Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

Prosodic features of awe: What makes the voice sound in awe?

Valérie Coulombe[1], *Vincent Martel-Sauvageau*[2], *Laura Monetta*[2], *Maxime Montembeault*[1]

[1] McGill University, Québec, Canada
[2] Université Laval, Québec, Canada
valerie.coulombe@mail.mcgill.ca

Abstract

Awe is a complex positive emotion whose vocal expression has received little attention compared with well-studied emotions such as joy or anger. Although listeners can identify awe from prosody alone, its acoustic profile remains largely unknown. This study characterizes the prosodic profile of awe and identifies the acoustic cues that support its recognition. Thirty-one Québec French speakers produced one-word utterances (/papa/) in response to eliciting scenarios for six emotions (anger, fear, sadness, pleasant surprise, joy, awe). Acoustic features were extracted from the eGeMAPS set. Results showed that awe was produced with longer duration, lower intensity, darker timbre, and stable vowel articulation compared to other emotions. Listener recognition of awe was best predicted by utterance duration, greater pitch variability, and reduced formant variability. This study contributes to the growing understanding of nuanced emotional expressions through prosody.

Index Terms: emotional prosody, awe, acoustic analysis, emotion dimensions

1. Introduction

Vocal emotions are conveyed through prosodic modulations of fundamental frequency (*f*0, perceived as pitch), energy or amplitude (perceived as loudness), temporal organization, and spectral/voice-quality cues [1]. These acoustic features can be conceptualized as probabilistic cues that link speakers' emotional states to listeners' perceptual inferences [2]. Indeed, classic encoding–decoding studies show that emotions such as anger, fear, joy, and sadness differ reliably in their prosodic patterns [1], [3]. In addition, a large-scale perceptual work further indicates that speech prosody can communicate at least a dozen distinct emotion categories, including awe, across two cultures as evidenced by listeners' judgments [4]. In line with this work, in a previous study, we found that awe expressed prosodically by non-actors was recognized above chance level by listeners, although the confusion pattern revealed misclassifications of awe intended expressions as pleasant surprise, joy, or sadness [5]. Despite perceptual evidence of an awe prosodic signature, its acoustic profile remains largely unspecified.

Awe has attracted growing interest in psychology and neuroscience [6], [7]. It is typically described as an emotional state combining feelings of elevation and calm [6], [8]. Experiences of awe are frequently elicited by nature, but also by art, moral beauty, and scientific ideas [6]-[8].

Beyond discrete emotion categories, regularities in acoustic profiles align with bi-dimensional models of emotion [1]-[3], which locate emotions in a space defined by valence (aversiveness/pleasantness) and arousal (activation level) [9]. Dimensional accounts generally place awe in the quadrant of positive valence, i.e., pleasant quality, and relatively low to moderate arousal [6]-[8]. This positioning suggests a possible explanation for previously observed misclassification patterns: the acoustic profile of awe may share properties with other positive emotions (e.g., joy, pleasant surprise) and low-arousal emotions (e.g., sadness) [5].

In this study, the same corpus as in [5] is analyzed acoustically, combining the previously collected perceptual data with newly extracted acoustic features. The present study aims to characterize the prosodic expression of awe in acted productions of a one-word utterance by 31 non-actor adult speakers of Québec French. Specifically, the first aim is (1) to identify the acoustic features and quantify how they are modulated when speakers intend to express awe compared with other discrete emotions varying in valence (positive vs. negative) and arousal (high vs. low). The second aim is (2) to determine which acoustic features predict listeners' recognition of awe in speech.

2. Method

2.1. Participants and procedure

2.1.1. Speakers and task

Participants were 31 native speakers of Québec French (17 women and 14 men), aged 41-76 years, recruited in Québec, Canada. None of them had a significant background in acting or singing.

Each speaker produced six emotions (anger, fear, sadness, pleasant surprise, joy, awe) on command on the same one-word utterance /papa/. For each emotion, two standardized scenarios were presented, resulting in 12 productions per speaker (31 × 12 = 372 tokens). The six emotions comprised three negative (anger, fear, sadness) and three positive (joy, pleasant surprise, awe) emotions, with four high-arousal and two relatively lower-arousal categories (sadness, awe). The 12 short scenarios described typical situations for the elicitation of the target emotion. For awe, the scenarios depicted a first experience of the northern lights and viewing a famous painting, based on prototypical awe-eliciting situations [8].

