Special Issue – Building Acoustics and Health

BUILDING ACOUSTICS

Building Acoustics
1–23

DOI: 10.1177/1351010X19896346
journals.sagepub.com/home/bua
SAGE

Effect of mode of delivery and background noise on speech characteristics of talkers in a classroom environment

Nithya Subramaniam and K Ramamurthy

Abstract

Global and acoustic–phonetic correlates of speech intelligibility are important measures of talker intelligibility. Speech characteristics have been evaluated using speech samples that involve reading pre-administered passages under controlled environments. In classrooms, lecturing is the mode of speech delivery. The presence of noise from ceiling fans and other external noise hinder the speech communication process. This study evaluated the talkers' speech characteristics utilizing recordings from graduate students under reading and lecturing modes and in the presence and absence of noise generated by ceiling fans. The acoustical conditions under which talkers delivered their speech were characterized using octave band U_{50} values. Global and acoustic–phonetic correlates of talker intelligibility were measured and the variation in correlates of speech intelligibility was statistically analyzed. The results revealed that talkers significantly modified their speech characteristics relevant to intelligibility across modes of speech delivery and in the presence of noise. Fundamental frequency measures such as F0-mean and F0-SD and durational measures such as speech and pause rates were all found to be higher for lecture mode of delivery. Talkers showed similar vowel-articulatory changes under the two modes of delivery. When lecturing in the presence of noise, talkers significantly reduced the length of pauses and also utilized a combination of vowel-articulatory strategies to overcome the presence of noise. The results suggest the need to investigate talkers' speech adaptation in real classroom environments in terms of correlates of speech intelligibility and to reconsider classroom acoustical guidelines in view of both listener and talker intelligibilities.

Keywords

Vowel space area, useful-to-detrimental ratio, speech intelligibility

Introduction

The role of speech and articulation in education is of paramount importance. Classroom acoustics research has evolved over the decades and has mainly focused on room and listener effects on

Building Technology and Construction Management Division, Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India

Corresponding author:
K Ramamurthy, Building Technology and Construction Management Division, Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600036, India.
Email: vivek@iitm.ac.in

intelligibility of speech. Majority of the studies have considered the speaker to be a fixed entity delivering speech at comfortable signal-to-noise ratios. Very few studies have tried to explore the changes in correlates of speech intelligibility in classrooms by studying the speakers' acoustic–phonetic and prosodic characteristics. For a given environment, intelligibility of speech varies from talker to talker.[1,2] If talkers adopt speaking styles that benefit listener perceptually, there could be radical changes in the teaching–learning experience in classrooms. Talkers elicit different speaking styles or modes consciously or unconsciously, depending on the communication circumstances, nature of the environment, listener characteristics, and their inherent speech characteristics. In evaluating speech intelligibility, the talker and the environment are the main influential factors.[3] It has been proposed in the literature that talkers vary their articulation along a hypo-to-hyper continuum in response to the communication circumstances[4] and also in response to degraded ambient conditions.[5] In classroom environments, during presentations/talks, talkers typically adopt the lecturing mode of speech delivery, which is free flowing and natural. The term "speaking mode" or "mode of delivery" has been used against speech task to differentiate the meaning in terms of the way speech is delivered: mode referring to natural free-flowing speech delivery as against task referring to reading a prescribed passage as instructed. Studies have found different influences of room acoustics, and type of discourse on speech levels of reading task compared to map-description task[6] and on fundamental frequency characteristics of speech.[7] Differences in vocal effort are seen between conversational task and when speaking in a classroom environment.[8] While it is widely acknowledged that talkers greatly vary in their intelligibility through speech modifications in terms of their speech sound pressure levels, pitch and durational measures that happen due to changes in speech production and delivery,[9,10] reading material,[11] environment,[8] and reverberation,[12] these studies have either utilized speech samples extracted during reading standard pre-administered texts or conversational tasks and when speaking in controlled environments. There have been no studies on how talkers articulate their speech in real classroom environments using samples from natural free-flowing lectures. First, a comparison of the variation between reading standard texts and natural lecturing is also brought out through this study, in terms of speech intelligibility correlates. Second, since communication happens in a closed environment, it is important to study the impact of the listening environment on speech. In spaces like schools and university classrooms, this becomes more of a concern when factors such as reverberation and noise not only affect speech intelligibility on their own,[13,14] but may also interact with individuals' vocal tracts to further influence intelligibility.[15] Degradation of speech by reverberation is primarily caused by the smearing of sounds in the temporal gaps in speech from excessive reflections.[16,17] Background noise (BGN) can influence the signal because noise causes the talker to modify his or her own speech pattern (the Lombard effect).[18] Talkers in the presence of noise have been reported to involuntarily compensate for BGN by changing various aspects of their speech[19] such as by increasing their fundamental frequency (F0) characteristics, sound intensity, and vowel duration. The standard deviation of the fundamental frequency (SD-F0) and other durational measures of voicing have also been identified as symptomatic indicators of vocal comfort in rooms.[20]

From the room acoustics perspective, useful-to-detrimental ratio (U_{50}) gives the relationship between beneficial sound energy from the talker and the detrimental sound energy including late-arriving energy and noise. In a study by Bradley,[21] U_{50} was found to be the most accurate room acoustic predictor of speech intelligibility among the three types of combined room acoustic predictors of intelligibility such as U_{50}, speech transmission index (STI), and articulation loss of consonants and explained 97% of the variance in speech intelligibility scores. A minimum 1 kHz U_{50} of +1 dB is recommended for satisfactory intelligibility in classrooms.[22] However, U_{50} in terms of frequency-weighted sums is reported to be ideal as this single number value is inclusive of the effects of useful and detrimental energy in all frequency bands of interest.[23]

Review

Studies relating speech intelligibility scores to speech characteristics of talkers have found different talker correlates of intelligibility in terms of (1) global and (2) acoustic–phonetic characteristics.[2,3,24,25] A brief review on the global characteristics such as fundamental frequency measures and durational measures such as pause length, pause rate, and speaking rate along with the acoustic–phonetic characteristics including vowel properties and vowel space measures from relevant studies are presented in the following.

Global speech characteristics

Fundamental frequency (F0). In studies on teacher speech characteristics from primary school environments, F0-mean was found to be higher during teaching when compared to day-to-day conversational speech.[26–29] In another room acoustical study, F0-mean was found to increase with distance at a rate of 3.8 Hz per double distance to the listener and is 4 Hz higher in anechoic conditions.[20] Although its direct influence on speech intelligibility is not clear,[2] with some studies reporting no changes in F0-mean between higher intelligibility and lower intelligibility speakers,[1] an increase in intelligibility was associated with an increase in F0-range.[10,24,30,31] The standard deviation of the fundamental frequency (F0-SD) also is a better measure to understand the variation present in speech over longer periods of time.[32] High levels of pitch variation, characterized by F0-SD, correlate with perceptions of speaker liveliness,[7,33] charisma,[34,35] and an increase in listeners' performance.[36] Speech that is delivered without pitch variation affects a listeners' ability to recall information and is not favored by listeners.[37] The speaker also risks conveying an impression of disengagement with the topic and the audience.[38] In short, intelligibility and clarity are associated with higher F0-mean, F0-SD, and F0-range, but the exact relationship between them remains largely unclear.

Durational measures: pause characteristics and speech rate. Pauses provide considerable help to listeners because they give them a chance to process what has been said before the talker says something more. This is especially true in noisy situations. An increase in the pause count and mean duration of pauses are typical of clear/slow speech and serve as acoustic cues to enhance intelligibility.[10,30,31] Influence of pause occurrence and speech intelligibility is not clear as artificially inserting pauses in conversational speech did not yield improved intelligibility scores,[39] while naturally occurring pauses in slow speech samples indicate a correlation between pause count and speech intelligibility.[10,30]

In the case of speech rate, slower rates of speech in general have been found to positively influence speech perception.[1,2,40–42] This positive effect of speaking rate on intelligibility is attributed to greater articulatory precision in slower speech rather than to the effect of segmental duration.[24] Although slow speech with increased pauses is not a substitute to the deficiencies in the listening environment or other speaker intrinsic or speaker-induced intelligibility deficits, it nevertheless enhances communication experience by allowing listeners extra time to process information. Speech rate is one of the critical factors contributing to vowel reduction by reducing vowel durations and affects the temporal and spectral characteristics of sounds.[42,43]

In room acoustical conditions with noise, an increase in the speech rate could be further detrimental to intelligibility as increases in speech rate require a corresponding increase in signal-to-noise ratio required to achieve 50% accuracy on speech intelligibility scores.[42] However, extreme slowing down of speaking rate is also not beneficial in the presence of BGN.[44]

Talkers vary in their strategies used to communicate more efficiently and successfully.[3,31] Inter-talker intelligibility studies have shown mixed results with no clear consensus on the influence of speech rate on intelligibility, with some studies indicating correlation between speech rate and speech intelligibility[2,40] while some others did not find any correlation between the two.[1,24]

Acoustic–phonetic characteristics

Vowel properties. Research supports the belief that intelligibility is linked to the properties of the first and second vowel formant frequencies F1–F2.[1,24] The formant frequencies F1 and F2 largely determine the acoustical quality of vowel sounds. F1 and F2 vary with speech styles with a general tendency for F1 to increase rather than F2.[19,45,46] However, in the case of speech produced in the presence of noise, both F1 and F2 differ. An increase in the mean vowel length has also been observed in clear speech in several studies[10,46–49] and is an important aspect of talker intelligibility.

