LIFESPANS & STYLES
Undergraduate Working Papers on Intraspeaker Variation

Volume 3, Issue 1 Article 4

2017

Style-Shifting in Vlogging: An Acoustic Analysis of "YouTube Voice"

Sarah Lee
sarah.l.lee0@gmail.com

ISSN: 2057-1720 doi: 10.2218/ls.v3i1.2017.1826

This paper is available at: http://journals.ed.ac.uk/lifespansstyles
Hosted by The University of Edinburgh Journal Hosting Service: http://journals.ed.ac.uk/

Style-Shifting in Vlogging: An Acoustic Analysis of "YouTube Voice"

Sarah Lee

This article demonstrates evidence of context-related style-shifting on YouTube. This was achieved by comparing the vowels of vlogger Phil Lester in multiple contexts (a solo vlog, collaborative vlog, gaming video, and live video). A mixed-model regression found significant differences between the more scripted solo vlog context and the less scripted gaming and live videos: in the former, Lester's FOOT/STRUT merger was reduced, and in the latter he showed considerable variability in his employment of the TRAP/BATH split. It is argued that this results from greater attention paid to level aspects of his Northern accent for an international audience.

1 Introduction

Past sociolinguistic research on language use online revolved around the use of text-based communication, compensating for the "limited presence […] of spoken language" on the Internet (Crystal 2001:9). Much of this research examines the construction of identity, as the Web's anonymous nature provides near-complete freedom in how users present themselves online (Rheingold 2000). As multimedia websites such as YouTube have become more popular, acoustic data on the Internet have also become more accessible. This side of the Internet represents a new context for investigating sociolinguistic theories such as intraspeaker variation. As such, the popularity of videoblogs ("vlogs")—and the vloggers that create them—present significant implications for research into these areas.

Vlogging typically features a single person speaking to a camera about a range of topics, including personal topics or those relating to the wider world (e.g., VlogBrothers 2008). Many vloggers have uploaded hundreds of videos of themselves speaking in the past decade, representing a sizeable corpus for both diachronic and synchronic research. As the medium has grown, it has also brought its own apparent linguistic features. Naomi Baron suggested a set of characteristics common to vloggers, including epenthesis and overstressed vowels, dubbed "YouTube voice" (discussed in Beck 2015, para. 7–11). Its purpose was suggested to be performative and to attract more viewers. This fits with Bell's (1992) theory of audience design, where individuals adjust their speech based on their perceived audience. As far as I know, there is no acoustic study of this claim about YouTube.

This paper analyses acoustic variation among the monophthongal vowels of one Northern British vlogger, Phil Lester. Videos were compared to see if style-shifting occurred between four contexts: a solo vlog, a collaborative vlog, a gaming video, and a live video. These were chosen on the assumption that they differ in level of spontaneity based on whether they feature scripted content or unscripted reactions to unexpected events. Consequently, the solo vlog was deemed least spontaneous, while the live video was deemed the most spontaneous. This paper presents a mixed-model regression analysis. The results indicate significant differences between contexts, particularly between the solo vlog and the live video. The solo vlog and collaborative vlogs were the most similar to each other, as were the gaming video and live video, indicating that style-shifting occurred due to context change rather than as a result of audience or number of interlocutors.

I investigate the potential explanations for this style-shifting by considering audience design (Bell 1984), speaker design (Schilling-Estes 1998), and attention paid to speech (Labov 1966). For reasons explored below, it is unlikely that the audience of each context differs enough to provoke a style shift, meaning that audience design does not fully explain this behaviour. Speaker design also fails to explain the observed style-shifting as being part of a performative register to evoke a persona of some kind: a movement towards RP in Lester's solo vlogs for this purpose is implausible based on perceptions of this dialect as "posh" or unlikeable (Garrett 2010:627).