2.1.2. Listeners and emotion ratings

In our previous study [5], five expert listeners rated all 372 productions, yielding 1,860 trials (372 × 5). On each trial, listeners selected one of six emotion labels (anger, fear, sadness, pleasant surprise, joy, awe) and indicated their confidence on a 7-point scale (1 = "did not recognize any emotion" to 7 = "absolute confidence"). Confidence ratings of 1 were treated as an absence of recognition; these responses were excluded from further analyses.

10.21437/SpeechProsody.2026-126

For each utterance, a composite recognition score was computed for listeners' attributions of awe, irrespective of the intended emotion, by summing confidence ratings (2–7) from the listeners in which the awe label was chosen.

2.1.3. Audio recording, preprocessing, and acoustic analyses

All audio recordings were conducted using a Tascam DR-40 recorder, either in a sound-attenuated booth or in a quiet room at the participant's home. Booth recordings utilized the recorder's built-in stereo microphones, while home sessions employed a Shure SM-10A head-worn microphone. All files were saved in WAV format (44.1 kHz sampling rate, 32-bit depth).

To enable later calibration of audio files, vocal intensity (in dB) was measured at the start of each recording session using a sound-level meter. Throughout the recording, the distance and angle between the speaker's mouth and the microphone were held constant. Audio files were subsequently amplitude-calibrated in Praat [10] to match the monitored intensity levels, yielding samples with vocal intensities comparable to the original productions. Following calibration, utterance boundaries were manually segmented in Praat [10] using waveform visualization, and for each participant, twelve single-utterance audio files were saved for subsequent acoustic analysis.

Subsequently, acoustic features were extracted using the extended eGeMAPS v2 feature set in openSMILE [11]. From each audio file, 88 features were obtained, summarizing statistics or functionals from 40 acoustic descriptors that capture frequency, energy/amplitude, spectral, temporal, and cepstral characteristics relevant to emotional speech.

2.2. Statistical analyses

All statistical analyses were conducted in Jamovi. The significance threshold was set at $p < .05$.

2.2.1. Encoding analyses

Encoding analyses examined how speakers modulated prosody to express intended emotion.

First, to identify candidate acoustic parameters that carried emotion-related variance, each of the 88 eGeMAPS features was entered separately into a linear mixed-effects model (LMM) with emotion as a fixed factor (six intended emotions) and speaker as a random intercept (n=31 speakers). Features that showed a significant effect of emotion were retained for further consideration.

Second, intercorrelations among the retained features were inspected. When two or more features were highly correlated ($|r| > .50$), they were treated as indexing the same underlying acoustic dimension; in those cases, only the most interpretable and theoretically motivated feature was retained. This yielded a reduced set of acoustic features for the encoding analysis, which will be reported in the results (section 3.1.1).

Finally, for each acoustic parameter retained, the LMM results were analyzed. Type III F-tests assessed the main effect of emotion. Estimated marginal means (EMMs) were extracted for each emotion and compared with EMMs for awe expressions to identify significant differences.

2.2.2. Decoding analyses

Decoding analyses examined the acoustic variations predicting listeners' perception of awe.

First, Pearson correlations were computed between the 88 eGeMAPS features and the awe recognition score. Features significantly correlated with awe recognition were retained as candidate predictors. As in the encoding analyses, intercorrelations among these candidates were examined, and when features were highly correlated ($|r| > .50$), a single representative measure was selected to avoid redundancy and multicollinearity. The final set of decoding predictors is reported in the results (Section 3.1.2).

Second, a multiple linear regression model was fit with awe recognition score as the dependent variable and the reduced set of predictors. This model was used to identify which acoustic cues increased or decreased the likelihood that listeners confidently labelled an utterance as awe, irrespective of the emotion intended by the speaker.

Third, a multinomial logistic regression was run on the recognized emotions (awe, joy, surprise, sadness, anger, fear) for a subset of tokens with at least three agreements among the five listeners (n=277). The subset of key predictors identified in the linear model was entered as fixed effects. Awe served as the reference category, allowing us to estimate how variation in these cues shifted listeners' choices away from or toward awe.

