5.098)	(3.773)	(0.290)
Constant	−1.822	6.354^{*}	22.212^{***}	-0.646^{***}
	(2.408)	(3.720)	(2.850)	(0.221)
Observations	1,462	656	597	1,462
R^2	0.004	0.020	0.023	
Adjusted R^2	−0.008	−0.006	−0.006	

Table 9:

	Dependent variable:		
	pitch_dev		
	(1)	(2)	(3)
vidlangFR	−0.490		
	(1.545)		
langmismatch		−1.079	
		(2.607)	
langmixENFR			−4.898
			(4.442)
langmixFREN			0.632
			(3.174)
langmixFRFR			−0.910
			(1.705)
Constant	0.171	0.144	0.276
	(0.642)	(0.600)	(0.649)
Observations	1,472	1,472	1,472
R^2	0.0001	0.0001	0.001
Adjusted R^2	−0.001	−0.001	−0.001
Residual Std. Error	22.403 (df = 1470)	22.402 (df = 1470)	22.407 (df = 1468)
Note:			*p<0.1; **p<0.05; ***p<0.01

Inductive and Exploratory Analyses: Tree and Forest Models, Cont.

Initial tree models using all variables available primarily relied on mean pitch in the video (No limitation on number of variables considered per split; Minimum of 10 cases per end node required):

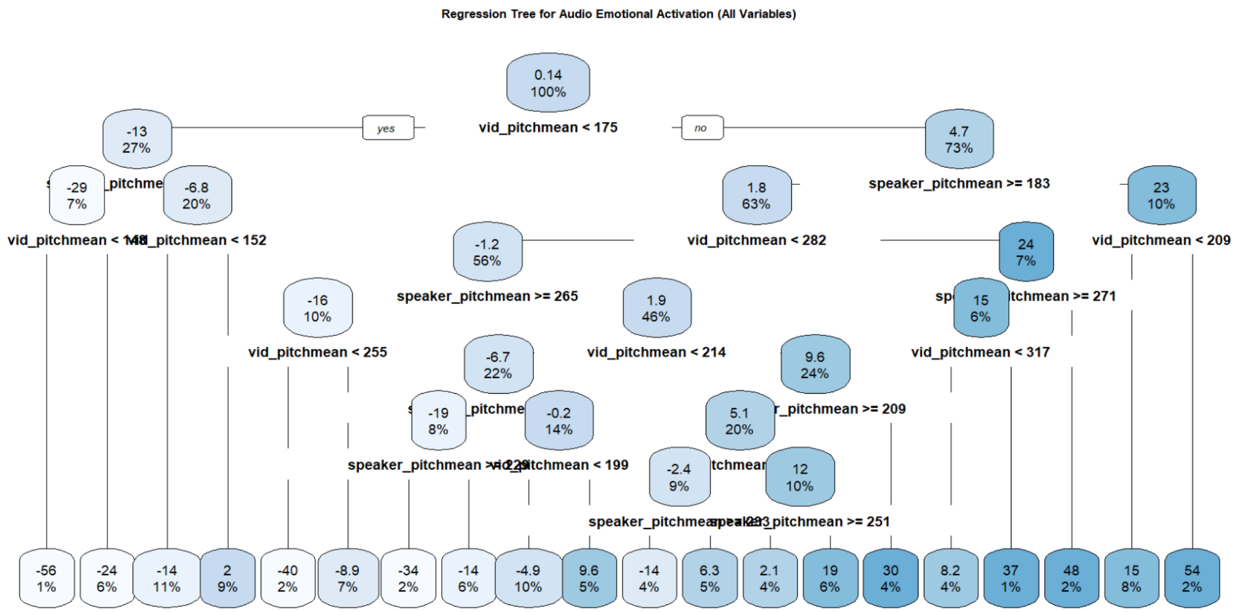


Figure 10: First tree, all variables including metadata.

A second tree that included metadata variables (speaker's information), but excluded the video's mean pitch, resembled the final audio-variable-only tree shown in the main analysis:

Additional tree and forest models predicting sentiment

Any tree or forest model built to predict sentiment relied heavily on speaker's party or ideology when metadata variables were included. Age, but not geographic region, and manually labelled emotional activation, also came up as top metadata variables of importance.