Labov's (1966) description of style-shifting as a result of attention paid to speech, on the other hand, is able to explain these observations. Comparing Lester's vowel data to formant data from speakers of RP, Standard Southern British English (SSBE), and Lancashire English did not show consistent shifts between Northern vowels and a system closer to that of RP/SSBE in the solo vlog (Hawkins and Midgley 2005, Ferragne and Pellegrino 2010). However, it was shown that Lester merged his FOOT/STRUT vowels considerably more in the live context. Similarly, although Lester showed evidence of a TRAP/BATH split in all contexts, he did not employ it in some linguistic contexts that were to be expected in Southern British accents, and these vowels were far more variable in the live context. Because these changes were reversed in less "spontaneous" contexts (e.g., the solo vlog videos), I argue that this is suggestive of a style-related change resulting from Lester paying closer attention to his speech when vlogging.

2 Background

Vlogs, described by Moor et al. (2010:1536) as the "video version of text-based weblogs" appear on first examination to be no different from existing media. Content-wise they may be scripted, involving acted skits akin to a television programme, or present a monologue directed to a single camera that more closely resembles a news broadcast. Despite these similarities, there exists a major difference distinguishing vlogs from other audio-visual media: typically, only a single individual writes, acts, and edits their own vlog. Additionally, whatever the format, the speaker is always observed by a distant, unknown audience "who does not take an active part in the spoken interaction" (Frobenius 2011:815). Consequently, the role of the individual and their relationship with this audience are far more important in this medium than in others.

This kind of intimate relationship between creator and audience has been present since the inception of vlogging, when technological limitations meant early vlogs during the mid-2000s were little more than teenagers speaking to webcams in their bedrooms. The earliest video from Phil Lester, for example, features precisely this (AmazingPhil 2006). Yet far from being a drawback, this one-on-one approach simulated the "interpersonal engagement of eye contact in face-to-face conversation" (Hutchby 2014:1). Small viewer numbers also increased interactivity of the videos, whether through responses to comments left on the video or a follow-up video response. The geographic and temporal distance was easily overcome in this way, allowing viewers to have an "authentic" and "intimate" experience (Ault 2014). This perception of authenticity is no doubt a large part of its appeal. Nonetheless, there seems to be a growing trend among vloggers to actively distance themselves from the persona they project online, instead opting to create exaggerated caricatures of themselves. An example is the gaming vlogger, Felix Kjellberg (Pewdiepie), who explains that he is "himself" in the videos but with "100 percent energy" (Shields 2013, para. 7). As vloggers risk a backlash from fans and sponsors if they make a "boring self-revelation […] that fails to move the reader" (Behar 1996, quoted in Lange 2007:1), this is unsurprising. A wrong move may lead to fallout from their subscribers or financial sponsors. At times, vlogging appears to require the careful balancing of keeping the "real self" distant and private, all the while maintaining a sense of authenticity.

Significantly, an entire linguistic style specific to YouTube appears recognisable even without subtle acoustic analysis—suggested features include the use of overstressed vowels and consonants, epenthesis, and aspiration. The motivation for any kind of intraspeaker linguistic variation may be a variety of factors, including sociodemographic characteristics, formality, and the level of integration with social groups (Barbu et al. 2013). In the case of vlogging, these features have been likened to an "intellectual used-car-salesman" (Beck 2015: para. 21). According to Frobenius (2014), its purpose is to attract more viewers, such that the YouTube Voice arises as a result of audience design (Bell 1984). Under this theory, vlogging audiences are "referees", or "third persons not physically present at an interaction but possessing such salience for a speaker that they influence language choice even in their absence" (Bell 1992:33). For example, research by Montgomery (1988) showed that British DJs accommodate for the absence of their referees with techniques normally used in face-to-face conversation such as greetings or interrogatives. Similar techniques are present in vlogs, exemplified in DeFranco (Philip DeFranco 2015) and Tran (communitychannel 2011), who open their videos with greetings and respond to individual comments left on their videos.

Another example of audience design theory concerns the modification of marked regional accents. Considering the international nature of YouTube, it would not be surprising to see some kind of levelling of the more marked features of Lester's Northern accent in some contexts. One survey shows that Lester's fans consist of approximately 45% American viewers while only 21% are from the United Kingdom; the rest are from other countries (Dan and Phil Survey 2015). Potentially, Lester may show deliberate reduction or exaggeration of certain features in an effort to remain intelligible and accessible to the greatest number of viewers. Conversely, this may lead to such changes being less present or even reversed entirely when Lester relaxes this focus on his speech. Based on this, the monophthongs used in this analysis are those relevant to differentiating Northern and Southern British accents, particularly the TRAP-BATH split and the FOOT-STRUT merger (Wells 1982).

