Language Variation and Change (2022), **34**, 305–329
doi:10.1017/S0954394522000151

Medium-shifting and intraspeaker variation in conversational interviews

Isaac L. Bleaman[1]*, Katie Cugno[2] and Annie Helms[1]

[1]University of California, Berkeley, USA and [2]San Francisco State University, USA
*Corresponding author. E-mail: bleaman@berkeley.edu

Abstract

We investigate the impact of medium of communication (in-person versus video) on intraspeaker variation in conversation—a process we refer to as *medium-shifting*. To quantify the effects of medium-shifting and understand its possible motivations, we analyze three variables that show intraspeaker effects of "clear" or "careful" speech: articulation rate, density-controlled vowel space area, and (ING). The data come from matched in-person and video-mediated interviews with thirty-three repeat guests from *The Late Show with Stephen Colbert*, recorded before and during the COVID-19 pandemic. Mixed-effects regression models show that compared to in-person interviews, video-mediated interviews involve a significantly lower articulation rate and larger vowel space, but no significant difference in (ING). The results suggest that speakers may engage in medium-shifting in order to enhance their intelligibility over video, for example, through more precise articulatory movements and greater contrast between phonemic vowel categories. The null effect of medium on (ING) further suggests that medium-shifting is a motivator of intraspeaker differences even within a single contextual style. An emergent extralinguistic factor affecting speaking behavior and choices, medium-shifting should be carefully considered especially when designing variationist research involving mixed media interviews.

Keywords: intraspeaker variation; medium-shifting; sociophonetics; speech rate; vowel space area; (ING); COVID-19 pandemic

As a result of the restrictions on travel and social gatherings during the COVID-19 pandemic, many have turned to video conferencing as a necessary and effective substitute for in-person communication. Scholars from a variety of disciplines have predicted that the increased reliance on video conferencing platforms including Zoom will be one of the lasting legacies of the pandemic, establishing a "new normal" for remote work (Hermann & Paris, 2020), distance learning (Deflem, 2021), telehealth (Keesara, Jonas, & Schulman, 2020), and other domains. Of course, the pandemic has also had an impact on the practice of linguistics, including how researchers elicit and record spoken language data. Recent projects in dialectology and variationist sociolinguistics have already demonstrated the efficacy of remote data collection using participants' personal devices, including smartphones (Hall-Lew, Cowie, Lai, Markl, McNulty, Liu, Llewellyn, Alex, Elliott, & Klingler, 2022; Leemann, Jeszenszky, Steiner, Studerus, & Messerli, 2020; Nesbitt & Watts, 2022; Sneller, Wagner, & Ye, 2022). Just as online language experiments have become a convenient and cost-effective alternative to those conducted in the laboratory, the practice of eliciting speech over the internet—including in video-mediated sociolinguistic interviews—is likely to remain popular in the future.

While the turn to video conferencing creates new opportunities for linguistic research, it also raises important theoretical and methodological questions. What do speakers do when faced with the need to communicate over video? If their speech behavior over video differs from their behavior during in-person conversation, how might we understand what motivates those changes? Thus far, research on the impact of video-mediated communication has focused on questions related to the reliability of acoustic phonetic measurements. Zhang, Jepson, Lohfink, and Arvaniti (2021) found that audio recordings from Zoom showed lower values for F1, F2, and F3 in vowels, and unexpected fluctuations in intensity, compared to analogous recordings made on a solid-state digital recording device with no file compression. Sanker, Babinski, Burns, Evans, Johns, Kim, Smith, Weber, and Bowern (2021) did not find any consistent overall differences in vowel formant frequency attributable to Zoom, though they note that there may still be meaningful effects that are vowel-specific; at the same time, the authors identify effects related to duration (shorter consonants and longer vowels, as determined by a forced aligner) and a higher signal-to-noise ratio, possibly stemming from Zoom's proprietary background noise reduction algorithm.[1] These findings should inspire caution when analyzing and interpreting raw acoustic measurements from recorded video calls, especially when relying on recordings from multiple mixed media in a single study.

As linguists continue to assess the methodological consequences of video conferencing for phonetic analysis, we must also recognize that video-mediated conversation is a *qualitatively* different experience than in-person conversation. The potential for delays, disruptions, and conversational misfires due to poor internet connectivity and hardware glitches all contribute to the perceived difficulty of communicating over video—problems that may also heighten speakers' awareness of the physical distance that separates them. These problems are exacerbated by the fact that, on most devices, interlocutors cannot make eye contact with one another (the camera is typically situated above or beside the screen) and, in most programs, including Zoom, speakers may be distracted by seeing *themselves* speaking. Even when internet connectivity is strong and devices are functioning correctly, the short electronic transmission lags that inevitably occur during video-mediated conversation have been shown to contribute to significantly longer transition times between conversational turns (Boland, Fonseca, Mermelstein, & Williamson, 2022). To compare in-person and video-mediated conversation, we must be able to appropriately theorize the impact of these different contexts on speaker behavior.

This article evaluates the impact of medium of communication (in-person versus video) on patterns of intraspeaker variation. We hypothesize that even when the discourse context and interlocutors are held constant, the transition from in-person to video-mediated conversation is likely to involve an increase in variants associated

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

with clear speech—a process we refer to as *medium-shifting*. Given the challenges of video-mediated communication outlined above, speakers may adopt compensation strategies that are meant to increase their intelligibility. Similar compensatory strategies have been posited to occur in other contexts, such as in the presence of ambient noise—a communicative phenomenon known as the "Lombard effect," which is assumed to be an automatic reflex and has been shown to involve increases in intensity, pitch, and duration (Castellanos, Benedí, & Casacuberta, 1996; Lau, 2008; Wassink, Wright, & Franklin, 2007). Beyond a concern for intelligibility, there may also be stylistic factors that promote clear or "careful" speech over video. For example, one might expect video-mediated communication to reflect a relatively higher degree of formality—perhaps a consequence of the distance between speakers—which has also been postulated to occur in telephone calls (Labov, Ash, & Boberg, 2006:36). Coupland's (1980) study of sociolinguistic variation at a travel agency in Cardiff, Wales, found that the rate of local variants was lower in telephone calls than in face-to-face communication, although that factor (which he refers to as "channel") cannot be distinguished from addressee because the telephone was only used with travel agents and tour operators. On the other hand, participants in Zoom conversations are often located in their own homes and casually dressed—factors that could contribute to more natural, intimate, and thus less formal conversations.[2]

In order to determine whether medium-shifting motivates patterns of intraspeaker variation, we analyze three different quantitative variables—articulation rate (a measure of speech tempo), density-controlled vowel space area, and (ING)—in a corpus of interviews with repeat guests from *The Late Show with Stephen Colbert*. The variables were selected because they are known to be modulated during clear speech, where "clear" refers either to intelligibility-related characteristics of speech production (slower articulation rate and larger vowel space area) or socially indexical notions of carefulness (a higher rate of the velar variant of [ING]). For each celebrity guest who was interviewed over Zoom in the early months of the COVID-19 pandemic (April 7 to June 11, 2020) and at least once before in the studio, we downloaded and transcribed both their Zoom interview and their most recent in-studio interview. This allowed us to analyze patterns of intraspeaker variation across the two mediums, while holding constant both the communicative context (a conversational interview) and the interlocutors. In addition to the guests' data, we also obtained a larger amount of data for the interviewer, Stephen Colbert, which we analyzed separately for two of the three variables.[3]

Our statistical analysis supports the hypothesis that speakers do engage in medium-shifting during the transition from in-person to video-mediated conversation. After controlling for relevant linguistic predictors as well as individual speaker-level differences, the use of video significantly favors a decrease in articulation rate and an increase in vowel space area. However, medium of communication is *not* selected as a significant predictor of variation in (ING). We interpret these findings to suggest that medium-shifting is a new contextual factor that motivates robust patterns of intraspeaker variation, particularly affecting features that are tied to speaker intelligibility. The null result of medium on (ING) does not rule out the possibility that medium-shifting is conceptually related to style-shifting, which has been argued

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

to be motivated by linguistic self-monitoring (Labov, 1972), audience design (Bell, 1984), the performance of identity and persona (Eckert, 2008), or a combination of social and cognitive factors (Sharma, 2018). However, it does suggest that the intelligibility-related effects of medium-shifting are robust enough to be seen even when there is no other evidence for stylistic differences, that is, within a single contextual style. Overall, the current study should inspire confidence that conversational interviews, including sociolinguistic interviews, can be conducted effectively with video conferencing software. However, because medium-shifting is a new source of intraspeaker variability, we additionally recommend that extra care be taken to ensure uniform methods of data collection across speakers and to avoid interpreting speaker behavior in online interviews as an accurate representation of their behavior in offline interviews.

