

The Sound of Inattention: Predicting Mind Wandering with Automatically Derived Features of Instructor Speech

Ian Gliser¹, Caitlin Mills¹, Nigel Bosch², Shelby Smith¹, Daniel Smilek³, Jeffrey D. Wammes⁴

¹ University of New Hampshire, Durham, NH, USA

² University of Illinois Urbana-Champaign, Champaign, IL, USA

³ University of Waterloo, Waterloo, Ontario, CA

⁴ Queen's University, Kingston, Ontario, CA

Abstract. Lecturing in a classroom environment is challenging - instructors are tasked with maintaining students' attention for extended periods of time while they are speaking. Previous work investigating the influence of speech on attention, however, has not yet been extended to instructor speech in live classroom lectures. In the current study, we automatically extracted acoustic features from live lectures to determine their association with rates of classroom mind-wandering (i.e., lack of student attention). Results indicated that five speech features reliably predicted classroom mind-wandering rates (Harmonics-to-Noise Ratio, Formant 1 Mean, Formant 2 Mean, Formant 3 Mean, and Jitter Standard Deviation). These speaker correlates of mind-wandering may be a foundation for developing a system to provide feedback in real-time for lecturers online and in the classroom. Such a system may prove to be highly beneficial in developing real-time tools to retain student attention, as well as informing other applications outside of the classroom.

Keywords: mind-wandering, attention, acoustics, speech, openSMILE

1 Introduction

In the classroom, lecturers are often faced with the challenging task of combatting frequent bouts of student inattention and disengagement. Such inattention often arises in the form of mind-wandering, defined here as thoughts unrelated to the task at hand (e.g., a classroom lecture [31-32]). When a student mind-wanders, they risk missing out on critical pieces of information and thus can develop an impoverished understanding of the learning material. It is therefore important to find ways to reduce the occurrence of mind-wandering and potentially mitigate its negative impact.

One way to minimize the potential negative influence of mind-wandering is by detecting and responding to it in real-time [20]. However, approaches to date have mostly focused on student-centered models of mind-wandering – where ongoing data specific to each learner (e.g., eye-gaze) are necessary to make predictions about

whether or not they are currently mind-wandering. These attempts have been successful in laboratory contexts, but they are not currently scalable to entire classrooms.

Here we adopt an environment-centered model instead, by focusing on subtle naturalistic fluctuations in the learning environment (i.e., the instructor’s speech). We test, for the first time, whether instructor speech patterns are related to classroom mind-wandering – potentially setting the foundation for the development environment-centered models of mind-wandering that can mitigate mind-wandering through scalable automated instructor feedback.

Environment-centered models seem based on the lessons we have already learned from laboratory cognitive psychology studies about when and why mind-wandering occurs. For example, mind-wandering tends to increase over the course of a task [33], and decrease when the task becomes more difficult (but see [13, 29]). Notably, even features like typeface seem to influence how often learners report mind-wandering: participants reported mind-wandering more often when reading a text in grey Comic Sans versus black Arial [11]. Although these studies demonstrate the potential malleability of mind-wandering, it is unclear if subtle environmental features (e.g., instructor behaviors, content changes, speech) may influence mind-wandering in live classrooms.

Here we directly examine how variations in the way information is transmitted through speech relates to students’ attention in classroom contexts. This study builds on the environment-centered approach adopted by Bosch et al. [3], which examined how fluctuations in instructor movements were found to successfully predict classroom mind-wandering rates. Our specific focus on the instructor’s speech fills an important gap in the literature, as very little research to date has been dedicated to quantifying and understanding how acoustical speech patterns influence student attention (e.g., rates of mind-wandering).

1.1 Background Literature

Acoustical features of speech have previously been linked to listener attention and information retention, albeit outside of the educational realm [4, 24]. For example, both the structure of speech (e.g., pitch contour and trajectory of source location) as well as prosodic quality (e.g., pitch and loudness) appear to reliably predict audience inattention [5, 10]. The emotional tone conveyed through acoustical features also seems to be an important aspect of speech; for example, there are clearly dissociable processing patterns in the brain when people hear angry versus neutral prosody [26]. These studies, though not conducted in the context of a lecture, highlight the potential for acoustic-prosodic features to impact information processing – making it likely that mind-wandering may also be influenced.