With all audio clips considered, the random forest version of the example tree shown has variable importance:

While all major parties had a similar *average* number of clips per unique speaker, the n_clips variable was removed from some models as it appeared to capture party affiliation. Most speakers with 20 or more clips each are Liberal (6 Liberals had 20+ clips; versus only 3 from the Conservatives and 1 NDP MP). This difference did not seem to matter for emotional activation-predicting models, but the Liberals are much more positive in sentiment.

Predicting sentiment within activated videos only, either by manual activation tag (RF2) or based on the audio (pitch deviation) cutoff (RF2B), negative audio clips still appeared to be more strongly activated, with higher pitch standard deviation, higher amplitude, and higher variation in amplitude.

Example trees for the former two models, respectively (RF2; RF2B) are as follows:

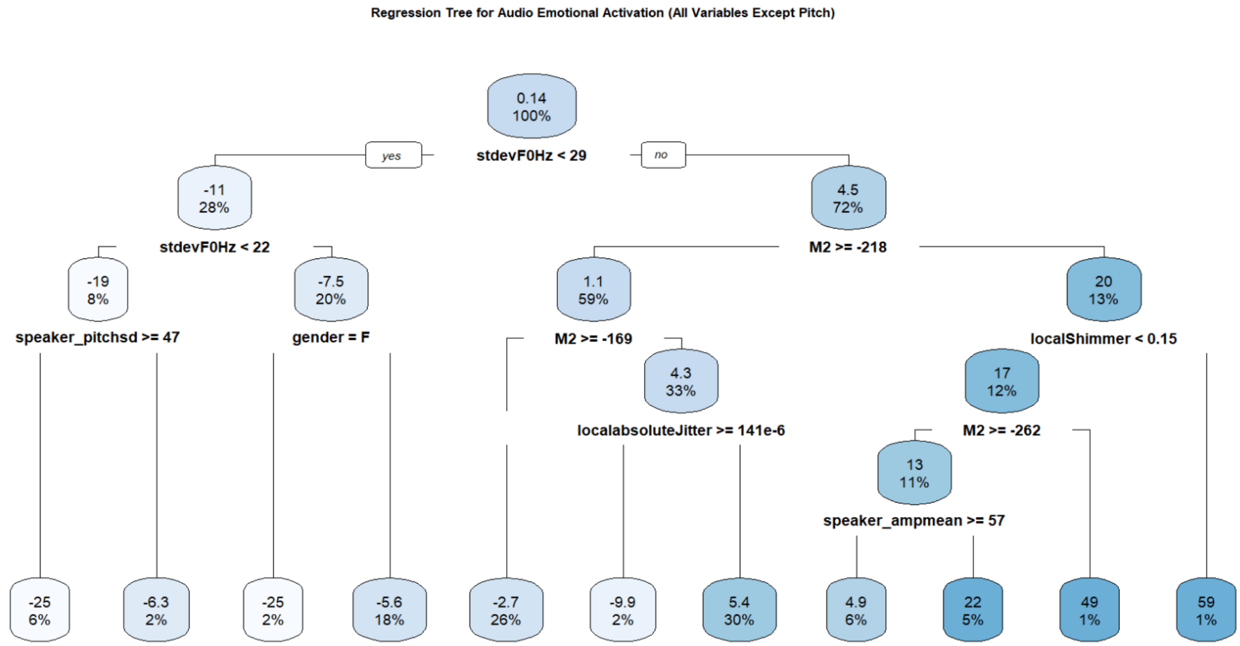


Figure 11: Second tree, most audio variables, and including metadata.