The paper is organized as follows. First, we discuss our methodology for compiling a corpus of in-person and video-mediated interviews, as well as the suitability of these televised interviews for research on intraspeaker variation. More detailed methodological considerations, including the extraction of variable tokens and phonetic measurements, are outlined in the subsequent sections, which discuss each of our three variables. Finally, we summarize our main findings regarding medium-shifting and speaker intelligibility and offer hypotheses for future research.

Methodology

The speech data for this study come from a corpus of matched interviews conducted on the popular late-night talk show *The Late Show with Stephen Colbert*. We compiled a list of all the guests who appeared on the show in the early months of the COVID-19 pandemic (episodes 900–935, airing April 7 to June 11, 2020) and at least one time on an earlier episode that was filmed in the studio. Interviews from the pandemic period were conducted using the popular Zoom video conferencing software, as Stephen Colbert explained in a segment that compiled celebrities' "slate" outtakes used to ensure synchronized audio and video during post-production.[4] Note that we only included as "repeat guests" those who were interviewed by Colbert in solo interviews, excluding those who appeared previously only as a musical performer or in group interviews with their creative collaborators. This yielded a diverse sample of thirty-three guests, including actors, comedians, filmmakers, authors, politicians, and journalists. These guests speak a number of different English varieties from the United States and abroad; however, because our study is focused on intra- rather than interspeaker variation, these differences in language background are not expected to have an impact on our analysis.

This corpus of interviews is appealing for a few reasons. First, it allows us to analyze a considerable amount of natural speech data from the early months of the pandemic, when the public was rapidly transitioning to the use of video as a replacement for in-person communication. Second, the interviews comprise a naturalistic real-time panel study in which the interlocutors, genre, and communicative goals are held constant; our focus on repeat guests means that the pairs of interlocutors had already interacted with each other offline at least once prior to their Zoom interview. In this way, we are able to control for many of the contextual factors that are known

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

to affect style variation within individuals. Third, the use of data from publicly accessible sources online facilitates the reproducibility of our methods. Of course, relying on data from celebrities, including professional performers, means that our findings and interpretations may not necessarily extend to other groups of everyday language users. However, the conversational nature of talk show interviews, in which participants are hypercooperative and readily engage in personal narratives (Loeb, 2015), means that the results of this research can inform hypotheses for studies of other communicative contexts in which video has recently become a popular alternative to in-person interaction.

To ensure that our corpus included every portion (i.e., before and after commercial breaks) of every relevant interview (i.e., during and before the pandemic), we consulted *The Late Show*'s official YouTube channel for playlists corresponding to each episode. In some cases, portions of a single interview were aired on two different evenings, but we coded them as being part of a single interview. We then downloaded every part of the guests' Zoom and in-studio interviews as video files using the command line program *youtube-dl* (Amine & M., 2021). We also used the program *FFmpeg* (*FFmpeg* Developers, 2021) to extract the audio track of each video as a mono .wav file, with a uniform sampling frequency of 44.1 kHz.

The interviews were originally transcribed as part of a project for a graduate seminar in sociolinguistics at UC Berkeley (Linguistics 250A, fall 2020). Each episode was transcribed in *ELAN* (Max Planck Institute for Psycholinguistics, 2021) using a uniform transcription protocol, with separate tiers for the interviewer and guest, and segmented into intonation groups. To the extent possible, transcribers were advised to exclude nonspeech noise (e.g., applause from the studio audience or music) from speech segments, as these can cause problems "downstream" for forced alignment. The three authors then completed any transcriptions that were missing and manually verified and hand-corrected all transcription files for typographical and segmentation errors. Except for tokens of the (ING) variable, which were auditorily transcribed as either <-ing> (the velar variant [ɪŋ]) or <-in'> (the alveolar variant [ɪn]), all transcriptions were written in standard English orthography.

To minimize alignment errors that could affect our measurements of articulation rate and vowel space area, we removed all conversational turns from the transcriptions that showed any amount of overlap across tiers (i.e., whenever a speech segment produced by Stephen Colbert or the guest overlapped with a speech segment on the other person's tier, we removed both). This is a conservative approach to data inclusion in that it likely eliminated genuinely overlapping speech segments as well as those that only seemed to overlap due to inaccuracies in our manual segmentation. Additionally, we removed all turns that contained partial words and those transcribed as either unknown or unclear. Finally, these pruned-down transcription files and the sound files for each interview were processed with the *Montreal Forced Aligner* (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017), using its pretrained acoustic model for English and the LibriSpeech lexicon. Note that our analyses of speech tempo and vowel space area make use of the pruned-down and forced-aligned transcriptions, while the analysis of (ING) uses the original orthographic transcriptions before the removal of overlapping segments.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Articulation rate

Background

One of the features we hypothesize to be a locus of intraspeaker variation under medium-shifting is speech tempo. Since the earliest studies in variationist sociolinguistics, speech tempo has been known to vary within individuals in accordance with contextual factors. Labov (1972:95) referred to increased speech tempo as one of the "channel cues" that could be used to diagnose whether a speaker has shifted from a more careful to a more casual style, suggesting that it is impacted by topic and by the level of rapport between speaker and interviewer. A number of studies have confirmed that speech tempo differs significantly across casual and careful (or reading) styles (e.g., Eskénazi, 1992; Laan, 1997), and it is also significantly affected by accommodation to one's interlocutor (e.g., Cohen Priva, Edelist, & Gleason, 2017). Furthermore, speech tempo has been shown to correlate with perceived intelligibility, where a slower speech rate is more characteristic of clearer speech (Picheny, Durlach, & Braida, 1989; Uchanski, Choi, Braida, Reed, & Durlach, 1996). In addition to these intraspeaker differences, a faster baseline speech tempo is also characteristic of younger speakers, men, and those from particular dialect regions (Jacewicz, Fox, O'Neill, & Salmons, 2009; Verhoeven, De Pauw, & Kloots, 2004), although the size of this effect varies across studies.

Adapting the methodology of Jacewicz et al. (2009), we operationalize speech tempo in this study as *articulation rate*, calculated by dividing the number of syllables in each speech segment by its duration in seconds. Because we ran the *Montreal Forced Aligner* using an English pronunciation lexicon written in ARPAbet format, we defined the number of syllables in a segment to be the number of phonemes ending in a digit (i.e., stressed or unstressed syllable nuclei). As explained above, all overlapping turns, partial words, unknown or unclear words, and nonspeech sounds were removed prior to running the MFA, which reduces the possibility of misaligned segments and phonemic transcription errors (e.g., the false start "*s*-" would be transcribed as if it were the name of the letter, [EH1 S], and thus inflate the syllable count). Additionally, after running the MFA, we excluded any speech segment that contained an out-of-dictionary word, a filled pause (such as *um* and *eh*), or a silent segment. This should make our results more comparable to previous analyses of articulation rate using read data.

If video-mediated communication favors patterns associated with clear speech, we hypothesize that both Stephen Colbert and his guests will show a lower articulation rate (indicative of slower speech) during their Zoom interviews compared to their in-studio interviews. We expect this effect to be independent of any inherent differences in speakers' baseline articulation rates, which we control for through a model with by-speaker random effects.

Analysis

We analyzed the variation in articulation rate (in syllables per second) across all speech segments in all interviews through a linear mixed-effects regression model using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in *R* (R Core

Team, 2020). Because Stephen Colbert accounted for much more data ($n = 3{,}202$ speech segments) than any of his guests, we modeled his articulation rate separately from that of the guests ($n = 10{,}710$ segments). Our model for the guests' data included a fixed effect for medium (in-studio versus Zoom) as well as by-speaker random slopes for medium, in order to control for any inherent by-speaker variability in medium-shifting.