Only a few studies have attempted to link acoustic features to mind-wandering specifically. Drummond & Litman [6] asked students to read a paragraph about biology aloud and then perform a learning task (either self-explanation or paraphrasing). Periodically, they were probed to report how frequently they experienced off-task thoughts on a scale from 1 (*all the time*) to 7 (*not at all*) during the task. Students’ responses were split into two categories, where 1–3 on the scale was “high” in zoning

out, and 5–7 was “low” in zoning out. They trained a supervised machine learning model on the students’ acoustic-prosodic features to classify low and high zone out, and achieved an accuracy of 64% in discriminating between the two. This study provides some evidence that individuals’ tendencies to mind-wander are related to acoustic-prosodic cues (e.g., percent of silence, pitch, energy) of their own speech. It remains unclear, however, how such acoustic-prosodic features extracted from a *speaker* influence mind-wandering for *listeners*. Establishing associations between these speaker features and listener attention may have direct applications for an environment-centered feedback system in a classroom.

In the current study we sought to bridge the gap in our understanding of possible associations between features of speech and classroom attention. Our goals were (1) to provide a proof-of-concept method for automatically analyzing classroom speech features from low-cost audio recordings, and (2) to elucidate the relationship between acoustical speech features and mind-wandering in the classroom.

To tackle the first goal, we automatically extracted speech features from classroom lecture recordings using an open source software package called open Speech and Music Interpretation by Large-Space Extraction (openSMILE) [9]. We selected a popular feature set provided by openSMILE – the Geneva Minimalistic Acoustic Parameter Set [8]. We then identified and extracted a set of theoretically-relevant acoustic features from nine live classroom lecture recordings.

Next, as a step toward identifying key features to use in an environment-centered model of mind-wandering, we assessed the relation of these features to mind wandering. We focused on mind-wandering because it is consistently reported to be negatively associated with performance and comprehension in complex learning environments [19, 23, 31], including university lectures [34-36]. We aligned speech features in time with students’ self-reported mind-wandering rates in order to probe this relationship. Below we describe our method for processing the audio recordings, how we arrived at a set of theoretically-relevant acoustic-prosodic variables, and how we tested for associations with mind wandering behavior.

2 Method

To address our two research goals, we collected data from multiple sources. As an overview, audio was extracted from low-cost video recordings of lectures, and students’ attention was polled using a computer application during these same lectures. The two data channels were temporally aligned at 500-second intervals for analyses. Each stage is outlined in greater detail below.

2.1 Classroom audio and self-reported mind wandering

We extracted audio from recordings of nine different lectures at the University of Waterloo. These lectures were delivered by three different instructors (three lectures each) who were teaching undergraduate psychology courses. The lectures were delivered during normal classroom meeting times and with no manipulations or interference related to our experiments. The lectures took place in two similar

classrooms, each with sloped, stadium-style seating and a stage with a podium for the teacher. The audio recording began at the same time as the lecture and lasted for the entirety of the class. For more details on data collection, please refer to Wammes et al. and Bosch et al. [3, 36].

Mind-wandering self-reports were collected from the students who participated in the study during the lecture ($N = 76$). Students who agreed to participate in the experiment downloaded an application onto their laptop that administered pseudo-randomly scheduled thought probes throughout the lecture. Specifically, the occurrence of each thought probe notification was individually randomized, with the constraint that probes appeared no more than five times throughout the lecture with a range of 15 and 25 minutes between probes. When a thought probe was scheduled, a small window appeared in the bottom right corner of their computer screen. This prompted participants to introspect about their mental state just prior to the probe, and report their current degree of mind-wandering on a continuous scale ranging from *Completely mind-wandering* to *Completely on task* (reverse scored to correspond numbers between 0 and 1, where higher values refer to more mind-wandering). They were informed that *mind-wandering* was defined as “thinking about unrelated concerns,” and *on task* was defined as “thinking about the lecture.”