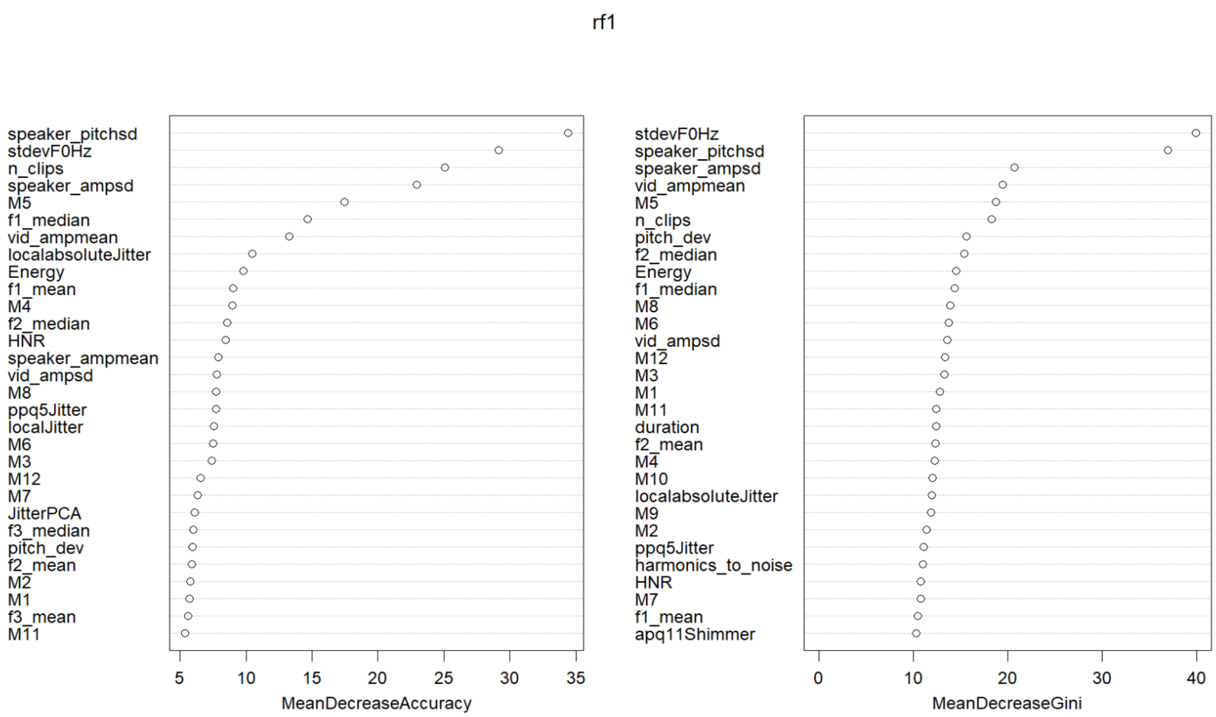


Figure 12: Automatic feature selection of predictors of positive (1) or negative (0) sentiment, all clips, audio variables only (mostly).

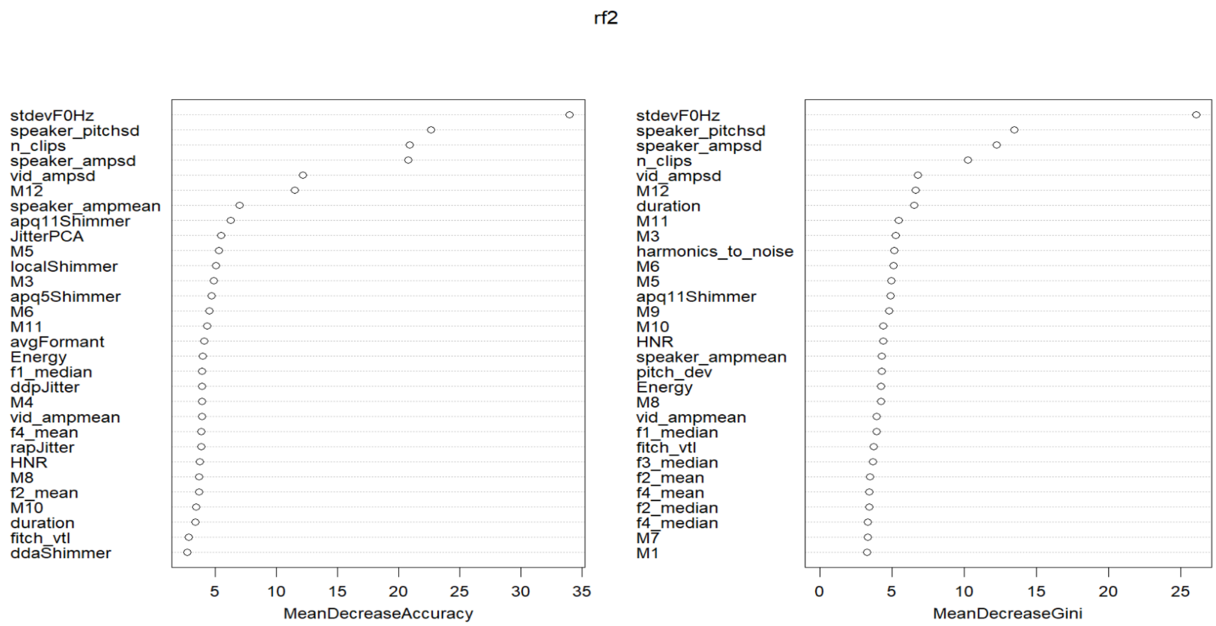


Figure 13: Predictors of positive or negative sentiment within activated clips only; activation determined by manual label.