The model summary, presented in Table 1, shows that articulation rate is significantly lower during Zoom interviews (a predicted value of 5.18 sylls/sec; 95% confidence interval [5.04, 5.31]) compared to in-studio interviews (5.58 sylls/sec; 95% CI [5.32, 5.83]); in other words, interview speech over Zoom is approximately 7% slower than in the studio. The predicted by-speaker effect of medium is visualized in Figure 1, showing that medium-shifting affects most guests (twenty-six of thirty-three) in the same direction, that is, toward a lower articulation rate over Zoom. This suggests a relatively uniform effect of medium-shifting regardless of individuals' dialect backgrounds, but one that is not categorical and thus cannot be attributed solely to differences, for example, in recording type across Zoom and in-studio interviews. The model for Colbert also shows a significant drop in articulation rate over Zoom (5.05 sylls/sec; 95% CI [4.94, 5.16]) when compared to in-studio interviews (5.52 sylls/sec; 95% CI [5.38, 5.67]), corresponding to speech that is approximately 9% slower.[5]

The lower articulation rate over Zoom is noteworthy in light of the fact that interviews involve the very same pairs of speakers, are similar in conversational style, and assume the same target audience of television viewers. Furthermore, because we obtained this result after excluding data from overlaps and interruptions (which are more characteristic of in-studio speech and presumably involve a faster local speech tempo) and from silences (which are more characteristic of video-mediated communication and involve a slower local speech tempo), the real effect of medium-shifting may be even more dramatic than our analysis suggests.

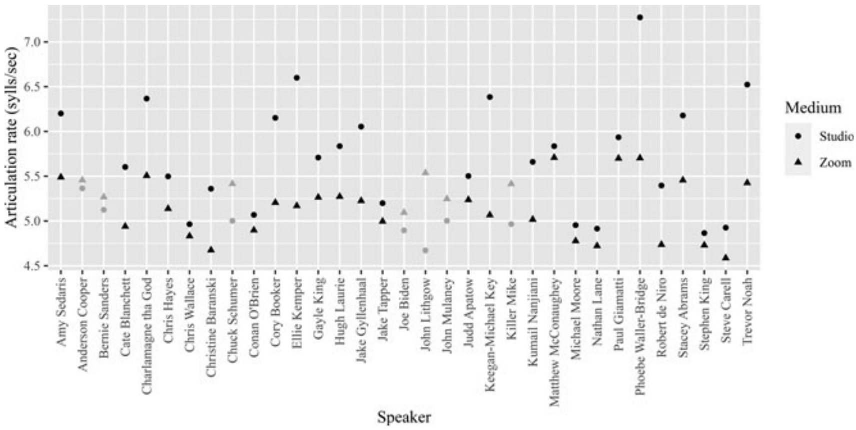


Figure 1. Predicted by-speaker random effect of medium of communication on variation in articulation rate among guests; darker points indicate speakers who conform to the direction of the overall predicted effect of medium.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Table 1. Summary of fixed effect of medium of communication from mixed-effects linear model predicting articulation rate in guests' data (n = 10,710 speech segments); rightmost column represents the mean articulation rate in the raw data for the factor level listed

	Estimate	Std. error	df	*t*-value	*p*-value		*N*	Mean
(Intercept)	5.575	0.131	32.308	42.653	<0.001	***	10,710	5.305
Medium (vs. in-studio)							3,162	5.525
Zoom	-0.396	0.125	32.326	-3.179	0.003	**	7,548	5.214

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

The predicted effect of medium-shifting obtained in this study, though significant, is considerably smaller than the effect reported in previous studies of speech rate differences across task types (e.g., 5.12 sylls/sec for casual spontaneous speech versus 3.40 sylls/sec for reading; Jacewicz et al., 2009:242). Furthermore, the reported articulation rate for casual speech falls within the range we obtained for Zoom and in-studio interviews (5.05-5.58 sylls/sec). These observations suggest that medium-shifting is a motivator of intraspeaker differences *within* a single speech style, a point to which we return below in connection with our analysis of (ING). What cannot be determined is whether the known effects of accommodation to an addressee might confound the observed effect of medium-shifting. For example, if one of the interview participants engages in medium-shifting (either the guest or Colbert), it cannot be determined whether a parallel effect for the interlocutor is due to accommodation or to medium-shifting in tandem. Because we do not yet have evidence from other speech variables about whether, and to what extent, these speakers engage in accommodation, we posit that accommodation in speech tempo might reinforce the overall effect of medium-shifting.

Vowel space area

Background

Variability in intelligibility within speakers and across different task types is a common phenomenon that has been termed "adaptive intelligibility" (McCloy, Wright, & Souza, 2012). Acoustic correlates of intelligibility include the distance between vowel categories (Neel, 2008), F1 range (Bradlow, Torretta, & Pisoni, 1996), F2 range (Hazan & Markham, 2004), vowel dispersion (Bradlow et al., 1996), the vowel space area of the polygon formed by vowel means (Neel, 2008), and the convex hull of all tokens (Luan, Wright, Ostendorf, & Levow, 2014; McCloy et al., 2012). For example, a relatively more dispersed vowel space contributes to fewer ambiguous vowel tokens, leading to higher intelligibility scores (Bradlow et al., 1996).

In line with this literature, vowel space area will be used in the present study as a proxy for intelligibility (and as another indication of clear speech), following the methodology described by Story and Bunton (2017). Vowel space area is measured with a convex hull, or the shape enclosing a given set of measurements, obtained at a specified density threshold. Here, density refers to the number of local F1/F2 pairs relative to the total number of F1/F2 measurements obtained from a speaker. In other words, density can be thought of as the relative time a speaker spends in a given part of the vowel space. As this method collects multiple formant measurements (entire formant trajectories) from individual vowel tokens to determine vowel space density, we can obtain a more accurate picture of a speaker's vowel space, and therefore intelligibility, produced over a longer duration of time.

As mentioned in the introduction, the findings of recent work on the acoustic impact of Zoom recordings are mixed. Zhang et al. (2021) compare acoustic phonetic measurements from simultaneous recordings using a recording device with external microphone, the Zoom video conferencing program, and a lossless mobile phone recording app. The researchers conclude that Zoom recordings yield lower values

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

for F1, F2, and F3, as well as fluctuations in intensity. Because vowel normalization procedures commonly make use of the third formant in vocal tract length calculations (Johnson, 2020), these distortions pose an issue for traditional methods of speaker normalization. Sanker et al. (2021), however, did not find any overall differences in formant measurements between Zoom and handheld recorders, though the authors note that the lack of significance is not indicative of reliability, but rather that different vowels are impacted in different ways.

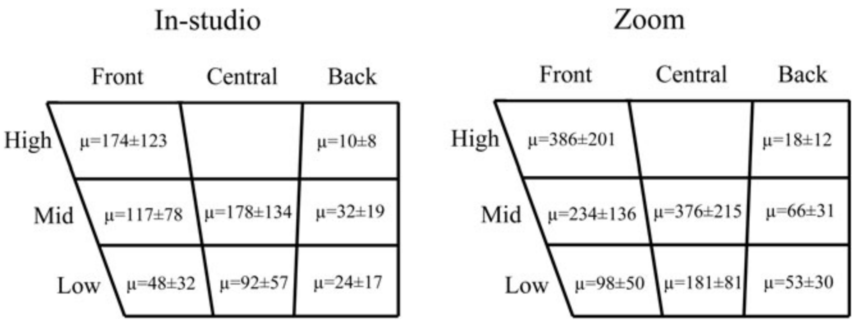
Freeman and De Decker (2021) examine acoustic measurements taken from three different video conferencing programs (Zoom, Microsoft Skype, and Microsoft Teams) and compare these measurements to values obtained from an H4n field recorder. Whereas both Skype and Teams are found to vary in formant accuracy across participant gender, Zoom was fairly accurate overall for both participant genders when compared to the recordings obtained from the field recorder. Overall, F1 and F2 values are found to be transmitted and recorded relatively faithfully in each of the three video conferencing programs, and the researchers recommend them as viable tools for phonetic analysis, especially when the focus of analysis is fairly broad (e.g., spatial arrangement of vowel categories, broad categorical determinations of mergers). Based on the results, the present study will make the assumption that comparisons between acoustic measurements taken from in-person recordings and Zoom recordings, where both are subsequently subjected to YouTube audio compression, are viable. Speaker normalization using F3 will not be undertaken, as distortions due to audio compression algorithms have been observed across this formant.