2.2 Audio Processing and Feature Extraction

In order to avoid interference from speech unrelated to lecture delivery, we used Audacity software to trim each audio clip to *only* include the instructor’s speech. Trimmed audio was then processed using openSMILE [10]. openSMILE is a flexible, open-source software package and audio toolkit capable of extracting a variety of different sound-based features, tailored for applications ranging from music to speech. The software extrapolates features based upon one’s chosen configuration package and returns information about the occurrences of the selected features [10]. In this experiment, the configuration package was an implementation of the GeMAPS, a set of acoustic parameters based upon recent acoustical speech research [8]. GeMAPS was selected as the configuration for openSMILE due to its minimalistic approach to affect-oriented audio feature extraction. These parameters are Pitch, Jitter, Formant 1, 2, and 3 frequencies (F1, F2, and F3, respectively), F1 bandwidth, Shimmer, Loudness, Harmonics-to-noise-ratio (HNR), Alpha ratio, Spectral slope of 0-500 Hz and 500-1500 Hz, F1, F2, and F3 relative energy, Harmonic difference H1-H2, and Harmonic difference H1-A3 [8].

Various relevant summary statistics of these basic parameters are also output by openSMILE. These include coefficient of variation (standard deviation normalized by the mean; SD) and mean for each parameter. For Loudness and Pitch, the following features were additionally included: 20th percentile, 50th percentile, 80th percentile, the range of 20th to 80th percentile, as well as the mean and SD of the slope rising signal and slope falling signal. Lastly, the mean of Spectral slopes (from 0-500 Hz and 500-1500Hz), the Alpha Ratio, and the Hammarberg Index were included for each recording, resulting in 56 total features for analysis.

Following extraction of audio recordings from lectures and feature extraction, we identified a subset of these GeMAPS features (described in more detail below) for our analysis.

2.3 Selecting Acoustic Features

We identified a set of theoretically-relevant acoustical characteristics based on previous literature. Specifically, we searched for features that have well-established relationships with psychologically-relevant constructs (see Table 1 for a full description of features and corresponding sources). Due to the lack of classroom-based investigations, the majority of literature review focused on studies examining how features of speech relate to attention and emotion, broadly conceived, in laboratory contexts. For example, emotion is considered to be a fundamental aspect of speech, as the delivery of emotional information is tied to inflection of the voice [21]. The following features were identified with the corresponding sources, as described in Table 1.

2.3.1 Formants

Formants are descriptions of the high regions of spectral energy that occur in discrete regions of frequency. F1, 2 and 3 are necessary for synthesis of vowels in speech; additionally, the presence of F3 is required for interpretable speech [37]. *F1 Bandwidth Mean* also describes the degree to which speech is nasally and thus is included here (see Table 1).

2.3.2 Voiced Segment Length, Loudness, Jitter and Shimmer

Loudness of Speech is vital in a lecture due to its important role in conveying information to all members of the audience as well as its relation to confident and precise speech [15, 18].

Voiced Segment Length is defined as the length of discrete units in a stream of speech, which is measured by recording the average periods of uninterrupted speech. Similar to loudness of speech, voiced segment length is a correlate of confidence and precision in speech [15, 18].

Shimmer Mean is described as the occurrences of fluctuations in loudness of speech. Somewhat analogous, *Jitter Standard Deviation* is defined as the fluctuations in pitch. Both shimmer and jitter have been found to be correlates of trembling and nervous speech [30].

2.3.3 Harmonics-to-Noise Ratio and Hammarberg Index

Harmonics-to-Noise Ratio (HNR) is the ratio of harmonic energy: the difference between fundamental formant (F0), first formant (F1), and second formant (F2). Previous research has found HNR to be a correlate of rough, uneven, and bumpy speech.

Table 1. Full list of selected acoustical speech features, description of selected features, rationale for selection and citations of relevant literature

<p>Formant 0 (F0) Mean. This feature characterizes the fundamental frequency of voice, which is critical in driving inflection and linked to prosody [18, 22, 27]</p> <p>Formant 1 (F1) Mean. The first harmonic formant is a determinant of prosodic quality, which drives speech reception [18, 26-27].</p> <p>Formant 2 (F2) Mean. The second harmonic formant uniquely defines sounds of speech and is an acoustic correlate of resonance and clearness of speech [18, 25-27]</p> <p>Formant 3 (F3) Mean. The third harmonic formant is present in vowel sounds and is fundamental to reception of clear speech [1, 18, 26-27]</p> <p>F1 Bandwidth Mean. This is the region of frequency in which amplitudes differ by less than 3 decibels from the center frequency. It is a determinant of nasally/honky qualities of speech [17-18].</p> <p>Loudness Mean. The average maximum volume of speech indicates more careful and precise speech and is correlated to confident speech as well as compliance [15, 18].</p> <p>Loudness Standard Deviation. See above (Loudness Mean).</p> <p>Jitter Standard Deviation. The standard deviation of pitch fluctuations is associated with trembling/tremorous voices, relating it to nervous voice [30].</p> <p>Shimmer Mean. The average fluctuations of speech loudness are also associated with trembling/tremorous voices, relating it to nervous voice [30].</p> <p>Voiced Segment Mean Length. The average length of discrete units in a stream of speech is a correlate of confident and compliant speech, indicative of precise and careful speech [15, 18].</p> <p>Harmonics-to-Noise Ratio. This is the ratio of harmonic energy difference between the fundamental formant (F0), first harmonic (F1) and second harmonic (F2). This is a correlate of rough, uneven, and bumpy speech [7].</p> <p>Hammarberg Index Mean. The difference in spectral energy between peaks in the 0.2 kHz and 2.5 kHz band [16] is a correlate of low percentile sadness and perceived attractiveness of the speaker [15, 18].</p>
--