Vowel working space. The $F1 \times F2$ vowel space is a robust predictor of speech intelligibility. This is of interest because the expansion and contraction of a talker's vowel space is indicative of within talker variability related to adaptive speech for improving intelligibility.[50] Vowel analysis from segmental study of speech is useful for predicting the speech intelligibility of talkers. Expanded vowel working areas have been reported in the literature for clear speech.[9,48,51–53] Although it is still unclear as to which properties of the vowel space correlate the best with intelligibility, overall size of the vowel space is found to be a reliable predictor of intelligibility.[47] Shrinkage of vowel spaces and vowel reduction[54] can perceptibly reduce intelligibility especially in degraded listening conditions.

Previous research studies have concentrated on training talkers to speak clearly, identifying factors influencing talker clarity, establishing correlates of talker intelligibility, and so on. Moreover, almost all the studies have used speech materials to elicit speech in a controlled manner to study properties of speech that influence talker intelligibility. A review of literature suggests that most speech tasks used in experiments only capture a part of the changes in speech mechanism relevant to speech intelligibility such as speech sound pressure level[11] and other speech parameters due to the effect of reverberation and noise[12,55] or have been done in simulated environments.[56] With respect to classroom environments, it becomes important to study any differences observed in global speech characteristics and vowel articulation characteristics of talkers from real classroom environments, when exposed to different room acoustical conditions and when using the natural mode of speech delivery, that is, lecturing. In this study, natural modifications in speech in terms of correlates of speech intelligibility were studied within and across talkers, when eliciting a typical lecture mode of speech as against reading a passage and also under typical classroom noise, reflecting the circumstances associated with communicative demands typical of classroom discourse.

The focus of this article is primarily to evaluate the influence of (1) mode of speech delivery and (2) presence of typical BGN, on correlates of talker intelligibility. The scope of the study is limited to one classroom under unoccupied conditions and a set of six talkers.

The procedure used to characterize the classroom using the U_{50} metric, the details of the talkers used in the study and their delivery modes, and the method of determination of global and acoustic–phonetic correlates of talker intelligibility is briefly explained in the next section.

Methodology

Room acoustic measures

The dimensions of the room were $9.1\,m \times 9.2\,m \times 3.6\,m$. The acoustical quality of the classroom used by talkers for presentations was characterized using the octave band U_{50} values. 1/1 octave

Table 1. Measured spectrum of balloon pops.

Frequency (Hz)	Magnitude (dB)
16	74
31.5	74.9
63	78.3
125	88.4
250	91.9
500	93.5
1000	92.7
2000	98
4000	94.3
8000	92.4
16,000	87.6

band measurements of reverberation times and BGN levels were carried out to derive the U_{50} metric. Reverberation times were measured using balloon pop impulse source method. The balloons used for generating the impulse sound were of above-average size and inflated to a level (40 cm in diameter with an error of ±1 cm) and were considered sufficient to produce considerable excitation across the frequency bands. The measured spectrum of the balloon pops is indicated in Table 1.

The octave band frequencies in the range of 125–8000 Hz were excited and the reverberation time (T_{20}) was measured using the integrating sound level meter Norsonic Type 118. Measurements were recorded at five typical student locations from the front, middle, and rear of the classroom. The 1 kHz reverberation time was 2 s under unoccupied conditions. The fenestration was kept open and the mechanical ventilation systems (ceiling fans) were turned off when evaluating differences between reading and lecture modes and turned on later to evaluate the influence of noise on speech under lecture mode. The spectrum of background ground with and without the operation of ceiling fans is indicated in Figure 1.

From the measured impulse responses in the classroom, octave band useful-to-detrimental ratios (U_{50}) were derived. The following equation (equation (1)) by Bradley[21] was used to arrive at octave band U_{50} values at typical receiver locations in the classroom.

Frequency weightings as per the method used by Choi[23] were used to derive the octave band U_{50} values by averaging the octave band values from 125 to 4000 Hz, considering that this range of octave band frequencies is home to all the vowel sounds and the mean U_{50} was evaluated as a frequency-weighted sum. The octave band U_{50} value evaluated at the center of the classroom was considered. U_{50} is given by equation (1) as follows

$$U_{50} = \left[\frac{10\log\left(\frac{1+r_h^2}{r^2}\right) - e^{\left(\frac{-0.69}{RT}\right)}}{e^{\left(\frac{-0.69}{RT}\right)} + 10^{\frac{SNR}{10}}} \right] \quad (1)$$

where "r_h" is the reverberation distance where the energy densities of the reflected and direct sounds become equal; "r" is the source–receiver distance; "RT" is the reverberation time; SNR is the signal-to-noise ratio at "r." Table 2 indicates the octave band U_{50} (125–8000 Hz) values for the classroom under quiet and noisy conditions.

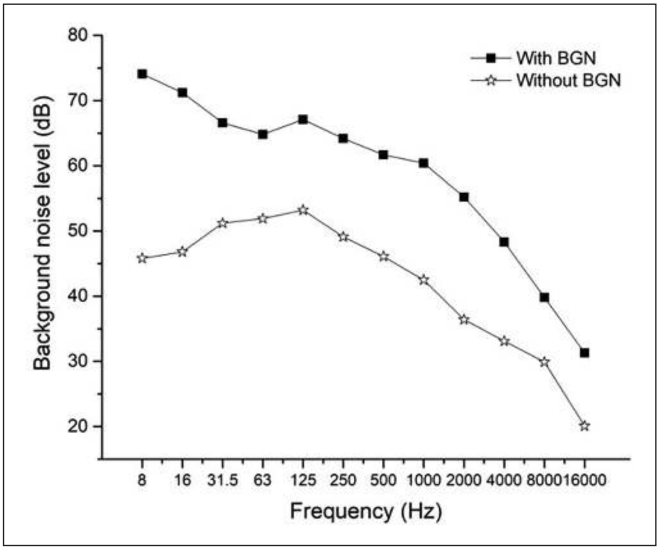

Figure 1. Noise spectrum with and without operation of ceiling fans.

Table 2. Octave band U_{50} values in classroom with and without background noise.

Frequency (Hz)	U_{50} (dB)	
	With background noise	Without background noise
125	–6.0	–2.6
250	–5.9	–1.9
500	–5.8	–1.6
1000	–5.7	–1.05
2000	–5.7	–0.67
4000	–5.7	–0.31
8000	–5.6	0.82
Frequency-weighted sum (250–4000) (AI method)[23]	–5.71	–0.86

AI: articulation index.

The average U_{50} values calculated as a frequency-weighted sum[23] change by an order of nearly 2 dB across the two conditions. The influence of the presence of noise apart from reverberation is seen across all frequency bands, with the 1-kHz U_{50} value ranging between −1.05 and −5.7 dB for quiet to noisy conditions, respectively.

Talkers, modes of speech delivery, and measures of talkers' speech characteristics

Six talkers (T1–T6) were randomly selected from a group of graduate students, with equal number of males and females in the age group of 22–30 years. The talker group has experience in giving seminars in class and in various forums and can be considered as early career teachers in higher educational institutions. The talkers assumed the typical lecturing position on the dais in the classrooms and delivered their talks. A headworn microphone was used, for a robust signal with less feedback issues and the microphone was placed at 5 cm from their mouth,[57,58] and their speech was recorded for the entire duration of their talk. The microphone output was connected directly to the audio card of the PC running on Windows 10. Recording was done on a single channel at a sampling rate of 16 bits/44.1 kHz using the Praat software interface.[59] Audience was not present during the studies and the experimenter alone was seated in front of the speaker and is a limitation of the study. However, the talkers were instructed to speak naturally as they would to a full audience in two modes of delivery: reading mode and lecture mode. The standardized speech material utilized for reading mode was the phonetically balanced Fairbanks'[60] *Rainbow passage*. The reading lasted for 120 s approximately. For the lecture mode, talkers were encouraged to give a mini-lecture for duration of approximately 30 min on their research area. The topic was of their choice and was done to induce confidence and to ensure fluency and flow of speech. They were instructed to talk naturally and were not restricted in any way. It was made sure that no talker delivered both modes on the same day. Reading and lecturing modes were not carried out for any talker on the same day so as not to influence their speech styles and to avoid speaker fatigue and habituation to the recording environment. Participants were blinded to the experimental conditions as the authors did not want the talkers to know that their modes of delivery were studied and compared. To avoid speculation and misconstrued vocal effort from talkers, the two modes were recorded on separate days. Although the recording durations were different for the two modes due to the nature of speech delivery involved, speech samples of 120 s duration were considered for analysis in both cases. Measures of fundamental frequency and duration were extracted and averaged across 30 s slices and the values were used to statistically analyze the data. For vowel space area (VSA) measures, seven tokens per vowel were used. A two-factor analysis of variance (ANOVA) was conducted on the influence of two composite variables (mode of delivery and status of noise) as the primary factor and the influence of speakers as the secondary factor on the selected speech parameters. Mode of delivery was either reading or lecturing and the status of noise was either present or absent. Six talkers were used for measuring each parameter and each speaker was replicated four times. The procedure for evaluating talkers' reading and lecturing modes of delivery in terms of some of the global and acoustic–phonetic measures of talker intelligibility are elucidated in the following.