If medium-shifting affects vowel space area, we hypothesize that the perceived difficulty of communication over Zoom will result in an intelligibility-based compensation strategy, in this case realized via a larger, less centralized vowel space area compared to in-studio speech. We further expect this effect to be independent of any differences in speakers' baseline vowel space area, which we control for through a model with by-speaker random intercepts.

Methodology

We calculate vowel space area by a convex hull obtained at specified density thresholds (following Story & Bunton, 2017) for each speaker in the dataset, including Stephen Colbert, within each medium of communication.[6] F1 and F2 values were collected at 5-millisecond intervals during the course of vowel production using an automated script in Python, utilizing the Burg method with a ceiling of 5500 Hz. Data were grouped by speaker, vowel category, and medium. Following a convention in numerous computational studies (e.g., Perez & Tah, 2020; Sainis, Srivastava, & Singh, 2018), outliers were defined based on the interquartile range (IQR), or the range between the first quartile (Q1) and the third quartile (Q3); formant measurements greater than Q3 + (1.5 x IQR) or less than Q1 – (1.5 x IQR) were removed. The remaining formant values were normalized with median scaling (following Story & Bunton, 2017), whereby each formant measurement is lowered by subtracting the median formant value of that speaker and then subsequently divided by the median. After normalization, the origin of the vowel space is at the median value of the formants and the F1 and F2 ranges are roughly constrained to [-1,1] (note that this

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press



In-studio

	Front	Central	Back
High	μ=174±123		μ=10±8
Mid	μ=117±78	μ=178±134	μ=32±19
Low	μ=48±32	μ=92±57	μ=24±17

Zoom

	Front	Central	Back
High	μ=386±201		μ=18±12
Mid	μ=234±136	μ=376±215	μ=66±31
Low	μ=98±50	μ=181±81	μ=53±30

Figure 2. Means and standard deviations in guests' data for vowel class counts across medium (in-studio and Zoom interviews).

normalization method does not make use of F3).[7] Excluding Stephen Colbert, who produced far more data than the other speakers, an average of 11,585 F1/F2 measurements from an average of 705 vowel tokens were analyzed from the in-studio interviews for each speaker, and an average of 25,540 F1/F2 measurements from an average of 1,477 vowels were analyzed from the Zoom interviews for each speaker. The average distribution of vowel classes across guests is shown in Figure 2. From Colbert's speech, we extracted 146,199 measurements from 7,911 vowels from the in-studio interviews and 282,719 measurements from 15,708 vowels from the Zoom interviews. Although peripheral vowels tend to be longer (Ladefoged & Johnson, 2015:105-7), the duration of individual vowel productions is not predicted to affect the vowel space area obtained by this methodology but will be accounted for in the regression model as detailed below.

Empty grids were generated for each speaker for each medium with discretized dimensions from -1 to 1, with an increment of 0.01. A field of view of radius 0.05 was centered on each point in the grid and the number of normalized formant values falling within that field of view was calculated as the density of that grid point. The resulting local density measurements were scaled relative to the largest local density and ranged from 0 to 1. The more frequently F1/F2 pairs occurred near a grid point in a given interview, the higher the scaled density value of that grid point. Using five different density cutoffs (0.10, 0.15, 0.20, 0.25, 0.30), five different area measurements were then obtained via a convex hull using the *SciPy* library (*SciPy* Developers, 2021) where the size of the convex hull was determined based on the density of F1/F2 measurements found within the shape. For example, a density cutoff of 0.25 means that the convex hull encircles F1/F2 coordinates that have a scaled density of 0.25 or higher. The higher the density cutoff the smaller the resulting convex hull and, consequently, vowel space area. In recognition that speech tempo and vowel space area may be correlated (e.g., Tsao, Weismer, & Iqbal, 2006), where faster articulation rate may favor a smaller vowel space area, the average vowel duration for each speaker from each medium was calculated and included as a factor in the regression model. For each density cutoff, vowel space area measurements for all speakers including Colbert were submitted to linear mixed-effect regression models, with medium of communication (Zoom versus in-studio) and average vowel duration as fixed effects and speaker as a random intercept.

Results and discussion

The outputs of the mixed-effects linear regression models that predict vowel space area across medium and average vowel duration at different density cutoffs are shown in Table 2. According to Story and Bunton (2017), a scaled density cutoff of 0.25 is suggested for obtaining convex hull measurements. However, vowel space areas obtained from a range of density cutoffs were submitted to regression models in order to demonstrate that the effect of medium on vowel space area is preserved at cutoffs that are both higher and lower than 0.25 ($p < 0.05$ for all). Note that the effect of medium is significant even when the influence of speech tempo (via average vowel duration) is factored into the model. At the 0.25 cutoff, the main effect of medium corresponds to a more than 14% predicted increase in vowel space area for Zoom interviews compared to in-studio interviews.

To visualize our analysis, Figure 3 shows the vowel space areas for three speakers by medium, as heatmaps at a density cutoff of 0.25. As described above, the convex hull (the overlaid outline) is the perimeter of the vowel space area, where interior F1/ F2 coordinates have a scaled local density of at least 0.25. Visual analysis suggests that the increase of vowel space area seen in the Zoom interviews is caused by relatively more peripheral vowel productions, allowing for greater acoustic distance between phonemic vowel categories. During the in-studio interviews, however, vowel productions are more centralized and there appears to be less acoustic distance between vowel categories. Distributions in the raw data for medium for each speaker are visualized in Figure 4, showing that medium-shifting affects most speakers (twenty-seven of thirty-four) in the same direction, that is, toward a larger vowel space area over Zoom. That not all the speakers are affected in the same direction or to the same magnitude indicates that vowel dispersion is not purely driven by Zoom's auditory compression algorithm but is rather influenced by individual, and potentially agentive, speaker variability.

As changes in vowel space area, and consequently vowel dispersion, have been correlated with variable intelligibility across task types (Bradlow et al., 1996; Luan et al., 2014; McCloy et al., 2012; Neel, 2008), these results suggest speakers are compensating for a perceived communicative difficulty over Zoom. If so, we would expect to find the "clear speech" effects of medium-shifting—including a lower articulation rate and an increased vowel space area—even in the absence of other evidence that there is a salient stylistic difference between in-studio and Zoom interviews. The analysis of the third and final variable, (ING), supports this view.

(ING)

Background

Our analysis of articulation rate and vowel space area have supported the notion that medium-shifting affects intraspeaker variation by promoting the use of clear speech features over video. While there was a significant reduction in articulation rate from in-studio to Zoom interviews, both contexts yielded a predicted value consistent with the articulation rate reported in the literature for casual speech. We interpret this as support for the hypothesis that medium-shifting is a predictor of intraspeaker

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Table 2. Regression coefficients for five different mixed-effects linear models predicting vowel space area in all speakers' ($n = 34$) data, with main effects for medium and average vowel duration and a random intercept for speaker; rightmost column represents the mean vowel space area in the raw data for the factor level listed

	Estimate	Std. error	df	*t*-value	*p*-value		*N*	Mean
Density-cutoff = 0.10								
(Intercept)	4.010	0.425	47.415	9.444	<0.001	***	68	3.627
Medium (vs. in-studio)							34	3.506
Zoom	0.269	0.102	36.064	2.646	0.012	*	34	3.747
Vowel duration	−6.010	4.989	46.625	−1.204	0.234		68	0.086
Density-cutoff = 0.15								
(Intercept)	3.587	0.421	68.000	8.516	<0.001	***	68	3.325
Medium (vs. in-studio)							34	3.174
Zoom	0.325	0.103	68.000	3.160	0.002	**	34	3.476
Vowel duration	−4.919	4.948	68.000	−0.994	0.324		68	0.086
Density-cutoff = 0.20								
(Intercept)	2.990	0.396	68.000	7.544	<0.001	***	68	3.066
Medium (vs. in-studio)							34	2.886
Zoom	0.366	0.097	68.000	3.788	<0.001	***	34	3.246
Vowel duration	−1.246	4.656	68.000	−0.268	0.790		68	0.086
Density-cutoff = 0.25								
(Intercept)	2.682	0.413	48.587	6.499	<0.001	***	68	2.861
Medium (vs. in-studio)							34	2.668
Zoom	0.386	0.096	36.698	4.007	<0.001	***	34	3.054