2.4 Predicting Student Mind-wandering with Acoustic Features

Speech features were processed and extrapolated from the audio recordings in 500 s epochs of time, whereas students' mind-wandering reports were sampled continuously throughout the lecture. To facilitate comparison between these two data channels, speech features were paired with mind-wandering reports within the same 500 sec window. To accomplish this, mind wandering reports were aggregated across participants within each time window from which acoustic features were derived. This resulted in 72 time windows per class, which we used in the analyses below. The average rating of mind-wandering (on a continuous scale between 0-1, where higher values mean more mind-wandering) was .499 ($SD = .287$).

Relationships between speech features and mind-wandering rates were assessed using linear mixed-effects models. We used the *lme4* package in R [2]. All models included a random effect of class to control for within-class variability in baseline mind-wandering. All models regressed mind-wandering on each acoustic feature of interest. We used restricted maximum likelihood estimation (REML) with unstructured covariance to avoid biasing the error variance. Tests of model significance were computed using a type II Wald chi-square test with a two-tailed α of .05 from the *car* package to take a conservative approach based on only estimates from the model [14].

3 Results

Descriptive statistics for each theoretically-relevant feature can be found in Table 2. Below we describe how each of these features related to classroom rates of mind-wandering. Effect sizes (i.e., standardized regression coefficients) can be found in Table 3. We checked for normality of the residuals (i.e. an assumption for linear regressions), and the residuals displayed a normal distribution.

3.1 Formants

Formant 0 Mean ($\beta = .047, p = .494$) and *Formant 1 Bandwidth Mean* ($\beta = .102, p = .156$) were not reliably related to mind-wandering. In contrast, *Formant 1 Mean* ($\beta = -.236, p = .006$), *Formant 2 Mean* ($\beta = -.188, p = .025$) and *Formant 3 Mean* ($\beta = -.239, p = .007$) were all significantly negatively related to classroom of mind-wandering.

3.2 Loudness, Voiced Segment Length, Jitter and Shimmer

Mind-wandering was not significantly related to *Loudness Mean* ($\beta = -.094, p = .128$) or *Loudness Standard Deviation* ($\beta = .024, p = .801$). The same non-significant patterns were observed between mind-wandering and *Voiced Segment Length Mean* ($\beta = .056, p = .482$) and *Shimmer Mean* ($\beta = -.051, p = .367$). However, *Jitter Standard Deviation* was significantly positively related to mind-wandering ($\beta = .097, p = .008$).

Table 2. Mean and standard deviation of speech features and mind wandering reports.

Features	<i>M</i>	<i>SD</i>
Formant 0 Mean	28.3	2.04
Formant 1 Mean	638	55.8
Formant 2 Mean	1683	56.0
Formant 3 Mean	2750	59.0
Formant 1 Bandwidth Mean	1190	34.0
Loudness Mean	.571	.105
Loudness Standard Deviation	.452	.080
Jitter Standard Deviation	1.56	.117
Shimmer Mean	1.40	.027
Voiced Segment Length Mean	.168	.049
Harmonics-to-Noise Ratio	2.01	1.74
Hammarberg Index Mean	.359	.892
# of Mind Wandering in Epoch	11.8	12.6

3.3 Harmonics-to-Noise Ratio and Hammarberg Index

Harmonics-to-noise Ratio was significantly positively related to mind-wandering ($\beta = .175, p = .009$), whereas *Hammarberg Index Mean* was not ($\beta = .071, p = .224$).