Fundamental frequency measures: F0-mean, F0-range, F0-SD. Fundamental frequency F0 was extracted with the help of Praat software[59] from the recordings at every 10 ms interval from voiced portions. An auto-correlation-based method was used with the Hanning windows with a length of 0.043 s, a pitch floor of 50 Hz, and a pitch ceiling of 300 Hz. F0-mean and F0-SD were computed for every 30 s interval of the lecture recordings and read passages, as well as for the whole duration of phonation. However, four 30-s samples from the middle portion of the lecture were used for analysis.

Durational measures: pause rate, pause length, and speech rate. Speech rate (syll/s), pause rate (pause/s), and mean pause length (ms) were evaluated for all talkers in both reading and lecture modes. Following the method elucidated in the literature,[30,31] pauses were fixed as silent periods of minimum 10 ms in duration. Pause rate was evaluated by dividing the pause frequency over the duration of total phonation. Mean pause rate was evaluated by dividing the total pause length over the pause frequency.

Vowel measures: VSA. Four-point vowel spaces for all the six talkers in the study were evaluated for all the six talkers. The waveforms for words containing the corner vowels /i/, /u/, /ɑ/, and /a/ that make up the quadrilateral were segmented and extracted from running speech samples recorded during both modes of delivery (reading and lecturing), using Praat software. Using Praat, the vowel segments were selected and saved for further analysis. Praat provides the parallel display of the

Table 3. F0-mean and F0-SD for talkers under reading and lecture mode.

Talker	Reading mode without BGN		Lecture mode without BGN		Difference between modes in ST (F0-mean)	Difference between modes in ST (F0-SD)
	F0-mean (Hz)	F0-SD (Hz)	F0-mean (Hz)	F0-SD (Hz)		
1	279	50	297	59	1.1	2.8
2	180	37	232	44	4.4	3.0
3	144	38	141	38	–0.4	0
4	261	46	279	56	1.2	3.4
5	278	54	281	56	0.2	0.6
6	153	49	149	29	–0.5	–9.1

SD: standard deviation; BGN: background noise; ST: semitones.

speech waveform, spectrogram, and the short spectral cross-section. Formant frequencies F1 and F2 were measured manually, by placing the mouse cursor over the steady-state portion of the vowel nuclei and values were recorded. Seven tokens for each vowel case were measured separately from reading and lecturing samples, for each of the six talkers and this worked out to 168 manual measurements totally. The mean F1 or F2 value of the seven tokens per vowel is taken as the formant pair in Hertz for the particular vowel. The formant frequency measures in Hertz were converted into the psycho-acoustical Bark scale using equation (2) given by Traunmüller[61]

$$F(\text{Barks}) = \left[\frac{26.81}{\left(1 + \left(\frac{1960}{F(\text{Hz})}\right)\right)} \right] - 0.53 \tag{2}$$

where F(Hz) is the value of the formant frequencies in Hertz and F(Barks) is the transformed Bark value.

To calculate the vowel working space, from the mean F1 and F2 frequencies, the F1, F2 coordinates for each vowel were fixed and the area of the polygon formed from the F1, F2 coordinates of each vowel was derived. The area of the polygon is the VSA. Inter-talker and intra-talker variations in VSAs were compared from speech samples from two different modes of delivery and in the presence and absence of noise.

Results and discussion

Fundamental frequency measures

Influence of mode of delivery of speech. Table 3 shows the fundamental frequency measures across talkers for two different speaking modes namely reading and lecturing. The difference in fundamental frequency measures between the two speaking modes is transformed from hertz to logarithmic semitones (ST) to perceptually normalize the speakers for easy comparison. There is a considerable inter-talker and intra-talker variation in the fundamental frequency characteristics between two modes of speech delivery. Majority of the talkers exhibit an increase in their fundamental frequency measures when delivering speech in lecture mode ($M = 230.5$, $SD = 63.9$)

Table 4. F0-mean and F0-SD of lecture with and without background noise (BGN).

Talker	Lecture mode (without BGN)		Lecture mode (with BGN)		Difference in ST (F0-mean)	Difference in ST (F0-SD)
	F0-mean (Hz)	F0-SD (Hz)	F0-mean (Hz)	F0-SD (Hz)		
1	297	59	279	50.1	−1.1	–2.8
2	232	44	244	59.4	0.9	5.2
3	141	38	164	64.9	2.6	9.3
4	279	56	326	62.6	2.7	1.9
5	281	56	285	58.6	0.3	0.8
6	149	29	231	48	7.6	8.7

SD: standard deviation; BGN: background noise; ST: semitones.

than in the reading mode (M = 216.4, SD = 58.8). An increase in F0-mean could be associated with increased intelligibility or an effect of vocal loading. The results of the two-factor ANOVA for F0-mean showed that the primary factor (reading vs lecture modes) was significantly different ($F(1, 36) = 101.8 > F_{crit} = 4.11$; $p < 0.0001$). Similarly, the speakers were also significantly different in their speech delivery ($F(5, 36) = 1431.5 > F_{crit} = 2.47$; $p < 0.0001$). There was a significant interaction between mode of delivery and speakers ($F(5, 36) = 36.6 > F_{crit} = 2.47$; $p < 0.0001$). Increase in liveliness of speech is characterized by increased excursions of F0 from its mean (F0-SD) and this is seen in lecture mode of delivery (M = 49.4, SD = 9.03) when compared to the reading mode (M = 47.8, SD = 6.59) for majority of the talkers. Talker 6 shows more liveliness in reading mode, indicating that reading always need not be associated with monotones. Results of the two-way ANOVA indicated that the primary factor (reading vs lecture mode) was not significantly different ($F(1, 36) = 1.9 < F_{crit} = 4.11$; $p = 0.17$), but speakers significantly modified the liveliness of their speech across the two modes of delivery ($F(5, 36) = 19.9 > F_{crit} = 2.47$; $p < 0.000$). There was also a significant interaction between the mode of delivery and speakers ($F(5, 36) = 10.9 > F_{crit} = 2.47$; $p < 0.000$). Also, female talkers in the group (Talkers 1, 4, and 5) exhibit more pitch variation (F0-SD) compared to male talkers as seen in another study.[2]

Influence of classroom BGN. Table 4 shows the fundamental frequency characteristics of all talkers in lecture mode with and without classroom BGN. When talkers are exposed to noise during lecture mode speech, majority of the talkers exhibit significant differences in F0-mean and F0-SD with an increase in the F0-mean (M = 281.1, SD = 36.8) and F0-SD (M = 82.4, SD = 31.5) in the presence of noise compared to F0-mean (M = 230.5, SD = 63.9) and F0-SD (M = 49.4, SD = 9.03) in lecture mode without BGN.

The results of the two-way ANOVA for F0-mean revealed significant effects of noise ($F(1, 36) = 694.5 > F_{crit} = 4.11$; $p < 0.000$) and speakers ($F(5, 36) = 442.7 > F_{crit} = 2.48$; $p < 0.000$), with a significant interaction between noise and speakers ($F(5, 36) = 116.3 > F_{crit} = 2.48$; $p < 0.000$). Similarly, the ANOVA results for F0-SD also showed significant effect of factor noise ($F(1, 36) = 906.2 > F_{crit} = 4.11$; $p < 0.000$) and speakers ($F(5, 36) = 106.2 > F_{crit} = 2.47$; $p < 0.000$) with a significant interaction between the two factors ($F(5, 36) = 230.4 > F_{crit} = 2.47$; $p < 0.000$). Overall, this suggests that talkers modulate their fundamental frequency characteristics in response to the ambient noise to enhance their intelligibility. This is in agreement with studies that have compared speech in controlled and real environments.[28,62]

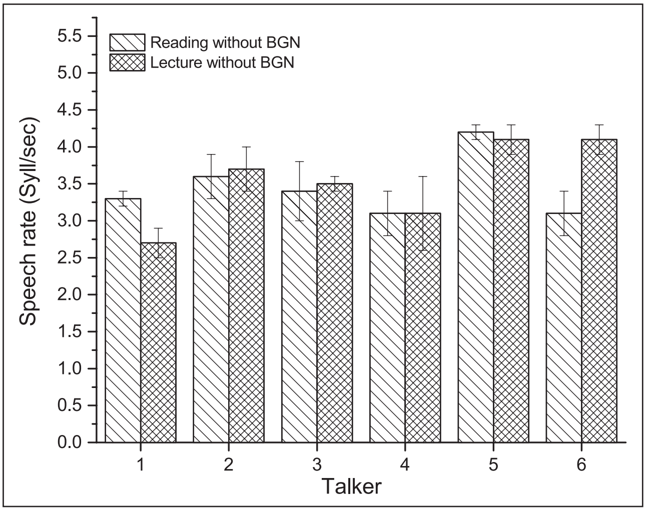

Figure 2. Speech rate across talkers for reading and lecture modes.