(*Continued*)

Table 2. (*Continued.*)

	Estimate	Std. error	df	*t*-value	*p*-value		*N*	Mean
Vowel duration	−0.165	4.850	47.878	−0.034	0.973		68	0.086
Density-cutoff = 0.30								
(Intercept)	2.669	0.418	51.709	6.392	<0.001	***	68	2.695
Medium (vs. in-studio)							34	2.491
Zoom	0.418	0.089	37.261	4.680	<0.001	***	34	2.899
Vowel duration	−2.122	4.911	51.252	−0.432	0.667		68	0.086

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

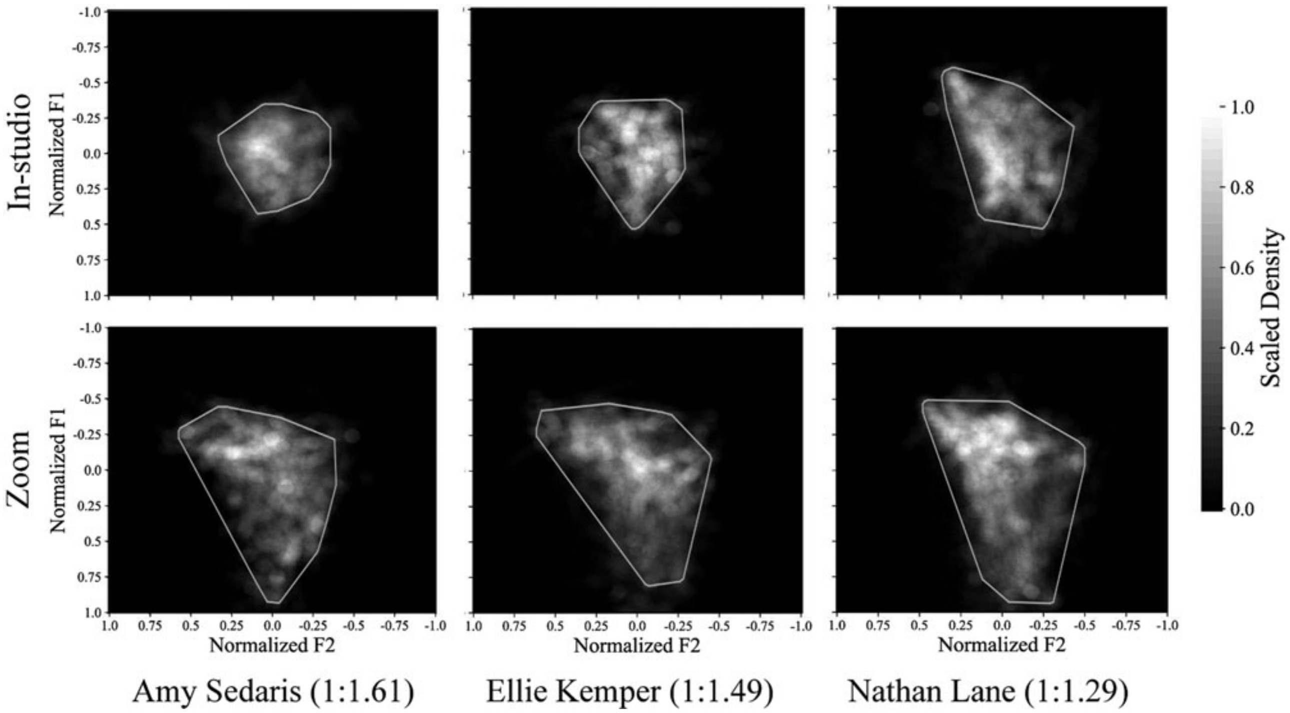


Figure 3. Heatmaps of vowel space areas at a 0.25 density cutoff with convex hull overlays, for three speakers in-studio (top row) and on Zoom (bottom row), with each speaker's areas represented as a ratio (in-studio:Zoom).

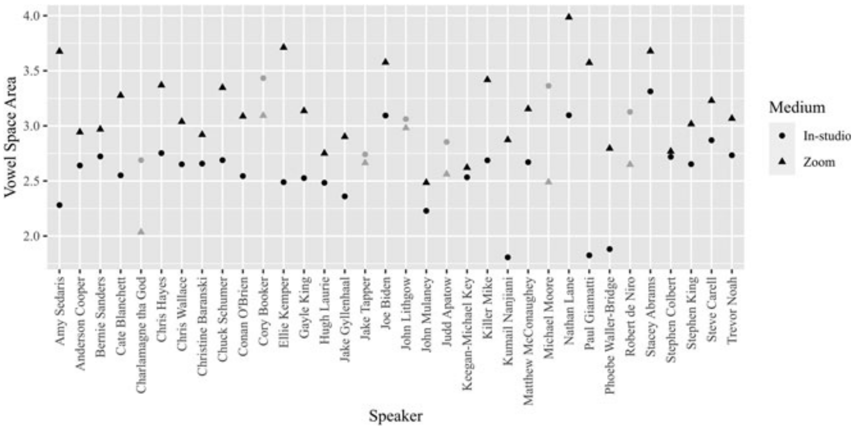


Figure 4. Distribution of vowel space area by medium of communication, based on speakers' raw data; darker points indicate speakers whose distribution conforms to the direction of the overall predicted effect of medium.

variation even within a single speech style. The scope of this article does not permit a detailed description of how in-studio and Zoom contexts might elicit different conversational styles. However, we analyze a third quantitative variable, (ING), in order to determine whether the "clear speech" effects described thus far might have a social motivation, rather than (or in addition to) an explanation rooted in intelligibility. A "staple" of variationist sociolinguistics (Hazen, 2008:117), pronunciation of word-final *-ing* (as in the progressive verb *running* or quantifier *nothing*) has been generally acknowledged to index style differences across varieties of English, including those in which the alveolar variant is especially frequent (Campbell-Kibler, 2007; Eckert, 2008; Fischer, 1958; Labov, 1966; see Hazen, 2008 for an in-depth overview). While a decrease in articulation rate and increase in vowel space area could be automatic reflexes of "clear speech," studies show that (ING) indexes more agentive and stylistically salient speaker choices. Because the velar variant [ɪŋ] is supported by standard orthography and associated with speakers of higher socioeconomic status (Hazen, 2008) and education/intelligence (Campbell-Kibler, 2009), it tends to dominate in "careful" or formal contexts, while the alveolar variant [ɪn] tends to index a "casual" or informal speaking style. A significant effect of medium of communication on (ING) variation could be taken as evidence that in-studio and Zoom interviews represent different styles. If so, the previous "clear speech" effects could similarly be viewed as diagnostics of style-shifting across seemingly similar interview contexts. There is a precedent for the view that these variables might pattern together. Kendall (2013:204) found that in a corpus of interviews with young African-American women from Washington, DC, the alveolar variant was significantly more probable during stretches of talk that exceeded the speaker's mean articulation rate by over 0.5 standard deviations. However, if medium of communication does not significantly constrain (ING), this could be interpreted as additional evidence that medium-shifting affects intraspeaker variation *within* speech styles. Our analysis supports the latter view, although we offer other interpretations for the null effect of medium on (ING).

Methodology

We extracted all tokens of word final *-ing* from the corpus of *Late Show* interviews, prior to the removal of overlapping speech segments required by our other phonetic analyses. We checked the list of all possible tokens to remove categorical items, such as transcribed descriptions like "{gesturing}" and proper names like "Keating." Following Hazen (2008), we then coded the part of speech, morphological status (suffix versus nonsuffix), and immediate environment (the following segment and word) for each token in the dataset. All tokens with a following velar (/k/ or /g/) were removed as a possible neutralization context. We also coded each token's dictionary word (i.e., the underlying lexical item, abstracting away from the particular variant produced) as well as that word's log-transformed frequency within the dataset.

Exploratory analysis revealed that part of speech and morphological status largely overlapped, and so in our analysis we rely on part of speech, which is a more informative factor. The part of speech labels we used were verb, gerund, quantifier, and other, where the latter category includes nouns and adjectives, which are known to favor the velar variant (Hazen, 2008), as well as the preposition *during*. Lexical frequency has previously been shown to correlate with (ING) variation, where lower frequency words are more often produced with the velar variant. This correlation has been posited to be socially motivated, as "the avoidance of [the alveolar variant] may result from speakers' desire for clarity due to the use of an uncommon word, therefore treating [the velar variant] as the more articulate form" (Forrest, 2017:152). Under this theory, using less frequent words could yield a relative increase in attention paid to speech, favoring use of the velar variant which is associated with a more monitored (careful) style. In addition to a main effect of frequency, our models also include an interaction term for frequency and medium in order to test if the effect of frequency (whether or not it actually encodes speech style) differs across in-studio and Zoom interviews.