3.4 Summary of Results

We show that subtle fluctuations in speech characteristics influence classroom mind-wandering. Findings indicate that higher speech interpretability (higher values of *Formant 1*, *Formant 2*, and *Formant 3 Mean*), stability of pitch inflection (lower *Jitter Standard Deviation*, and the smoothness and evenness of speech (lower *Harmonics-to-noise Ratio*) were associated with lower rates of self-reported mind-wandering. This same pattern of results was unchanged when analyses were repeated with the number of mind-wandering reports as a control variable.

Table 3. Linear mixed effects models of relevant acoustical speech features. Standardized regression coefficients and p -values listed for each model with significant values ($p < .05$) bolded.

Features	β	p
Formant 0 Mean	.047	.494
Formant 1 Mean	-.236	.006
Formant 2 Mean	-.188	.025
Formant 3 Mean	-.239	.007
Formant 1 Bandwidth Mean	.102	.156
Loudness Mean	-.094	.128
Loudness Standard Deviation	.024	.801
Jitter Standard Deviation	.097	.008
Shimmer Mean	-.051	.367
Voiced Segment Length Mean	.056	.482
Harmonics-to-Noise Ratio	.175	.009
Hammarberg Index Mean	.071	.224

4 Discussion

To date, little research has been devoted to environment-centered models of mind-wandering in classrooms. However, understanding how subtle variations in classroom lectures relate to student attention is important given that students frequently report mind-wandering while listening to lectures. We addressed this gap by assessing the relationship between acoustical speech features and mind-wandering in a classroom setting. We first detailed a method for automatically extracting a set of psychologically-relevant acoustical speech features from low-cost video recordings in live classrooms. We then related these features to students' self-reported mind-wandering across nine lectures. The data indicate that acoustical characteristics of the instructor's speech matter: we observed significant relationships between mind-wandering and *Formant 1*, *Formant 2*, and *Formant 3 Means* as well as *Jitter Standard Deviation* and *Harmonics-to-noise Ratio*. Of these features, *Jitter Standard Deviation* and *Harmonics-to-noise Ratio* were positively related to student mind-wandering, whereas *Formants 1, 2, and 3 Means* were negatively related to mind-wandering.

The negative relationships seen for *Formant 1*, *Formant 2*, and *Formant 3 Means* suggest that students paid more attention when speech was clearer. These three formants are associated with better speech clarity and negative correlates of raspy or hard-to-hear speech [18, 26-27]. From a cognitive standpoint, mind-wandering may be more likely to occur when speech is less clear because discerning the content of the

speech becomes more difficult; students may not have the cognitive resources available to match these increased task demands [13, 38].

The positive relationship observed for *Jitter Standard Deviation* provide some insight into the role of pitch changes in mind-wandering; that is, as the volatility of the instructor's pitch increased, students reported higher rates of mind-wandering. Prior work revealed an association between jitter and trembling or nervous speech [30]. Thus, mind-wandering occurrences may increase when nervousness becomes detectable in an instructor speech. This relationship may be due to how students interpret of the prosody speech; additional sub textual emotional information that is actually irrelevant to the lecture may influence their attention. A similar positive relationship was found for *Harmonics-to-Noise Ratio Mean*: mind wandering increased along with increased magnitude of *Harmonics-to-Noise Ratio* (i.e., rough, uneven speech [15, 18]). In the context of our findings, this suggests that mind-wandering is more likely to occur when the instructor's speech is more rough and uneven. Taken together, these finding highlight promising avenues for improving classroom lectures given that overall clarity of speech may be a simple way to increase student attention.

Our findings serve as a potential basis for an environment-centered system to address mind-wandering that does not require any intrusive or high-cost student measurements. Current developments for such a system are in their infancy and rarely explore raw audio data. For example, Schneider, Borner, Van Rosmalen & Specht [28] developed a real-time feedback system which analyzes nonverbal and verbal behaviors and provides feedback to speakers. This system, while found to increase performance and confidence in speaking, focused on more basic features than the ones used here, such as arm-crossing and volume. This system seeks to ensure that fluid speech is maintained and does not currently incorporate potential predictors of listener attention such as mind-wandering. Our findings suggest that interventions may be effective by targeting acoustical characteristics of instructor speech to improve lecture delivery and reduce classroom mind-wandering.