Durational measures

Influence of mode of delivery of speech

Speech rate. Figure 2 indicates the speech rate (syll/s) across all talkers when eliciting speech under reading and lecture modes. The results of the two-way ANOVA for speech rates indicated that the primary factor (reading vs lecture) was not significantly different ($F(1, 36) = 0.177 < F_{crit} = 4.11$; $p = 0.67$), but speakers significantly altered their speech delivery ($F(5, 36) = 11.5 > F_{crit} = 2.47$; $p < 0.000$). There was also a significant interaction between the mode of delivery and speakers ($F(5, 36) = 3.75$; $p < 0.01$). Speech rate varies across talkers with marginal intra-talker variation between reading and lecture style of speaking. Inter-talker variation may also be attributed to differences in reading fluency, cognitive skills, speaking skills, and so on.

Pause rate. Figure 3 shows the mean pause rate (pauses per second) for all talkers across the two modes of speech delivery. Talkers vary their pause characteristics with almost all talkers exhibiting an increase in the rate of pauses under lecture mode ($M = 2.8$, $SD = 0.4$) when compared to reading mode ($M = 2.5$, $SD = 0.4$).

The results of the two-way ANOVA for pause rate revealed that the primary factor (reading vs lecture) was significantly different ($F(1, 36) = 10.24 > F_{crit} = 4.11$; $p < 0.01$). The speakers were also significantly different in their way of speaking ($F(5, 36) = 6.86 > F_{crit} = 2.47$; $p < 0.001$). However, there was no significant interaction between the factors ($F(5, 36) = 0.61 < F_{crit} = 2.47$; $p = 0.68$).

Reading does not require as much cognitive processing/thinking as lecturing does and so the insertion of pauses in read text is only dependent on reading fluency and oral reproduction capability of the talkers. In the case of lecturing mode, talkers need to think and talk to deliver a meaningful lecture and the insertion of frequent pauses in this mode of speech gives the talker time for cognitive processing and delivery of speech.

Mean pause length. Figure 4 shows the mean pause length (ms) for all six talkers in reading and lecturing modes. Majority of the talkers show a significant decrease in mean pause lengths

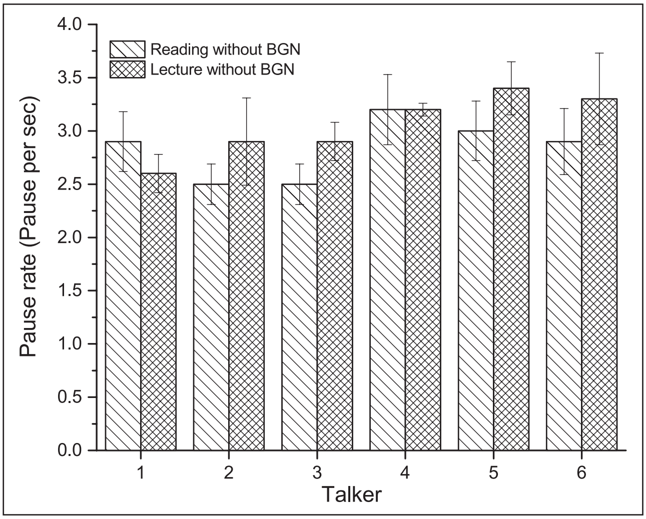

Figure 3. Pause rate across talkers for reading and lecture modes.

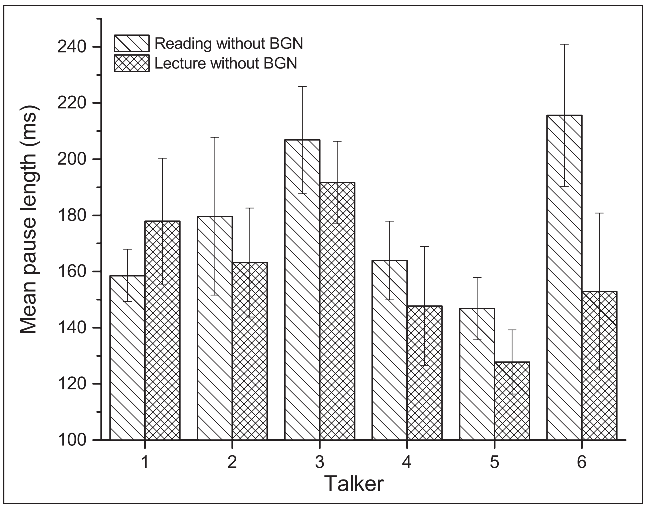

Figure 4. Mean pause length of talkers across speaking conditions.

in lecture mode (M = 150.8, SD = 36.3) when compared to reading mode (M = 180.2, SD = 42.6). Two-way ANOVA results for mean pause length reveal that the factor, mode of delivery, was significantly different ($F(1, 36) = 12 > F_{crit} = 4.11$; $p < 0.001$) and the speakers were also significantly different in their method of delivery ($F(5, 36) = 5.9 > F_{crit} = 2.47$; $p < 0.001$). There was also a significant interaction effect between the speakers and mode of delivery ($F(5, 36) = 4.69 > F_{crit} = 2.47$; $p < 0.01$).

Although pause rate has increased, the mean pause lengths are reduced in lecture mode resulting in more frequent pauses of shorter duration for majority of the talkers. Talkers typically use “filled pauses” and interject them into their speech to give themselves sufficient time to think

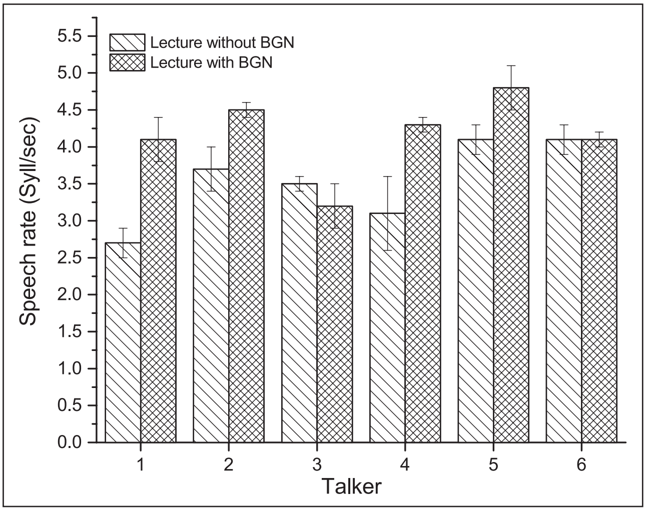

Figure 5. Speech rate across talkers for lecture mode with and without background noise.

about what next to say. The reduction in mean pause lengths in lectures could be attributed to this phenomenon.

Influence of classroom back ground noise

Speech rate. Figure 5 shows the variation in speech rates of talkers when eliciting lectures in the presence and absence of classroom BGN.

Speech rate of talkers in lecture mode (M = 4.12, SD = 0.64) in the presence of classroom BGN is higher compared to speech rate in lecture mode without BGN (M = 3.5, SD = 0.64), contradictory to popular belief that Lombard speech[18] (speech in noise) is characterized by lower speech rates. The Lombard effect is triggered in the presence of noise that contains similar speech frequencies. The octave band U_{50} values (Table 2) reflect the influence of noise across relevant speech frequencies and the influence of noise on talkers shows significant effects. Two-way ANOVA results for speech rate indicate that the primary factor (lecture in noise vs quiet) was significantly different ($F(1, 36) = 24.9 > F_{crit} = 4.11$; $p < 0.001$). Speakers also were significantly different in their speech delivery ($F(5, 36) = 9.9 > F_{crit} = 2.47$; $p < 0.001$). There was also a significant interaction between noise and speakers ($F(5, 36) = 4.52 > F_{crit} = 2.47$; $p < 0.01$). This reflects on talkers' capability to adapt to degraded speaking conditions.

Pause rate. Looking at pause characteristics of talkers when eliciting speech in lecture mode in the absence and presence of noise from Figure 6, there is a significant increase in pause rate (M = 3.73, SD = 0.47) across all talkers when speaking in the presence of noise ($F(1, 36) = 112.5 > F_{crit} = 4.11$; $p < 0.001$) with speakers being significantly different in their speech delivery ($F(5, 36) = 3.67 > F_{crit} = 2.47$; $p < 0.01$). There was also a significant interaction between noise and speakers ($F(5, 36) = 8.76 > F_{crit} = 2.47$; $p < 0.000$). The noise present may increase the cognitive effort of the talker, increasing the instances of "filled pauses" and which results in increased rate of speech and pauses per second.