We expected to find a significantly higher probability of the alveolar variant for verbs compared to the other grammatical categories. We also expected to find that lower frequency lexical items would favor the velar variant, potentially as a byproduct of speakers' adopting a more "clear" or careful style when producing lower frequency items. With regard to medium of communication, we envisioned three different possibilities: Zoom-mediated interviews might favor the use of the velar variant in light of the physical distance between interlocutors, the lack of a supportive studio audience, and the fact that speakers might feel more self-conscious when seeing themselves speak on the screen. Alternatively, the intimate nature of the conversations, the familiarity between Colbert and his guests, and the fact that the guests were usually located in their homes and dressed much more casually than during their prior interviews might contribute to a less formal context overall and thus an increase in the alveolar variant. In either scenario, the effect of medium on (ING) could be interpreted as a consequence of style-shifting. A third possibility is that Zoom and in-studio interviews do *not* differ significantly for (ING), either as a main effect or

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

in interaction with lexical frequency. If so, there would be no clear evidence that medium-shifting necessarily co-occurs with style-shifting.

Results and discussion

There were 3,655 tokens of (ING) in our dataset, after excluding pre-velar tokens. Usage of the alveolar variant [ɪn] was low for most speakers (12.1% among guests; 15.2% among Colbert), and even categorically absent for some—reflecting an adherence by the speakers in this corpus to a standard prescribing the velar variant.

As with articulation rate, we modeled the variation in (ING) separately for guests and for Colbert, who alone produced 33.3% of the total token count. Our model for guests included fixed effects for part of speech (verb, gerund, quantifier, and other), log word frequency (within the corpus), medium of communication (in-studio versus Zoom), the interaction of log word frequency and medium, as well as random intercepts for speaker and word.[8] Our model for Colbert had all of the same fixed effects and a by-word random intercept.

Summaries of the fixed effects from the two models are shown in Table 3. Note that the relevant predictors are shared across the two models, and all pattern in the same direction. Verbs are predicted to have the highest probability of the alveolar variant (for guests: 0.05; for Colbert: 0.16; probabilities back-transformed from log odds), followed by gerunds, quantifiers, and then other items. Log frequency is positively correlated with the use of the alveolar variant in both datasets (and conversely, less frequent tokens favor the velar variant). Medium does not emerge as a significant predictor in either model ($p = 0.10$ for guests; $p = 0.60$ for Colbert), which is reflected in the similar predicted probability of the alveolar variant across the two mediums as well as the large amount of overlap in the error bars (Figure 5). Finally, the interaction between log frequency and medium is also not a significant predictor of variation ($p = 0.32$ for guests; $p = 0.93$ for Colbert).[9]

The null effect of medium on (ING) could be interpreted in a number of different ways. One possibility is that most of our speakers are already performing near or at ceiling for use of the velar variant in their in-studio interviews, and thus they could not show any sizable increase when shifting to Zoom. For this reason, we also produced two alternative models using subsets of the guests' data: one excluding the fully categorical users of the velar variant, and another excluding the guests who produced five or fewer tokens of the alveolar variant. Although medium still did not emerge as a significant predictor of the variation in these other models, there simply may not be enough variability in (ING) to rule out the possibility of a true effect of medium. Additional data from other speakers, especially from those who vary more in their productions of (ING), could shed light on the role of medium as a predictor of variation.

A second possibility is that there is, in fact, no real effect of medium on (ING) variation. Because (ING) is so strongly correlated with speech style, this result would be expected if in-studio and Zoom interviews are not stylistically differentiated contexts. Under this scenario, the "clear speech" effects of medium-shifting identified for articulation rate and vowel space area would not be due to style-shifting toward more "careful" speech over Zoom. If these arise in order to improve one's

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Table 3. Regression coefficients for mixed-effects logistic models predicting the alveolar variant [ɪn] in (a) the data from guests and (b) the data from Stephen Colbert alone; rightmost column represents the mean rate of [ɪn] in the raw data for the factor level listed, if categorical

(a) **Guests** (n = 2,438)

	Estimate	Std. error	*z*-value	*p*-value		*N*	Mean
(Intercept)	−3.706	0.507	−7.31	<0.001	***	2,438	0.121
Part of speech (vs. verb)						1,206	0.176
gerund	−0.752	0.209	−3.60	<0.001	***	665	0.102
quantifier	−2.611	0.653	−4.00	<0.001	***	267	0.045
other	−2.931	0.610	−4.81	<0.001	***	300	0.013
Medium (vs. in-studio)						967	0.119
Zoom	−0.738	0.448	−1.65	0.100		1,471	0.123
Log word frequency	0.330	0.109	3.02	0.003	**	2,438	
Medium:Log word frequency	0.111	0.111	1.00	0.317		2,438	

(b) **Stephen Colbert** (n = 1,217)

	Estimate	Std. error	*z* value	*p* value		*N*	Mean
(Intercept)	−3.138	0.544	−5.77	<0.001	***	1,217	0.152
Part of speech (vs. verb)						562	0.224
gerund	−0.438	0.227	−1.93	0.054	.	349	0.146
quantifier	−2.809	0.908	−3.09	0.002	**	161	0.043
other	−3.336	1.060	−3.15	0.002	**	145	0.007
Medium (vs. in-studio)						587	0.157
Zoom	−0.275	0.527	−0.52	0.602		630	0.148
Log word frequency	0.493	0.143	3.44	<0.001	***	1,217	
Medium:Log word frequency	0.012	0.124	0.09	0.925		1,217	

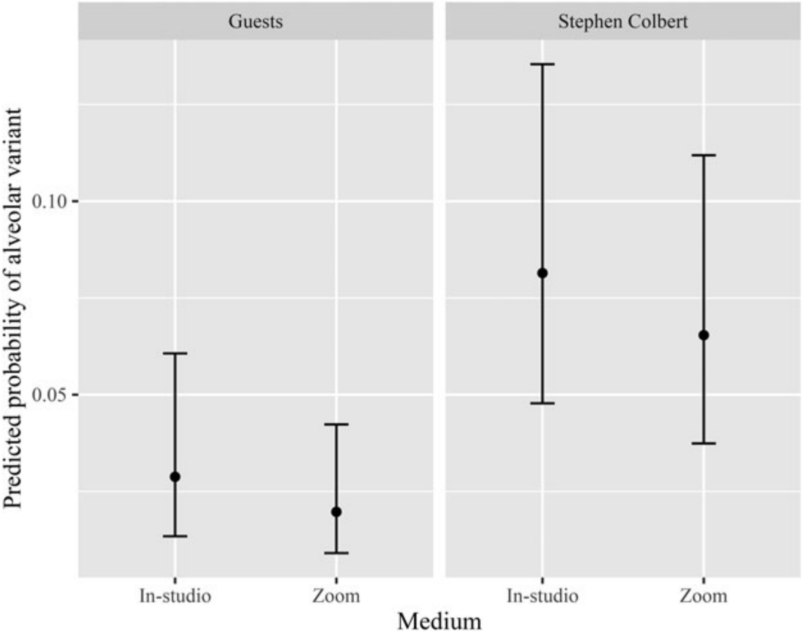


Figure 5. Predicted probability of the alveolar variant [ɪn] by medium from two statistical models (for all guests and for Stephen Colbert); the effect of medium is not significant in either model.

intelligibility when communicating over video, then medium-shifting is expected to affect intraspeaker variation regardless of style. Further research could be conducted to determine whether the effects of medium-shifting found here in a corpus of conversational (though not sociolinguistic) interviews are replicated for other speech styles, or whether they might be amplified or tempered depending on the style.

At the very least, the null result of medium for (ING) suggests that this variable is relatively uninformative for understanding the impact of medium-shifting on intraspeaker variation in conversation. The role of medium-shifting is more strongly demonstrated by our studies of articulation rate and vowel space area, where it appears to be operative even within a single speech style.