3. Results

The results are presented in two parts, corresponding to the study's main objectives. First, we examine how speakers modulated prosodic features when expressing awe, compared to other emotions (encoding analyses). Second, we identify which acoustic features predicted listeners' perception of awe (decoding analyses).

Table 1: *Acoustic features retained for analysis*

Acoustic feature	**Short definition**	**Perceptual correlate**
Mean fundamental frequency (f_0)	Average f_0 over voiced frames	Overall pitch height
f_0 range and coefficient of variation (CoV)	*Range*: Difference between the 80th and 20th percentile of f_0 over voiced frames *CoV*: Standard deviation of f_0 divided by mean f_0 across voiced frames	Pitch variation, intonation range
Equivalent sound level	Mean acoustic energy	Loudness, vocal intensity
Duration	Total time length of the utterance	Speech length
Alpha ratio	Ratio of summed energy from 50–1000 Hz and 1–5 kHz over voiced frames (spectral balance)	Timbre: dark (low–frequency–dominant) vs bright (high–frequency–dominant)
First formant (F1) CoV	Standard deviation of F1 frequency divided by the mean F1 frequency across voiced frames	Articulatory instability (variability in vowel quality)

3.1.1. Encoding awe in speech prosody

The feature-selection process for the emotion-encoding analyses identified 29 acoustic measures that showed significant emotion-related variance. After removing highly

correlated measures, six parameters remained: mean f_0, f_0 range, equivalent sound level, utterance duration, alpha ratio, and F1 coefficient of variation (CoV). Table 1 summarizes these measures and their perceptual correlates.

Figure 1 shows how awe expressions differed from the other emotions on these six acoustic features. Compared to the other emotions, awe was produced with a low pitch, although it remained higher than sadness (mean f_0; Figure 1a), and with a wide pitch range, albeit narrower than pleasant surprise (f_0 range; Figure 1b). In terms of loudness, awe productions were relatively soft, being quieter than joy, surprise, anger, and fear, yet slightly louder than sadness (equivalent sound level; Figure 1c). Awe was produced with longer durations than the other five emotions (Figure 1d). In terms of spectral balance, awe exhibited a lower alpha ratio than anger, fear, surprise, and joy, indicating comparatively reduced high-frequency energy (Figure 1e). Finally, awe showed reduced formant variability (lower F1 CoV) relative to the other emotions, except for joy, which did not differ significantly from awe (Figure 1f).

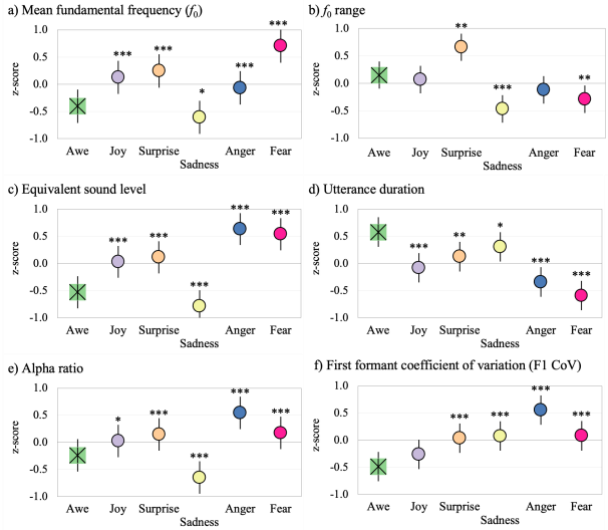

Figure 1: *Prosodic modulation for six acoustic features across emotions (z-scored estimated marginal means and 95% confidence intervals). Asterisks denote significant differences from awe (* p < .05; ** p < .01; *** p < .001).*

3.1.2. *Decoding awe in speech prosody*

In the decoding analyses, 33 of the 88 eGeMAPS features showed significant correlations with awe recognition scores. After removing redundant features with high intercorrelations, 20 acoustic features remained as candidate predictors. A linear regression using these predictors identified three significant predictors of awe recognition: duration, f_0 CoV, and F1 CoV (see Table 1 for feature definitions).