Mean pause length. Figure 7 shows that the presence of noise results in a decrease in pause length (M = 129.5, SD = 27) and this can also be partially attributed to higher instances of "filled pauses"

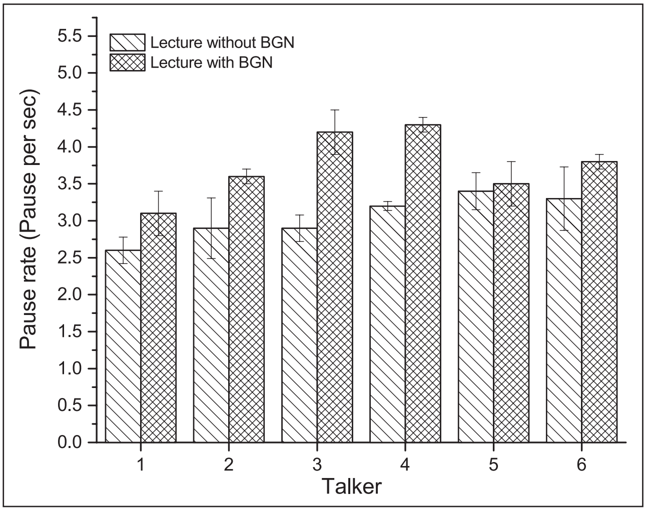

Figure 6. Pause rate across talkers for lecture mode with and without background noise.

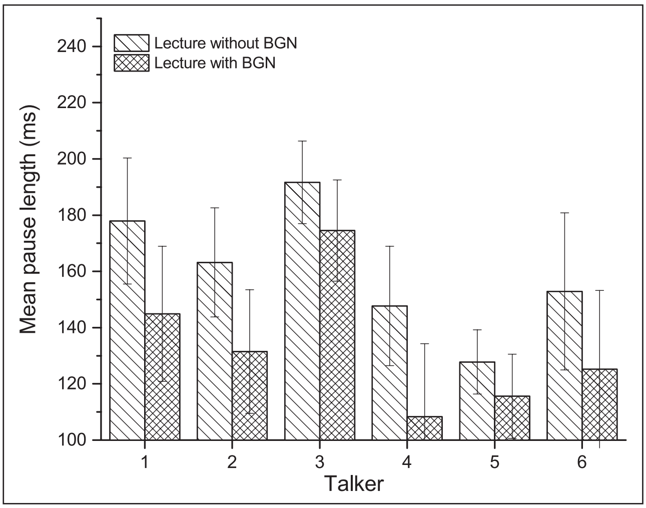

Figure 7. Mean pause length for lecture mode with and without background noise.

that are typical of lecture mode of speech. The two-way ANOVA results reveal that the primary factors (lecture with and without BGN) were significantly different ($F(1, 36) = 9.70 > F_{crit} = 4.11$; $p < 0.01$). The speakers were also significantly different in their method of speech delivery ($F(5, 36) = 7 > F_{crit} = 2.47$; $p < 0.000$). There was also a significant interaction between noise and speakers ($F(5, 36) = 2.81 > F_{crit} = 2.47$; $p < 0.05$).

Intra-talker variation in vowel space properties

Influence of mode of delivery. Figure 8 shows the intra-talker difference in vowel articulation between the two speech modes as quantified by the VSA.

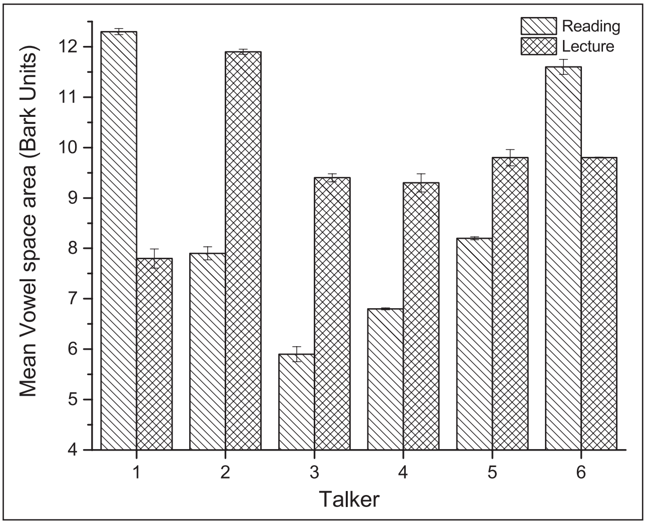

Figure 8. Intra-talker variation in vowel space area between reading and lecture modes.

The intra-talker difference in VSA between reading and lecture mode varied from −4.0 to 4.5 bark units, with higher VSA in lecture mode (M = 9.87, SD = 3.71), when compared to reading mode (M = 9.00, SD = 4.14). Talker 1 shows maximum vowel articulation in reading mode and Talker 2 shows the maximum vowel articulation in lecture mode of delivery. Talker 1 shows maximum variation between the two speech modes, whereas Talker 5 shows the least. Considering that the VSA is a robust indicator of talker intelligibility, it is seen that talker intelligibility could be better in lecturing mode. Talkers 2–5 show increased vowel articulation in lecture mode, whereas Talkers 1 and 6 show better articulation in reading mode. This indicates that talkers modify their speech characteristics differently with change in speaking mode and are in agreement with previous studies.[51,63]

Figure 9 illustrates the variation in the articulatory definitions of vowels, between reading and lecturing modes of speech delivery for each of the six talkers separately (Figure 9(a)–(f)). All talkers show variation in their utterances in terms of vowel formant frequencies, F1 and F2, between reading and lecture mode of deliveries. For four out of six talkers, VSA reduces in reading mode compared to lecture mode, so there is shrinkage of vowel space for majority of talkers in reading mode compared to lecture mode.

A considerable increase in the mean F1 for both the front vowels /i/ (M = 3.74, SD = 0.74) and /a/ (M = 7.9, SD = 0.61) is seen for all talkers in lecture mode, whereas an increase in the mean F2 is also seen for front vowel /a/ (M = 9.4, SD = 1.2) along with the other low vowel /ɑ/ (M = 11.6, SD = 0.9). The difference in vowel /i/ articulation ($F(1, 24) = 11.16 > F_{crit} = 4.25$; $p < 0.01$) and vowel /a/ articulation ($F(1, 24) = 6.59 > F_{crit} = 4.25$; $p < 0.05$) between the two modes of delivery was found to be significant as per the results of the two-way ANOVA with replications. The change in F2 of high vowels (/i/, /u/) was marginal compared to the change in F2 of low vowels (/a/, /ɑ/). The difference in the mean articulatory vowel positions in the F1 × F2 plane in lecture mode of delivery as shown in Figure 9(a)–(f) indicates different ways of articulation from reading mode. For example, the increased mean F1 and F2 of vowel /i/ in lecture mode of delivery indicates more open and more front articulatory movement and increased mean F1 of low vowels indicates that talkers open their mouths wider for lecturing than reading.

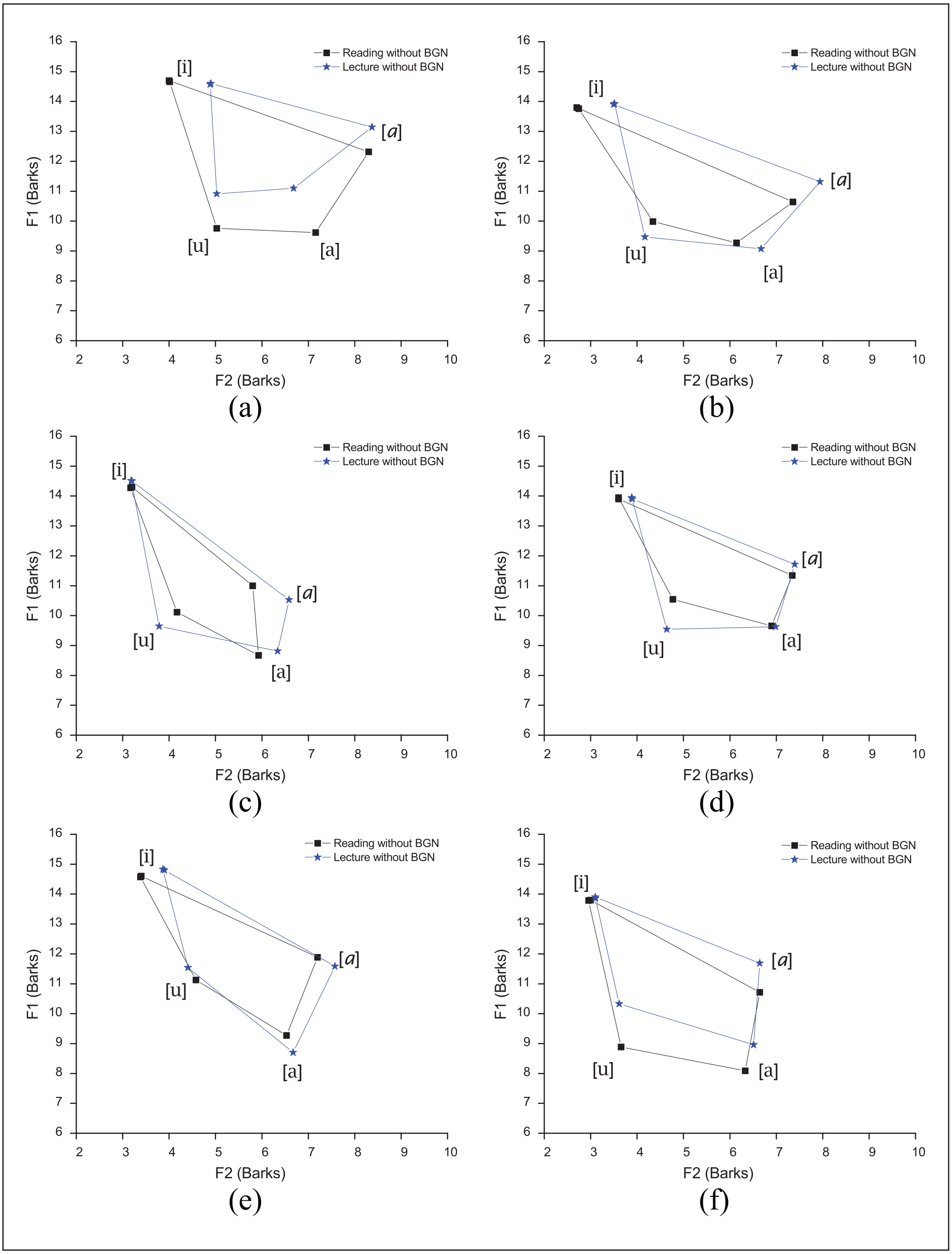

Figure 9. Intra-talker variation in VSA between reading and lecture modes: (a) Talker 1, (b) Talker 2, (c) Talker 3, (d) Talker 4, (e) Talker 5, and (f) Talker 6.