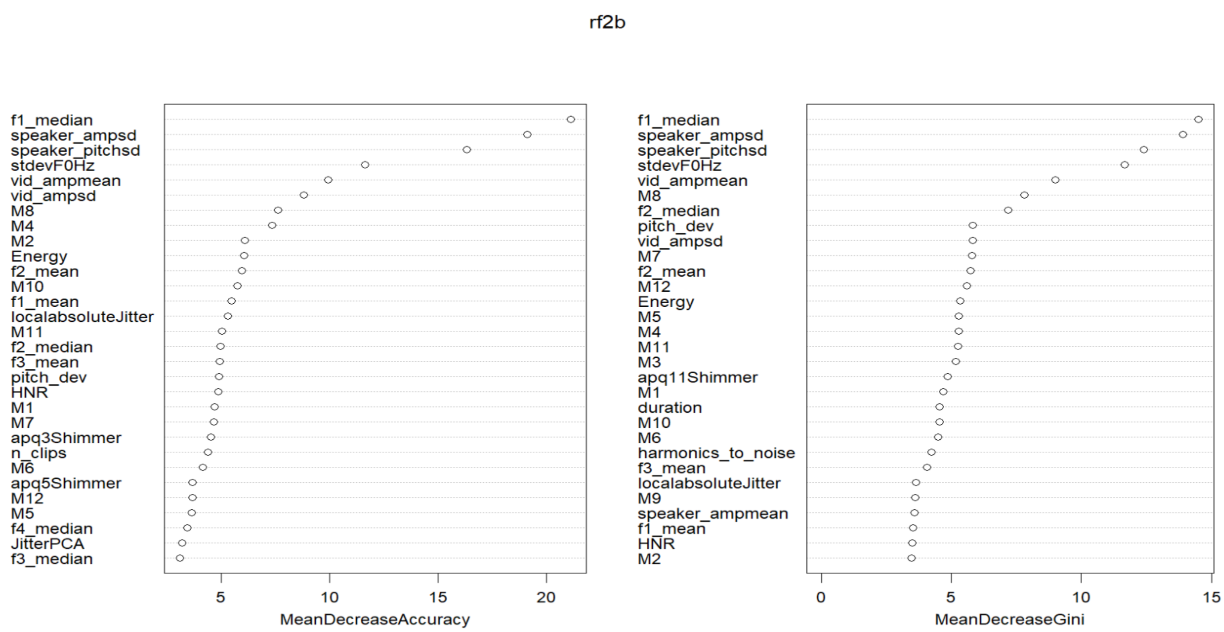


Figure 14: Predictors of positive or negative sentiment within activated clips only; activation determined by audio cutoff.