Among these, utterance duration was the strongest predictor ($\beta = 0.76$, $t = 5.65$, $p < .001$): longer utterances were associated with higher awe recognition scores. Pitch variability, measured by f_0 CoV, was also positively associated with awe judgments ($\beta = 0.13$, $t = 2.44$, $p = .015$): greater f_0 variability slightly increased the likelihood of awe recognition. In contrast, F1 CoV (a measure of formant variability; higher values indicate less stable vowel articulation) negatively predicted awe recognition ($\beta = -0.14$, $t = -3.09$, $p = .002$): greater formant variability was associated with lower awe recognition.

To examine how these cues shape listeners' emotion attributions, a multinomial logistic regression was fit on recognized emotion for tokens with at least 3 agreements among the 5 listeners (n=277), using awe as the reference category and duration, f_0 CoV, and F1 CoV as predictors. Omnibus likelihood-ratio tests showed that all three predictors significantly improved model fit (f_0 CoV: $\chi^2(5) = 86.23$, $p < .001$; duration: $\chi^2(5) = 117.45$, $p < .001$; F1 CoV: $\chi^2(5) = 17.50$, $p = .004$). Figure 2 illustrates how the probabilities of each recognized emotion change as a function of the three acoustic predictors.

Across all emotion categories, duration had negative coefficients (ORs = 0.06–0.22), indicating that longer utterances strongly decreased the odds of joy, surprise, sadness, anger, or fear attribution relative to awe. For joy, sadness, anger, and fear, f_0 CoV also had negative coefficients (ORs = 0.22–0.36), such that greater pitch variability favored awe over these emotions; in contrast, f_0 CoV did not reliably distinguish surprise from awe. F1 CoV showed the opposite pattern: positive coefficients for surprise, sadness, anger, and fear (ORs = 1.93–3.37; marginal for joy, $p = .061$) indicate that greater formant variability increased the odds of these labels compared with awe.

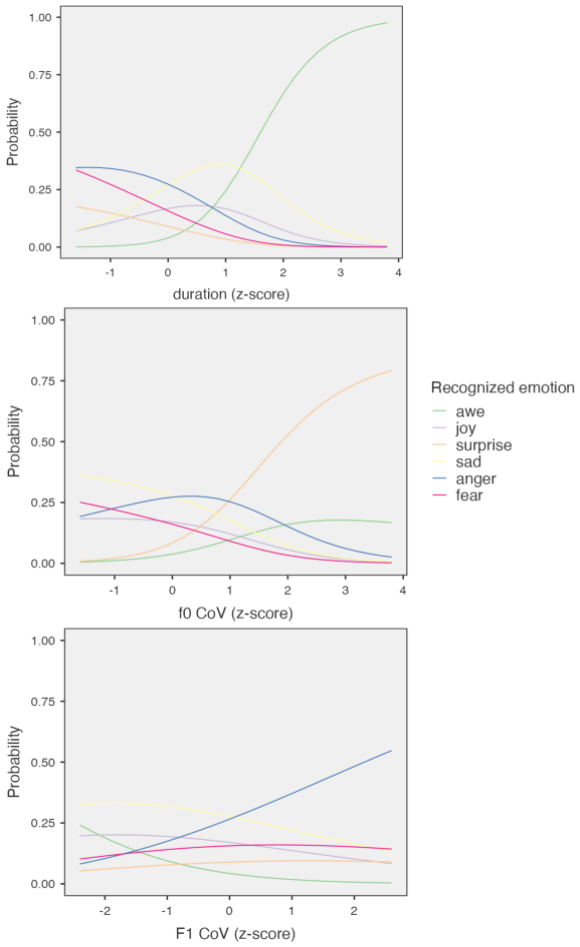

Figure 2: *Probabilities of emotion attribution as a function of duration, fundamental frequency (f_0) coefficient of variation (COV), and first formant (F1) COV.*

4. Discussion

This study investigated how awe is expressed in speech prosody and how listeners recognize it, using acted one-word utterances in six emotional contexts. Two main questions guided the analysis: (1) How do speakers modulate prosodic features to express awe compared to other discrete emotions? (2) Which acoustic features predict listeners' recognition of awe?