Also, a decrease in the mean F1 and increase in the mean F2 of vowel /u/ in lecture mode indicate less lip-rounding or closed articulatory movement. VSA alone may not be sufficient to

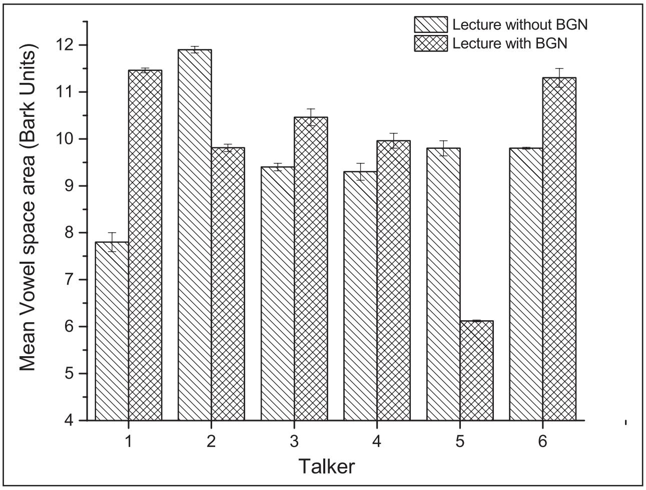

Figure 10. Intra-talker variation in VSA with and without background noise.

understand the intra-talker variability in vowel articulation between the two modes of speech delivery, as for the similar VSAs, the articulation of vowels in the F1 × F2 plane may differ.

Influence of classroom BGN. Figure 10 presents the intra-talker variation in VSAs between lecture modes with and without BGN.

Majority of the talkers exhibit higher VSAs in the presence of BGN in the classroom. Talkers 2 and 5 show reduction in their vowel articulation in the presence of noise.

Mean VSA across talkers in lecture mode with and without BGN varies from 7.8 to 11.9 bark units and from 6.1 to 11.5 bark units for lecture mode with BGN. With a relatively narrower range of VSA for lecture mode in the presence of BGN, an overall vowel reduction happens across talkers in their lecture delivery in the presence of noise. The mean VSA across talkers for lecture mode with and without BGN are 9.9 and 9.7 bark units, respectively.

The intra-talker difference in VSA between the two conditions varies from −3.6 to 3.6 bark units, with a mean difference of 0.2 bark units between the two conditions. The intra-talker difference in VSA between the two conditions varies from −3.6 to 3.6 bark units, with a mean difference of 0.2 bark units between the two conditions. However, the individual differences in vowel articulation between the two speaking conditions (with and without BGN) are analyzed from the individual vowel space diagrams for each individual talker presented in Figure 11.

Inter-talker variation in vowel space properties

Influence of mode of delivery

Reading and lecture modes. From the range of F1 and F2 values obtained using seven tokens per vowel for all talkers, for the four cardinal vowels for all speakers, the mean VSAs were determined across reading and lecture modes, as presented in Figures 12 and 13, respectively.

The mean VSA across talkers for reading mode is 8.8 bark units and for lecture mode is 9.7 bark units. The mean VSAs in reading mode range from 5.9 to 12.3 bark units and those in lecture mode across all talkers range from 7.8 to 11.9 bark units, with greater variation in reading mode compared to lecture mode for all talkers. All talkers show similar articulatory modifications in the F1 dimension for

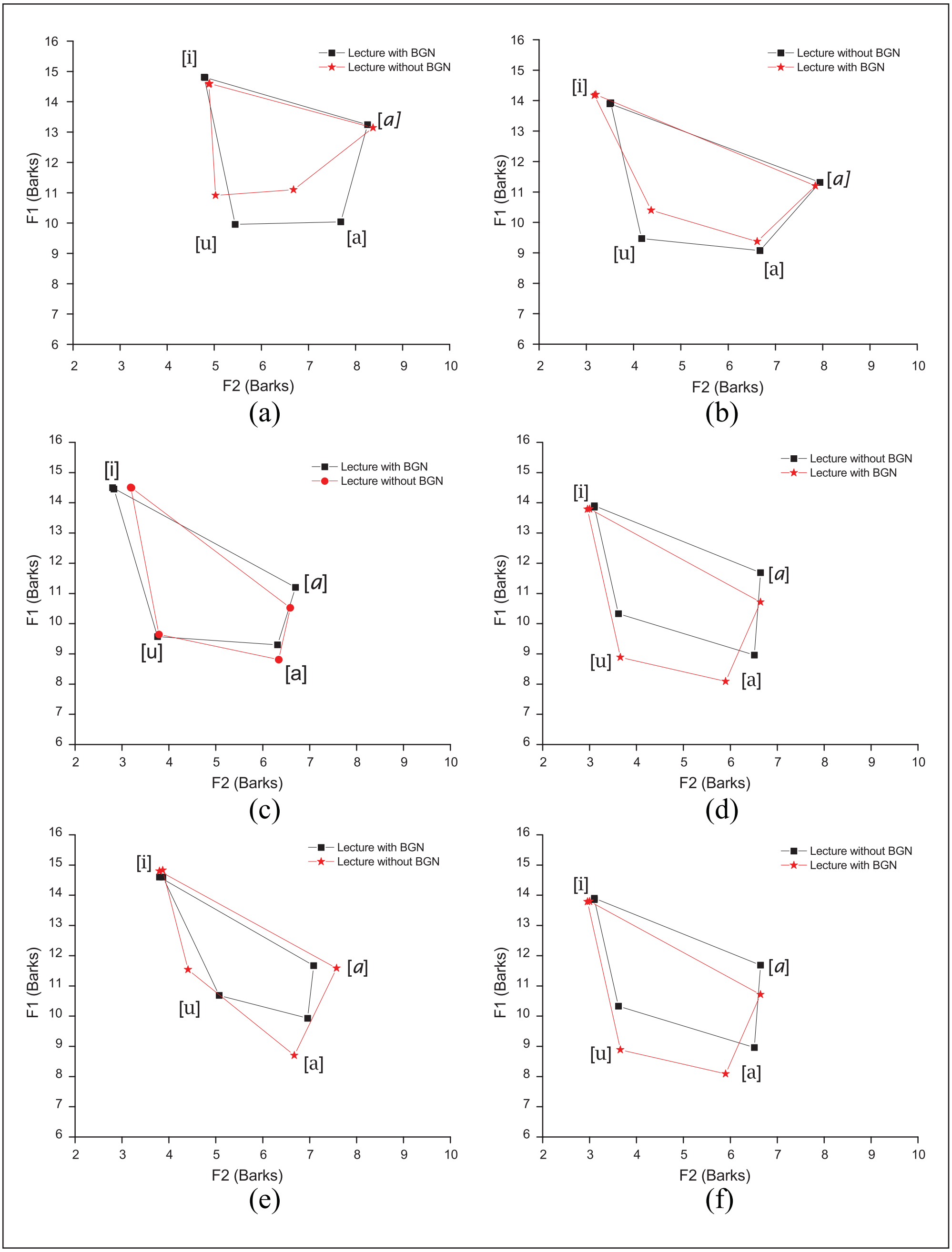

Figure 11. Intra-talker variation in vowel space for lecture mode with and without noise: (a) Talker 1, (b) Talker 2, (c) Talker 3, (d) Talker 4, (e) Talker 5, and (f) Talker 6.

vowels /i/ and /a/, with greater variation in mean F1 of vowel /i/ in lecture mode and greater variation in mean F1 of vowel /a/ in reading mode. The results of the two-way ANOVA indicated no significant effects of mode of delivery on VSA ($F(1, 36)=0.78 < F_{crit}=4.11$; $p=0.38$); however, the speakers

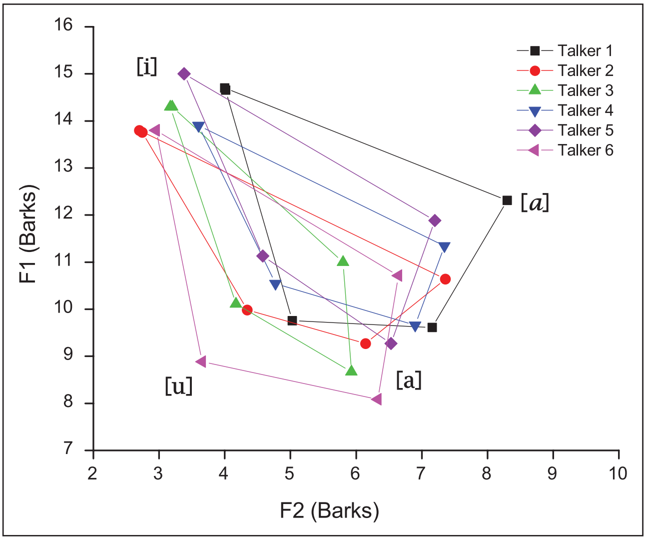

Figure 12. Inter-talker variation in VSA in reading mode without BGN.