Our results, when combined with other environment-centered features like movement [3], may provide a crucial step toward the development of a multi-modal feedback system where acoustic properties are considered along with motion, content, and other sources of classroom data. Future research in this area may benefit by repeating the experiment while controlling for lecture content, as content of the presented information is likely to influence off-task though. Next steps include integrating and testing automatically extracted acoustic features into a real-time system. Such a system would require the intake of data in a specified time window, buffering the data in order to allow time for openSMILE computation, then feeding the buffered data to openSMILE for computation and prediction of student attention. Finally, the system will need to return a report that is easily accessible and interpretable by the instructor. While these steps are substantive, they are all attainable and the efficacy of using both video and auditory features has been established. As these tools develop, they provide a promising direction for online real-time intervention in educational settings.

5 References

1. Barrichelo-Lindström V, Behlau M (2009) Resonant Voice in Acting Students: Perceptual and Acoustic Correlates of the Trained Y-Buzz by Lessac. *J Voice* 23:603–609. <https://doi.org/10.1016/j.jvoice.2007.12.001>
2. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67:201–210. <https://doi.org/10.18637/jss.v067.i01>
3. Bosch, N., Mills, C., Wammes, J. D., & Smilek, D. (2018). Quantifying classroom instructor dynamics with computer vision. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018)*. Berlin Heidelberg: Springer.
4. Darwin CJ, Hukin RW (2000) Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *J Acoust Soc Am* 107:970–977. <https://doi.org/10.1121/1.428278>
5. Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: Functional roles and interpretations. *Front. Hum. Neurosci.* 8. <http://doi.org/10.3389/fnhum.2014.00311>
6. Drummond J, Litman D (2010) In the zone: Towards detecting student zoning out using supervised machine learning. *International Conference on Intelligent Tutoring Systems 2010*: 306–308. https://doi.org/10.1007/978-3-642-13437-1_53
7. Eskenazi L, Childers DG, Hicks DM (1990) Acoustic correlates of vocal quality. *J Speech Hear Res* 33:298–306. <https://doi.org/10.1044/jshr.3302.298>
8. Eyben F, Scherer KR, Schuller BW, et al (2016a) The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Trans Affect Comput* 7:190–202. <https://doi.org/10.1109/TAFFC.2015.2457417>
9. Eyben F, Weninger F, Groß F, Schuller B (2013) Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor Categories and Subject Descriptors. *Proc 21st ACM Int Conf Multimed* 835–838. <https://doi.org/10.1145/2502081.2502224>
10. Eyben F, Wöllmer M, Schuller BB, et al (2015) OPENSOURCE: open-Source Media Interpretation by Large feature-space Extraction. *MM'10 - Proc ACM Multimed 2010 Int Conf* 1–65. <https://doi.org/10.1145/1873951.1874246>
11. Faber M, Mills C, Kopp K, D’Mello S (2017) The effect of disfluency on mind wandering during text comprehension. *Psychon Bull Rev* 24:914–919. <https://doi.org/10.3758/s13423-016-1153-z>
12. Faber M, Radvansky GA, D’Mello SK (2018) Driven to distraction: A lack of change gives rise to mind wandering. *Cognition* 173:133–137. <https://doi.org/10.1016/j.cognition.2018.01.007>
13. Feng S, D’Mello S, Graesser AC (2013) Mind wandering while reading easy and difficult texts. *Psychon Bull Rev* 20:586–592. <https://doi.org/10.3758/s13423-012-0367-y>
14. Fox J, Friendly M, Weisberg S (2013) Hypothesis Tests for Multivariate Linear Models Using the car Package. *R Journal*, Vol. 5, No. 1, p. 39-52.
15. Fernandez Gallardo, L. & B. Weiss (2017). “Perceived Interpersonal Speaker Attributes and their Acoustic Features”. *Proc. Phonetic & Phonologie*, pp. 61–64.
16. Hammarberg B, Fritzell B, Gaufin J, et al (1980) Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol* 90:441–451. <https://doi.org/10.3109/00016488009131746>
17. Lee ASY, Ciocca V, Whitehill TL (2003) Acoustic correlates of hypernasality. *Clin Linguist Phonetics* 17:259–264. <https://doi.org/10.1080/0269920031000080091>
18. Memon SA (2018) Acoustic Correlates of the Voice Qualifiers Summarizing the Perception of Qatar on Twitter View project Algorithms for Speaker Profiling <https://doi.org/10.13140/RG.2.2.24759.57764>