The results support the idea that awe has a distinct acoustic signature. Compared to other emotions, awe was produced with longer durations, lower intensity, moderate pitch height and range, darker timbre, and stable vowel articulation. This combination of acoustic features differs from other positive emotions (joy, surprise) and a low-arousal emotion, i.e., sadness, but also shares some similarities. These comparisons between awe and other emotions might contribute to our understanding of where awe fits within the valence-arousal acoustic space. Awe appears to occupy a region of positive valence and low-to-moderate arousal, combining a slow, soft, and stable vocal signature.

From the listener's perspective, three acoustic cues significantly predicted awe recognition: utterance duration, pitch variability (f_0 CoV), and formant stability (F1 CoV). Duration emerged as the strongest predictor, with longer utterances more likely to be labeled as awe. This temporal expansion may reflect the contemplative nature of awe or an emotional pause prompted by beauty, immensity, or transcendence. Pitch variability also contributed positively to awe recognition, while increased formant variability, indicating unstable articulation, reduced it. Together, these cues suggest that awe is perceived when speech sounds are elongated, moderately expressive in pitch, and vocally controlled.

These insights may inform future work in mapping prosodic features to emotions.

4.1.1. Limitations and future work

The study used short, acted utterances of a single word in a controlled setting, which limits generalizability to natural speech and to other linguistic units. However, using standardized materials allowed for prosodic comparisons across emotions, isolating prosodic from segmental differences in acoustic features.

Future research should explore spontaneous expressions of awe in speech, longer utterances that might expand the range of prosodic cues used, and different linguistic or cultural contexts to determine whether the acoustic profile of awe is consistent across contexts and populations.

5. Conclusion

This study provides a preliminary description of the acoustic profile of awe in speech prosody and identifies the cues that guide its recognition by listeners. The findings suggest that awe is prosodically distinct from other discrete emotions, marked by temporal expansion, softness, and articulatory stability.

6. Acknowledgements

We thank Zihao Geng for the help with implementing the OpenSmile tool for acoustic analyses. The first author was supported by a postdoctoral grant from Fonds de recherche du Québec (https://doi.org/10.69777/370132).

7. References

[1] P. Larrouy-Maestri, D. Poeppel, and M. D. Pell, "The Sound of Emotional Prosody: Nearly 3 Decades of Research and Future Directions," *Perspectives on Psychological Science*, Jan. 2024, doi: https://doi.org/10.1177/17456916231217722.

[2] K. R. Scherer, "Vocal affect expression: A review and a model for future research," *Psychological Bulletin*, vol. 99, no. 2, pp. 143–165, 1986, doi: https://doi.org/10.1037/0033-2909.99.2.143.

[3] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614–636, 1996, doi: https://doi.org/10.1037/0022-3514.70.3.614.

[4] A. S. Cowen, P. Laukka, H. A. Elfenbein, R. Liu, and D. Keltner, "The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures," Nature Human Behaviour, vol. 3, no. 4, pp. 369–382, Apr. 2019, doi: https://doi.org/10.1038/s41562-019-0533-6.

[5] V. Coulombe, V. Martel-Sauvageau, and L. Monetta, "The Expression of Vocal Emotions in Cognitively Healthy Adult Speakers: Impact of Emotion Category, Gender, and Age," Journal of Nonverbal Behavior, Sep. 2024, doi: https://doi.org/10.1007/s10919-024-00472-x.

[6] J. C. Chen *et al.*, "'Awe-scillations': EEG spectral and complexity representations of awe," *bioRxiv (Cold Spring Harbor Laboratory)*, Nov. 2025, doi: https://doi.org/10.1101/2025.11.20.689403.

[7] V. E. Sturm *et al.*, "Big smile, small self: Awe walks promote prosocial positive emotions in older adults," *emotion*, vol. 22, no. 5, Sep. 2020, doi: https://doi.org/10.1037/emo0000876.

[8] D. Keltner and J. Haidt, "Approaching awe, a moral, spiritual, and aesthetic emotion," *Cognition and Emotion*, vol. 17, no. 2, pp. 297–314, Jan. 2003, doi: https://doi.org/10.1080/02699930302297.

[9] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: https://doi.org/10.1037/h0077714

[10] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer," (Version 6.4.07) [Computer program], 2024. Retrieved March 17, 2024, from http://www.praat.org/.

[11] F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190–202, Apr. 2016, doi: https://doi.org/10.1109/taffc.2015.2457417.