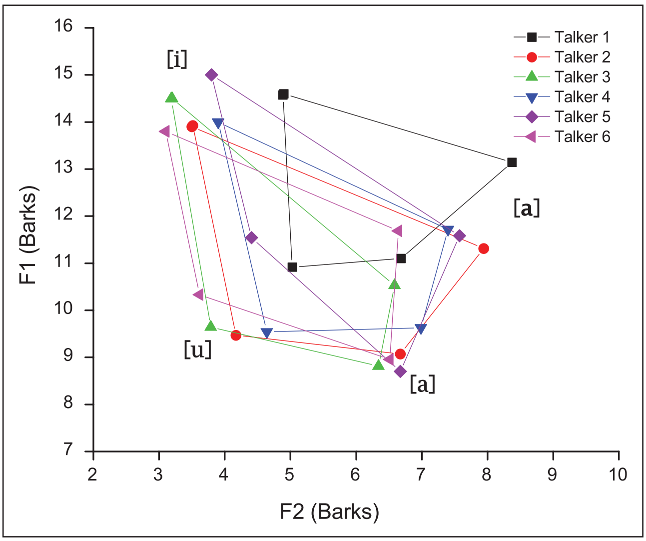

Figure 13. Inter-talker variation in VSA in lecture mode without BGN.

differed significantly in their vowel articulation ($F(5, 36)=2.66>F_{crit}=2.47$; $p<0.05$) and there was also a significant interaction between mode of delivery and speakers ($F(5, 36)=2.9>F_{crit}=2.47$; $p<0.05$). This indicates that overall, talkers exhibit more front and closed articulatory movement in lecture style of speaking, but open their mouths wider in the reading mode of delivery.

Influence of classroom BGN. To further evaluate the influence of the presence of noise on vowel working space, the VSAs were measured for all talkers when subject to classroom noise (lower

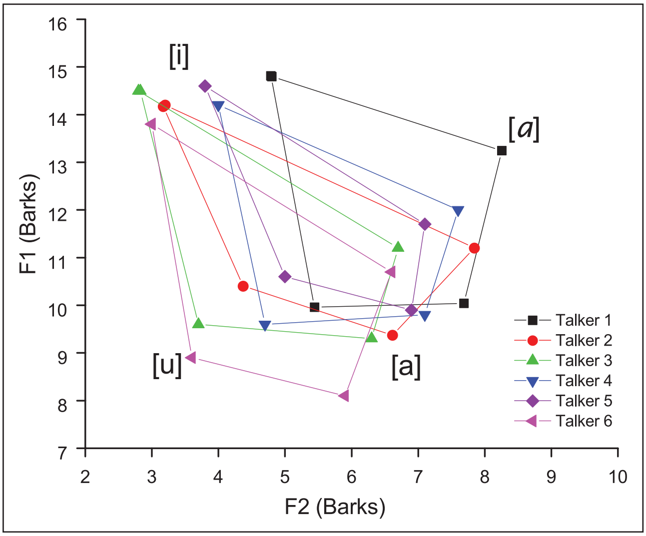

Figure 14. Inter-talker variation in mean VSA for lecture mode with BGN.

U_{50}). Figure 14 shows the variation in mean vowel working spaces across talkers when lecturing in the presence of classroom noise. The mean VSAs across talkers when speaking in lecture mode without BGN varies from 7.8 to 11.9 bark units and from 6.7 to 11.5 when lecturing with BGN. Mean VSAs across talkers increase from 9.7 to 9.9 when BGN is present. The variation in mean VSAs across talkers is greater when lecturing with BGN indicating that talkers articulate in different ways in the presence of noise to overcome it and to maintain optimum level intelligibility. Talkers' speech response in the presence of noise is as expected and in agreement with results from previous studies.[3,64]

The results of the two-way ANOVA for F1 of vowel /i/ and /ɑ/ show that the primary factor (with and without BGN) was significantly different ($F(1, 36)=4.58>F_{crit}=4.11$; $p<0.05$). The speakers also were significantly different in their vowel articulation in the presence of noise ($F(5, 36)=15.8>F_{crit}=2.47$; $p<0.000$). This indicates that talkers use a combination of articulatory strategies to overcome the presence of noise and its interference in their speech.

For talkers with intrinsically lower intelligibility, the presence of noise could be further detrimental. Majority of the talkers adapt their speech and modify their vowel spaces to enhance their intelligibility in detrimental conditions. VSA by itself may not be a robust indicator of speech intelligibility when comparing talkers, as for similar VSAs, the vowel spaces could be different with different articulatory positions of the vowels in the $F1 \times F2$ plane, which can perceptibly alter speech intelligibility.

Conclusion

This study investigated the acoustic–phonetic characteristics of six talkers for two different modes of speech delivery in a typical classroom environment. The study also investigated whether speech modifications made in the classroom under lecture mode style of speaking differ in the presence of typical classroom BGN (operation of ceiling fans). The results provide information about the sources of variability and their effects on the speech intelligibility correlate. Specifically, change in the mode of delivery (reading/lecturing) was found to significantly influence speech intelligibility correlates such as F0-mean, pause rate, and mean pause length. The presence/absence of BGN was found to significantly influence F0-mean, F0-SD, speech rate, pause rate, and mean pause length. Talkers significantly varied their speech in terms of all the correlates studied. The main aim of the study was to evaluate the significant differences in the talkers' speech characteristics as a result of adopting lecturing mode compared to reading texts and when lecturing under typical classroom BGN. The results from the study are in agreement with previous studies indicating the influence of speaking style and room on vocal effort and speech production.[65,66] The salient conclusions from the study are as follows:

1. Fundamental frequency characteristics of talkers in terms of F0-mean and liveliness (F0-SD) are higher when the mode of delivery is lecture style and further increase in the presence of noise.
2. Speech rates and pause rates are lower for reading modes and higher in lecture modes of speech.
3. Mean pause length significantly reduces in lectures delivered in the presence of BGN.
4. Talkers show similar vowel-articulatory changes under different modes of speaking.
5. When delivering lectures in the presence of BGN, talkers utilize a combination of articulatory strategies in different vowels to overcome the presence of interfering BGN.

Considering that correlates of intelligibility were significantly higher for lecture mode of delivery in noise, it may be prudent to evaluate speech characteristics of classroom talkers using running speech from lecture recordings in natural classroom environments rather than from pre-defined and pre-administered passages, sentences, and words. Moreover, talker speech characteristics that are relevant for classroom discourse and student engagement need to be ascertained by examining different phonetic and prosodic features of the talkers' speech in varying classroom acoustical conditions. In future, relationships between room correlates of intelligibility such as U_{50} and STI and talker correlates of intelligibility need to be explored systematically to rank the acoustical quality of classrooms that are ideal for both listening and speaking.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Nithya Subramaniam https://orcid.org/0000-0003-0751-5026
K Ramamurthy https://orcid.org/0000-0001-6177-6230

References

1. Bond ZS and Moore TJ. A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Commun* 1994; 14: 325–337.
2. Bradlow AR, Torretta GM and Pisoni DB. Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun* 1996; 20: 255–272.
3. Hazan V and Baker R. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J Acoust Soc Am* 2011; 130: 2139–2152.