19. Mills C, Graesser A, Risko EF, D’Mello SK (2017) Cognitive coupling during reading. *J Exp Psychol Gen* 146:872–883. <https://doi.org/10.1037/xge0000309>
20. Mills C, Gregg J, Bixler R, D’Mello SK (2020) Eye-Mind reader: an intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human-Computer Interaction* 1–27. <https://doi.org/10.1080/07370024.2020.1716762>
21. Murray IR, Arnott JL (1993) Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J Acoust Soc Am* 93:1097–1108. <https://doi.org/10.1121/1.405558>
22. Paeschke A, Kienast M, Sendlmeier WF, Berlin TU (1999) F0contours in emotional speech. *Proc. 14th Int. Congr. Phonetic Sci.*
23. Randall JG, Oswald FL, Beier ME (2014) Mind-wandering, cognition, and performance: a theory-driven meta-analysis of attention regulation. *Psychol Bull* 140:1411–1431. <https://doi.org/10.1037/a0037428>
24. Rimmele JM, Zion Golumbic E, Schröger E, Poeppel D (2015) The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68:144–154. <https://doi.org/10.1016/j.cortex.2014.12.014>
25. Robb M, Blomgren M (1997) Analysis of F2 transitions in the speech of stutterers and nonstutterers. *J Fluency Disord* 22:1–16. [https://doi.org/10.1016/S0094-730X\(96\)00016-2](https://doi.org/10.1016/S0094-730X(96)00016-2)
26. Sander D, Grandjean D, Pourtois G, et al (2005) Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage* 28:848–858. <https://doi.org/10.1016/j.neuroimage.2005.06.023>
27. Scherer KR, London H, Wolf JJ (1973) The voice of confidence: Paralinguistic cues and audience evaluation. *J Res Pers* 7:31–44. [https://doi.org/10.1016/0092-6566\(73\)90030-5](https://doi.org/10.1016/0092-6566(73)90030-5)
28. Schneider J, Borner D, Van Rosmalen P, Specht M (2016) Can You Help Me with My Pitch? Studying a Tool for Real-Time Automated Feedback. *IEEE Trans Learn Technol* 9:318–327. <https://doi.org/10.1109/TLT.2016.2627043>
29. Seli P, Konishi M, Risko EF, Smilek D (2018) The role of task difficulty in theoretical accounts of mind wandering. *Conscious Cogn* 65:255–262. <https://doi.org/10.1016/j.concog.2018.08.005>
30. Shao J, MacCallum JK, Zhang Y, et al (2010) Acoustic analysis of the tremulous voice: Assessing the utility of the correlation dimension and perturbation parameters. *J Commun Disord* 43:35–44. <https://doi.org/10.1016/j.jcomdis.2009.09.001>
31. Smallwood J, Fishman DJ, Schooler JW (2007) Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychon Bull Rev* pp 230–236 <https://doi.org/10.3758/bf03194057>
32. Smallwood J, Schooler JW (2006) The restless mind. *Psychol Bull* 132:946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
33. Thomson DR, Seli P, Besner D, Smilek D (2014) On the link between mind wandering and task performance over time. *Conscious Cogn* 27:14–26. <https://doi.org/10.1016/j.concog.2014.04.001>
34. Wammes JD, Seli P, Cheyne JA, et al (2016) Mind wandering during lectures II: Relation to academic performance. *Scholarsh Teach Learn Psychol* 2:33–48. <https://doi.org/10.1037/stl0000055>
35. Wammes JD, Smilek D (2017) Examining the Influence of Lecture Format on Degree of Mind Wandering. *J Appl Res Mem Cogn* 6:174–184. <https://doi.org/10.1016/j.jarmac.2017.01.015>

36. Wammes JD, Ralph BCW, Mills C, et al (2019) Disengagement during lectures: Media multitasking and mind wandering in university classrooms. *Comput Educ.* <https://doi.org/10.1016/j.compedu.2018.12.007>
37. Weiss B, Burkhardt F (2010) Voice attributes affecting likability perception. *Proc 11th Annu Conf Int Speech Commun Assoc INTERSPEECH 2010*
38. Xu J, Metcalfe J (2016) Studying in the region of proximal learning reduces mind wandering. *Mem Cognit* 44:681–695. <https://doi.org/10.3758/s13421-016-0589-8>