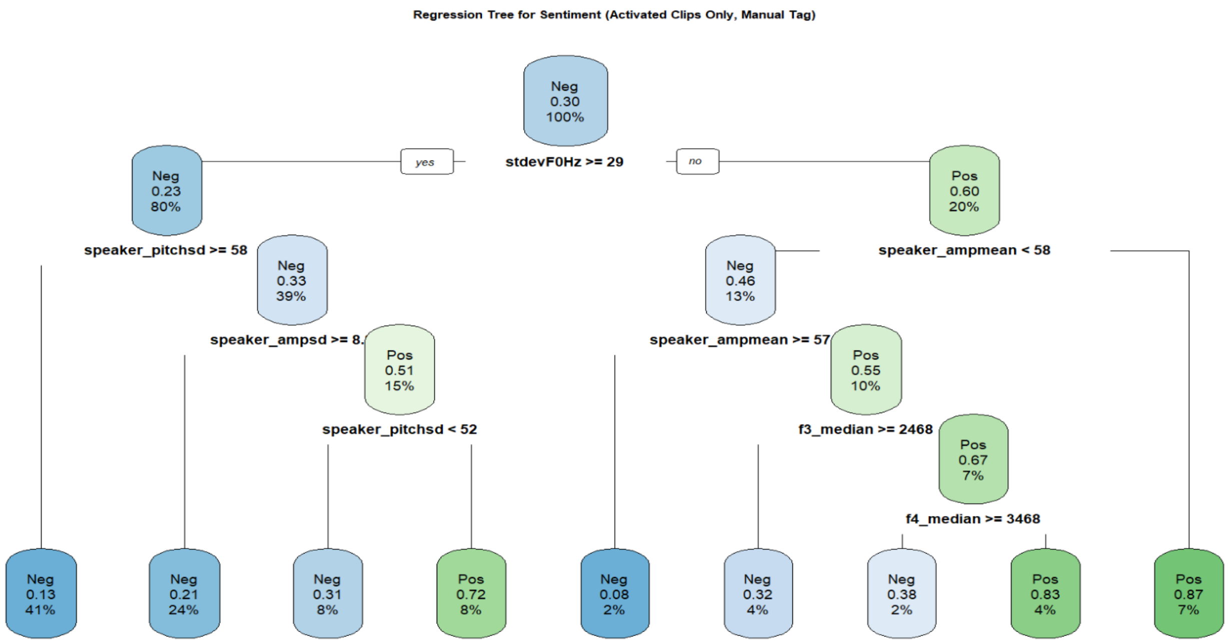


Figure 15: Predictors of positive or negative sentiment within activated clips only; activation determined by manual label; example tree.

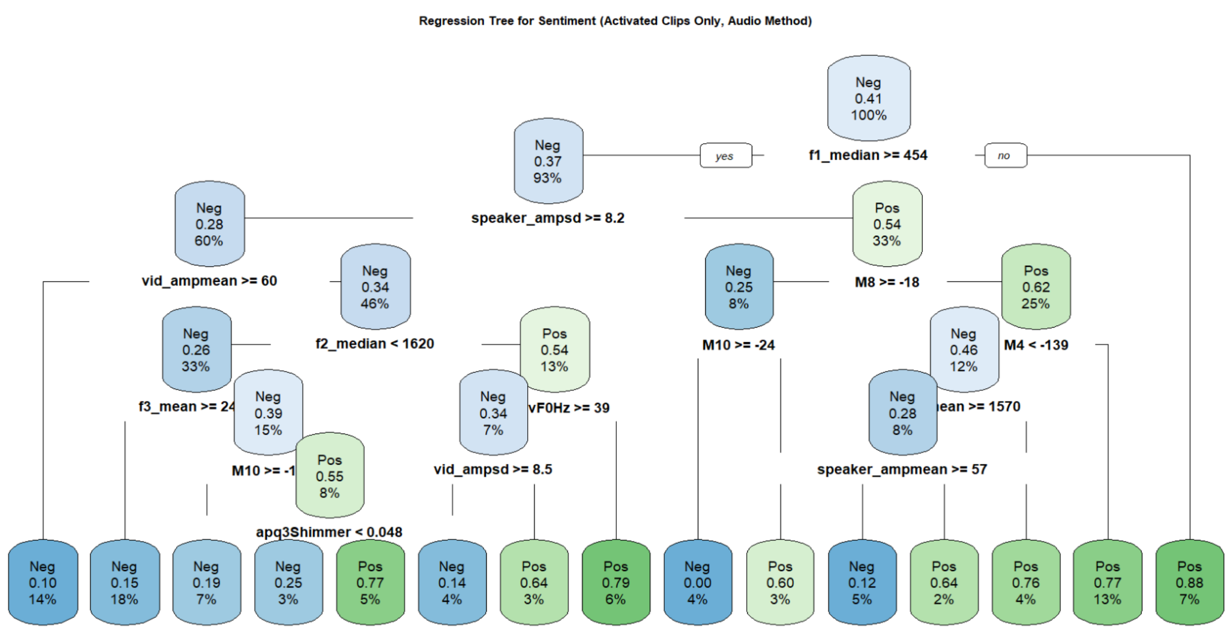


Figure 16: Predictors of positive or negative sentiment within activated clips only; activation determined by audio cutoff; example tree.