Discussion and conclusions

We investigated the effect of medium of communication (in-person versus video) on intraspeaker variation in a corpus of interviews recorded before and during the COVID-19 pandemic. We hypothesized that speakers may be motivated to produce clearer speech over Zoom due to the perceived difficulty of video-mediated compared to in-person communication. In order to test this hypothesis, we examined three variables: articulation rate (an operationalization of speech tempo), density-controlled vowel space area, and (ING). While all three variables are affected by “clear speech,” only the third variable relates primarily to socially significant stylistic choices rather than intelligibility. Our analysis showed that articulation rate was significantly lower

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

in the Zoom interviews, and density-controlled vowel space area was significantly larger. However, the analysis of (ING) production across in-studio and Zoom interviews did not find medium to be a significant predictor of variation. The analyses of articulation rate and vowel space area support the hypothesis that medium-shifting plays a role in intraspeaker variation, where speakers may be motivated to enhance their intelligibility over Zoom through more precise articulatory movements and greater contrast between phonemic vowels (see McCloy et al., 2012). The null effect of medium on (ING) suggests that these significant differences can be operational even within a single conversational style.

Although an analysis of other variables is beyond the scope of the present article, we hypothesize other features correlated with intelligibility to likewise be affected by medium-shifting, such as loudness and intonational contours. Our results complement the findings of other very recent studies that have examined the effects of video-mediated communication on common speech variables. Notably, Kang and Nycz (2021) studied the effects of medium (in-person versus video) among speakers of Korean in a spot-the-difference task. They compared peripheral vowel production and stop production and found that peripheral vowels become more peripheral, voice onset time (VOT) increases, and the pitch space of stop production is expanded in Zoom conversation. The differences in methodological approach between Kang and Nycz (2021) and the present study—including language of interaction, entertainment-oriented conversation versus speech task, presence versus absence of audience, and celebrity versus noncelebrity status of speakers—demonstrate how pervasive medium-shifting is as a general motivator of intraspeaker behavior. We expect future work to reveal additional variables that are similarly affected by medium-shifting. We also hypothesize the effects of medium-shifting to be seen across other speech events that take place either in-person or over video, such as classroom instruction or social gatherings among friends.

The data in the present study come from thirty-four people who are diverse in terms of occupation, place of origin, variety of English spoken, race, and gender. Despite the diversity seen in the speaker pool, the effects on articulation rate and density-controlled vowel space area are rather consistent, suggesting that medium-shifting is a general linguistic strategy and not one confined to US varieties of English. Although all of the interview participants from *The Late Show* are in the public eye, our results are suggestive of communicative strategies that should be further investigated with panel data from noncelebrity speakers (see Kang & Nycz, 2021).

As all the interviews analyzed were recorded within the first four months of the COVID-19 pandemic, it is important to ask whether the effect of medium-shifting on linguistic variables is a lasting consequence. It is plausible that with increased exposure to video conferencing programs like Zoom, medium-shifting will no longer influence production to the same extent, if at all. Additionally, speakers may be affected differently according to age, where younger speakers more familiar with the technology will not evidence effects of medium-shifting to the same degree as older speakers. Although the differences in articulation rate and vowel space area are statistically significant across medium of communication, they may not be perceptually salient. Accordingly, we propose the effects of medium-shifting to be local, possibly without any long-term impact on how people speak when removed from a video conferencing setting.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

As linguists increasingly rely on video-mediated communication to collect speech data, including after the pandemic, they should be aware that medium-shifting, like other sources of intraspeaker variability, can affect common speech features, including vowel acoustics. Therefore, the potential effects of medium-shifting should be considered carefully before using video conferencing platforms to collect phonetic measurements, perform certain normalization methods, or study variables that may be affected by speech tempo. Our findings about the effects of medium-shifting should inspire caution before incorporating mixed media interviews into a single variationist study, even if those interviews are otherwise similar in genre and style. This research contributes theoretically to our understanding of the range of factors that can affect intraspeaker variation. The sudden increase in the use of video conferencing technology during the COVID-19 pandemic has created new ways for language users to engage with one another in conversation; the choice of medium thus constitutes a new extralinguistic factor that affects speaker behavior and speaker choices.

Acknowledgments. We are grateful to Ernesto Gutiérrez, Aurora Martinez Kane, Yevgeniy Melguy, Ben Papadopoulos, Dakota Robinson, Daniel Schaefer, and Oliver Whitmore for collaborating with us on interview transcriptions. We would also like to thank Keith Johnson, members of the Sociolinguistics Lab at Berkeley, and audiences at New Ways of Analyzing Variation 49 and the 2022 Linguistic Society of America meeting for feedback on earlier versions of this project. All three authors were involved in designing the research and writing the manuscript. ILB processed and aligned the interview data, conducted the study of articulation rate, and analyzed the (ING) data. KC oversaw the annotation of the (ING) data. AH conducted the study of vowel space area.

Competing interests. The authors declare none.

Notes

[1] The solid-state recording device used as a baseline in these studies is the Zoom Handy Recorder (model H6 in Zhang et al., 2021; H4n in Sanker et al., 2021), which is popular among sociolinguists and phoneticians working in the field. This is not to be confused with the unrelated Zoom video conferencing program.

[2] This view is reflected in metadiscourse from a recent interview from *The Late Show with Stephen Colbert*, the show from which we obtained the raw data for this study. Fellow talk show hosts Stephen Colbert and Trevor Noah agreed that the absence of a live studio audience during Zoom interviews meant that the conversations were less "performative" (SC), "more natural" (SC), and characterized by a heightened sense of "smallness" or intimacy (TN). The interview aired on October 29, 2021 and was posted to YouTube: https://www.youtube.com/watch?v=Zf8ykuUU_Ds.

[3] For the analysis of vowel space area, each speaker contributes one data point (a measure of area) for each medium of communication. Therefore, there is no statistical motivation to separate Colbert's data from those of the guests.

[4] The compilation video of guests' "slates" aired in March 2021 and was posted to YouTube: https://www.youtube.com/watch?v=FOvc7mjDQkU. For another clip that attests to the show's use of the Zoom software, see footnote 2.

[5] This model is functionally equivalent to a t-test, as it contains just one fixed effect for medium.

[6] An implementation of our methodology is provided at https://github.com/anniehelms/vsd.

[7] Note also that meaningful differences in vowel space size are not erased with median normalization since the only true constraint on the normalized values is -1, which corresponds to the asymptotic limit of 0 Hz.

[8] We also built a more complex model with by-speaker random slopes for medium, but it resulted in a slightly worse fit to the data.

[9] We tested more parsimonious versions of the two models with the interaction term removed; this had no impact either on the direction or on the (non-)significance of the remaining factors.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

References

Amine, Remita & M., Sergey. 2021. Youtube-dl [computer program]. https://youtube-dl.org/.

Bates, Douglas, Mächler, Martin, Bolker, Benjamin M. & Walker, Steven C. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204. doi:10.1017/S004740450001037X.

Boland, Julie E., Fonseca, Pedro, Mermelstein, Ilana & Williamson, Myles. 2022. Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General* 151(6). 1272–82. doi:10.1037/xge0001150.

Bradlow, Ann R., Torretta, Gina M. & Pisoni, David B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20(3–4). 255–72. doi:10.1016/S0167-6393(96)00063-5.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82(1). 32–64. doi:10.1215/00031283-2007-002.

Campbell-Kibler, Kathryn. 2009. The nature of sociolinguistic perception. *Language Variation and Change* 21(1). 135–56. doi:10.1017/S0954394509000052.

Castellanos, Antonio, Benedí, José-Miguel & Casacuberta, Francisco. 1996. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication* 20(1–2). 23–35. doi:10.1016/S0167-6393(96)00042-8.

Cohen Priva, Uriel, Edelist, Lee & Gleason, Emily. 2017. Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *Journal of the Acoustical Society of America* 141(5). 2989–96. doi:10.1121/1.4982199.

Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9(1). 1–12. doi:10.1017/S0047404500007752.

Deflem, Mathieu. 2021. The right to teach in a hyper-digital age: Legal protections for (post-)pandemic concerns. *Society* 58(3). 204–12. doi:10.1007/s12115-021-00584-w.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–76. doi:10.1111/j.1467-9841.2008.00374.x.