4. Lindblom B. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle WJ and Marchal A (eds) *Speech production and speech modelling*. Dordrecht: Springer, 1990, pp. 403–439.
5. Perkell JS, Guenther FH, Lane H, et al. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *J Phon* 2000; 28: 233–272.
6. Pelegrín-García D and Brunskog J. Speakers' comfort and voice level variation in classrooms: laboratory research. *J Acoust Soc Am* 2012; 132: 249–260.
7. Traunmüller H and Eriksson A. The frequency range of the voice fundamental in the speech of male and female adults. *Dep Linguist Univ Stock* 1994; 97: 1905191–1905195.
8. Emma PG, Greta M, Arianna A, et al. Spatial benefits on speech intelligibility in real classroom acoustics under energetic and informational masking noise, http://www.sea-acustica.es/fileadmin//INTERNOISE_2019/Fchrs/Proceedings/1574.pdf
9. Ferguson SH and Kewley-Port D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 2002; 112: 259–271.
10. Bradlow AR, Kraus N and Hayes E. Speaking clearly for children with learning disabilities. *J Speech Lang Hear Res* 2003; 46: 80–97.
11. Castellana A, Carullo A, Astolfi A, et al. Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. *J Acoust Soc Am* 2017; 141: 2353–2363.
12. Astolfi A, Castellana A, Puglisi GE, et al. Speech level parameters in very low and excessive reverberation measured with a contact-sensor-based device and a headworn microphone. *J Acoust Soc Am* 2019; 145: 2540–2551.
13. Knudsen VO and Harris CM. *Acoustical designing in architecture*. Hoboken, NJ: John Wiley & Sons, 1980.
14. Lochner JPA and Burger JF. The influence of reflections on auditorium acoustics. *J Sound Vib* 1964; 1: 426–454.
15. Pollard KA, Tran PK and Letowski T. The effect of vocal and demographic traits on speech intelligibility over bone conduction. *J Acoust Soc Am* 2015; 137: 2060–2069.
16. MacDonald Bolt RH. Theory of speech masking by reverberation. *J Acoust Soc Am* 1949; 435: 577–580.
17. Houtgast T, Steeneken H and Plomp R. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acta Acust United Acust* 1980; 46: 60–72.
18. Lane H and Bernard T. The Lombard sign and the role of hearing in speech. *J Speech Hear Res* 1971; 14: 677–709.
19. Summers W, Van Pisoni DB, Bernacki RH, et al. Effects of noise on speech production: acoustic and perceptual analyses. *J Acoust Soc Am* 1988; 84: 917–928.
20. Pelegrín-García D, Brunskog J, Lyberg-Åhlander V, et al. Measurement and prediction of voice support and room gain in school classrooms. *J Acoust Soc Am* 2011; 131: 194–204.
21. Bradley JS. Predictors of speech intelligibility in rooms. *J Acoust Soc Am* 1986; 80: 837–845.
22. Bradley JS. *Uniform derivation of optimum conditions for speech in rooms*, 1985, https://nrc-publications.canada.ca/eng/view/accepted/?id=8afe9439-61b0-46bb-b283-389ad4403405
23. Choi Y-J. Comparison of two types of combined measures, STI and U50, for predicting speech intelligibility in classrooms. *Arch Acoust* 2017; 42: 527–532.
24. Hazan V and Markham D. Acoustic-phonetic correlates of talker intelligibility for adults and children. *J Acoust Soc Am* 2004; 116: 3108–3118.
25. Amano-Kusumoto A. *Relationship between acoustic features and speech intelligibility*. Portland, OR: Oregon Health & Sciences University, 2010.
26. Astolfi A, Carullo A, Pavese L, et al. Duration of voicing and silence periods of continuous speech in different acoustic environments. *J Acoust Soc Am* 2015; 137: 565–579.
27. Södersten M, Granqvist S, Hammarberg B, et al. Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural DAT recordings. *J Voice* 2002; 16: 356–371.
28. Lindstrom F, Ohlsson AC, Sjöholm J, et al. Mean F0 values obtained through standard phrase pronunciation compared with values obtained from the normal work environment: a study on teacher and child voices performed in a preschool environment. *J Voice* 2010; 24: 319–323.

29. Szabo Portela A, Hammarberg B and Södersten M. Speaking fundamental frequency and phonation time during work and leisure time in vocally healthy preschool teachers measured with a voice accumulator. *Folia Phoniatr Logop* 2014; 65: 84–90.
30. Picheny MA, Durlach NI and Braida LD. Speaking clearly for the hard of hearing II. *J Speech Lang Hear Res* 1986; 29: 434.
31. Krause JC and Braida LD. Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am* 2004; 115: 362–378.
32. Hincks R. Standard deviation of F_0 in student monologue. In: *Fonetik* 2004, 2003, pp. 132–135, https://pdfs.semanticscholar.org/80aa/46df3c05c15ed08c380d32e89ac3f81f003f.pdf
33. Hincks R. Processing the prosody of oral presentations. In: *Proceedings of the InSTIL/ICALL symposium computer assisted language learning*, Venice: Unipress, 17–19 June 2004, pp. 63–69.
34. Rosenberg A and Hirschberg J. Acoustic/prosodic and lexical correlates of charismatic speech, http://www.cs.columbia.edu/nlp/papers/2005/rosenberg_hirschberg_05.pdf
35. Strangert E. What makes a good speaker ? Subjective ratings and acoustic measurements. In: *Interspeech*, Brisbane, QLD, Australia, 22–26 September 2008, pp. 1688–1691. Brisbane, QLD, Australia: ISCA.
36. Drullman R and Bronkhorst AW. Speech perception and talker segregation: effects of level, pitch, and tactile support with multiple simultaneous talkers. *J Acoust Soc Am* 2004; 116: 3090–3098.
37. Hahn L. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Q* 2004; 38: 201–223.
38. Pickering L. The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English Specif Purp* 2004; 23: 19–43.
39. Uchanski RM, Choi SS, Braida LD, et al. Speaking clearly for the hard of hearing IV: further studies of the role of speaking rate. *J Speech Hear Res* 1996; 39: 494–509.
40. Cox RM, Alexander GC and Gilmore C. Intelligibility of average talkers in typical listening environments. *J Acoust Soc Am* 1987; 81: 1598–1608.
41. Bradlow AR and Pisoni DB. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *J Acoust Soc Am* 1999; 106: 2074–2085.
42. Adams EM and Moore RE. Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *J Am Acad Audiol* 2009; 20: 28–39.
43. Wingfield A and Ducharme JL. Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *J Gerontol B Psychol Sci Soc Sci* 1999; 54: P199–P202.
44. Picheny MA, Durlach NI and Braida LD. Speaking clearly for the hard of hearing III. *J Speech Lang Hear Res* 1989; 32: 600.
45. Bond ZS, Moore TJ and Gable B. Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *J Acoust Soc Am* 1989; 85: 907–912.
46. Lié Nard J-S and Di Benedetto M-G. Effect of vocal effort on spectral properties of vowels. *J Acoust Soc Am* 1999; 106: 411–422.
47. Amano-Kusumoto A and Hosom J. A review of research on speech intelligibility and correlations with acoustic features. *Science* 2011; 80: 1–16.
48. Perkell JS, Zandipour M, Matthies ML, et al. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *J Acoust Soc Am* 2002; 112: 1627–1641.
49. Smiljanić R and Bradlow AR. Temporal organization of English clear and conversational speech. *J Acoust Soc Am* 2008; 124: 3171–3182.
50. Picheny MA, Durlach NI and Braida LD. Speaking clearly for the hard of hearing I. *J Speech Lang Hear Res* 1985; 28: 96–103.
51. Ferguson SH and Kewley-Port D. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J Speech Lang Hear Res* 2007; 50: 1241–1255.
52. Helfer KS. Auditory and auditory-visual recognition of clear and conversational speech by older adults. *J Am Acad Audiol* 1998; 9: 234–242.
53. Maclay H and Osgood CE. Hesitation phenomena in spontaneous English speech. *WORD* 1959; 15: 19–44.

54. Miller GA, Heise GA and Lichten W. The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 1951; 41: 329–335.
55. Cantor Cutiva LC, Puglisi GE, Astolfi A, et al. Four-day follow-up study on the self-reported voice condition and noise condition of teachers: relationship between vocal parameters and classroom acoustics. *J Voice* 2017; 31: 120.e1–120.e8.
56. Cipriano M, Astolfi A and Pelegrín-García D. Combined effect of noise and room acoustics on vocal effort in simulated classrooms. *J Acoust Soc Am* 2017; 141: EL51–EL56.
57. Šrámková H, Granqvist S, Herbst CT, et al. The softest sound levels of the human voice in normal subjects. *J Acoust Soc Am* 2015; 137: 407–418.
58. Astolfi A, Castellana A, Carullo A, et al. Uncertainty of speech level parameters measured with a contact-sensor-based device and a headworn microphone. *J Acoust Soc Am* 2018; 143: EL496–EL502.
59. Paul B and van Vincent H. Speak and unSpeak with PRAAT. *Glot Int* 2001; 5: 341–347.
60. Fairbanks G. *Voice and articulation drillbook*. 2nd ed. New York: Harper & Row, 1960.
61. Traunmüller H. Analytical expressions for the tonotopic sensory scale. *J Acoust Soc Am* 1990; 88: 97.
62. Puglisi GE, Astolfi A, Cantor Cutiva LC, et al. Four-day-follow-up study on the voice monitoring of primary school teachers: relationships with conversational task and classroom acoustics. *J Acoust Soc Am* 2017; 141: 441.
63. Ferguson SH. Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. *J Acoust Soc Am* 2004; 116: 2365–2373.
64. Cooke M and Lu Y. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *J Acoust Soc Am* 2010; 128: 2059–2069.
65. Pelegrin-Garcia D, Lyberg-Åhlander V, Rydell R, et al. Influence of classroom acoustics on the voice levels of teachers with and without voice problems: a field study. *Proc Mtgs Acoust* 2010; 11: 060001.
66. Bottalico P, Graetzer S and Hunter EJ. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *J Acoust Soc Am* 2016; 139: 2870–2879.