Eskénazi, Maxine. 1992. Changing speech styles: Strategies in read speech and casual and careful spontaneous speech. Presentation at the International Conference on Spoken Language Processing, Banff, Canada. https://eric.ed.gov/?id=ED356511.

FFmpeg Developers. 2021. FFmpeg [computer program]. https://ffmpeg.org/.

Fischer, John L. 1958. Social influences on the choice of a linguistic variant. *Word* 14(1). 47–56. doi:10.1080/00437956.1958.11659655.

Forrest, Jon. 2017. The dynamic interaction between lexical and contextual frequency: A case study of (ING). *Language Variation and Change* 29(2). 129–56. doi:10.1017/S0954394517000072.

Freeman, Valerie & De Decker, Paul. 2021. Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *Journal of the Acoustical Society of America* 149(2). 1211–23. doi:10.1121/10.0003529.

Hall-Lew, Lauren, Cowie, Claire, Lai, Catherine, Markl, Nina, McNulty, Stephen Joseph, Liu, Shan-Jan Sarah, Llewellyn, Clare, Alex, Beatrice, Elliott, Zuzana & Klingler, Anita. 2022. The Lothian Diary Project: Sociolinguistic methods during the COVID-19 lockdown. *Linguistics Vanguard* 8(s3). 321–30. doi:10.1515/lingvan-2021-0053.

Hazan, Valerie & Markham, Duncan. 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America* 116(5). 3108–18. doi:10.1121/1.1806826.

Hazen, Kirk. 2008. (ING): A vernacular baseline for English in Appalachia. *American Speech* 83(2). 116–40. doi:10.1215/00031283-2008-008.

Hermann, Inge & Paris, Cody Morris. 2020. Digital Nomadism: The nexus of remote working and travel mobility. *Information Technology & Tourism* 22(3). 329–34. doi:10.1007/s40558-020-00188-w.

Jacewicz, Ewa, Fox, Robert A., O'Neill, Caitlin & Salmons, Joseph. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 21(2). 233–56. doi:10.1017/S0954394509990093.

Johnson, Keith. 2020. The ΔF method of vocal tract length normalization for vowels. *Laboratory Phonology* 11(1). 1–16. doi:10.5334/labphon.196.

Kang, Yoojin & Nycz, Jennifer. 2021. The sociophonetics of video-mediated vs. in-person interactions. Presentation at New Ways of Analyzing Variation 49, Austin, TX (Zoom). https://www.youtube.com/watch?v=OCDJngCL0oE.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Keesara, Sirina, Jonas, Andrea & Schulman, Kevin. 2020. Covid-19 and health care's digital revolution. *New England Journal of Medicine* 382(23). 1–3. doi:10.1056/NEJMp2005835.

Kendall, Tyler. 2013. *Speech rate, pause, and sociolinguistic variation: Studies in corpus sociophonetics*. New York: Palgrave Macmillan.

Laan, Gitta P.M. 1997. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* 22(1). 43–65. doi:10.1016/S0167-6393(97)00012-5.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William. 1972. The isolation of contextual styles. In William Labov (ed.), *Sociolinguistic patterns*, chap. 3, 70–109. Philadelphia: University of Pennsylvania Press.

Labov, William, Ash, Sharon & Boberg, Charles. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Ladefoged, Peter & Johnson, Keith. 2015. *A course in phonetics*. Stamford, CT: Cengage Learning, 7th edition.

Lau, Priscilla. 2008. The Lombard effect as a communicative phenomenon. *UC Berkeley Phonology Lab Annual Report* 4. 1–9. doi:10.5070/P719j8j0b6.

Leemann, Adrian, Jeszenszky, Péter, Steiner, Carina, Studerus, Melanie & Messerli, Jan. 2020. Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard* 6(s3). 1–16. doi:10.1515/lingvan-2020-0061.

Loeb, Laura. 2015. The celebrity talk show: Norms and practices. *Discourse, Context & Media* 10. 27–35. doi:10.1016/j.dcm.2015.05.009.

Luan, Yi, Wright, Richard, Ostendorf, Mari & Levow, Gina-Anne. 2014. Relating automatic vowel space estimates to talker intelligibility. In *Proceedings of INTERSPEECH 2014*, 2238–42. doi:10.21437/interspeech.2014-246.

Max Planck Institute for Psycholinguistics. 2021. ELAN [computer program]. https://archive.mpi.nl/tla/elan.

McAuliffe, Michael, Socolof, Michaela, Mihuc, Sarah, Wagner, Michael & Sonderegger, Morgan. 2017. Montreal Forced Aligner [computer program]. http://montrealcorpustools.github.io/Montreal-Forced-Aligner/.

McCloy, Daniel, Wright, Richard & Souza, Pamela. 2012. Modeling intrinsic intelligibility variation: Vowel-space size and structure. *Proceedings of Meetings on Acoustics* 18. 1–19. doi:10.1121/1.4870070.

Neel, Amy T. 2008. Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research* 51(3). 574–85. doi:10.1044/1092-4388(2008/041).

Nesbitt, Monica & Watts, Akiah. 2022. Socially distanced but virtually connected: Pandemic fieldwork with Black Bostonians. *Linguistics Vanguard* 8(s3). 343–52. doi:10.1515/lingvan-2021-0049.

Perez, Husein & Tah, Joseph H. M. 2020. Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using t-SNE. *Mathematics* 8(5). 1–18. doi:10.3390/math8050662.

Picheny, M. A., Durlach, N. I. & Braida, L. D. 1989. Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research* 32(3). 600–603. doi:10.1044/jshr.3203.600.

R Core Team. 2020. R: A language and environment for statistical computing [computer program]. https://www.R-project.org/.

Sainis, Nachiket, Srivastava, Durgesh & Singh, Rajeshwar. 2018. Feature classification and outlier detection to increased accuracy in intrusion detection system. *International Journal of Applied Engineering Research* 13(10). 7249–55.

Sanker, Chelsea, Babinski, Sarah, Burns, Roslyn, Evans, Marisha, Johns, Jeremy, Kim, Juhyae, Smith, Slater, Weber, Natalie & Bowern, Claire. 2021. (Don't) try this at home! The effects of recording devices and software on phonetic analysis. *Language* 97(4). e360–e382. doi:10.1353/lan.2021.0075.

SciPy Developers. 2021. SciPy [computer program]. https://scipy.org/.

Sharma, Devyani. 2018. Style dominance: Attention, audience, and the 'real me.' *Language in Society* 47(1). 1–31. doi:10.1017/S0047404517000835.

Sneller, Betsy, Wagner, Suzanne Evans & Ye, Yongqing. 2022. MI Diaries: Ethical and practical challenges. *Linguistics Vanguard* 8(s3) 307–19. doi:10.1515/lingvan-2021-0051.

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press

Story, Brad H. & Bunton, Kate. 2017. Vowel space density as an indicator of speech performance. *Journal of the Acoustical Society of America* 141(5). EL458–EL464. doi:10.1121/1.4983342.

Tsao, Ying-Chiao, Weismer, Gary & Iqbal, Kamran. 2006. The effect of intertalker speech rate variation on acoustic vowel space. *Journal of the Acoustical Society of America* 119(2). 1074–82. doi:10.1121/1.2149774.

Uchanski, Rosalie M., Choi, Sunkyung S., Braida, Louis D., Reed, Charlotte M. & Durlach, Nathaniel I. 1996. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research* 39(3). 494–509. doi:10.1044/jshr.3903.494.

Verhoeven, Jo, De Pauw, Guy & Kloots, Hanne. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3). 297–308. doi:10.1177/00238309040470030401.

Wassink, Alicia Beckford, Wright, Richard A. & Franklin, Amber D. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35(3). 363–79. doi:10.1016/j.wocn.2006.07.002.

Zhang, Cong, Jepson, Kathleen, Lohfink, Georg & Arvaniti, Amalia. 2021. Comparing acoustic analyses of speech data collected remotely. *Journal of the Acoustical Society of America* 149(6). 3910–3916. doi:10.1121/10.0005132.

Cite this article: Bleaman IL, Cugno K, Helms A (2022). Medium-shifting and intraspeaker variation in conversational interviews. *Language Variation and Change* **34**, 305–329. https://doi.org/10.1017/S0954394522000151

https://doi.org/10.1017/S0954394522000151 Published online by Cambridge University Press