However, to state that audience design is the only factor at play here may be an oversimplification and neglects the speaker's agency, an obviously important aspect in a media context that offers so much individual creative control. Furthermore, this argument is weakened in the face of context-related style-shifting for Lester, as it is debatable whether the audience differs enough between contexts to elicit intraspeaker variation. For example, Lester's collaborative vlogs are hosted on the same channel as his solo vlogs, while his gaming videos are uploaded to a different channel (created in 2014) and advertised from his solo vlogging channel. His live videos are quite different from the vlogging and gaming; they are generally hosted on the website YouNow, although in recent times Lester has begun livestreaming directly from his solo vlogging YouTube channel (LessAmazingPhil 2017). Live videos primarily consist of Lester answering questions from Twitter fans or those submitted in the chatroom that appears alongside the live video. Despite these differences, suggesting that there are substantial differences in audiences is dubious, as the biggest draw is arguably Lester himself, with fans watching these videos in supplement to his "main" solo vlogging channel. The recent appearance of live

videos on Lester's main account may cause differences in later years, but for the period of analysis (2015), the platform is kept constant enough that it is unlikely to be the primary motivation for style-shifting.

An alternative explanation is attention to speech. Labov (1984:3) argues that more "casual" speech styles emerge when the minimum amount of attention is paid to speech. Otherwise, "careful" speech arises, often leading to increased usage of a prestigious linguistic variant. This was shown in Labov's (1966) department store study on the production of final or preconsonantal /r/: the perceived prestige of this variant meant usage increased in more emphatic speech styles. For YouTube, casual/careful speech distinctions seem to arise directly from the affordances of each video context. For example, the solo vlog analysed in this paper features little spontaneity in the sense of Lester reacting to external events; in contrast, a gaming video requires a reaction to the game footage in real time. Likewise, a live video features the YouTuber reacting in real time to events for the audience to watch—there is no editing in post-production. If Lester shows evidence of attention-based style-shifting, we would expect to see an increased usage of prestige variants in less spontaneous contexts: specifically, his vowels may become more similar to those of RP, which has been suggested to be the main prestige variety in Britain (Garrett et al. 2003). However, categorising speech in YouTube videos as casual or careful is much more ambiguous than in a typical sociolinguistic interview, particularly as previous literature has often relied on the presence of a narrative as an indicator of casual speech. As vlogs often feature scripted narrative, applying this criterion is potentially misleading. Furthermore, the casual/careful speech model often relies on the formality of the situation; however, the previously stated perception of vlogs as "intimate" suggests none of the videos studied in this paper can be categorised as especially formal.

A final model which incorporates the speaker's intentions and yet also considers the audience is speaker design—the use of certain linguistic styles for a particular communicative purpose. Schilling-Estes (1998), for example, identified a "performative register" among North Carolina Ocracoke speakers who highlighted aspects of their relic dialect by raising the nucleus of their /ay/ diphthong. More recently, Coupland (2001) showed DJs in Cardiff using different local dialects based on the content of their jokes, as an attempt to project group membership and their own personal identity. The speaker design perspective affords more agency to the speaker and accounts for the more performative aspect of YouTube, taking it to be the result of Lester projecting a "persona similar to that of [his] interlocutors" (Coupland 1984:65). Notably, there are issues with this approach too. Given that the YouTube audience is generally anonymous, and as mentioned, very international, it is initially unclear exactly what speech community Lester might align himself with. One possibility is that if Lester showed increased usage of more Northern-like vowels in his scripted vlogging videos (as opposed to the live videos), he may be seeking to emphasise this aspect of his identity to appear more authentic and likeable (Garrett et al. 2003).

In the following analysis, I demonstrate how video context correlates significantly with Lester's vowel production, indicating context-based style-shifting. Secondly, I compare Lester's formant values to data from RP, SSBE, and Lancashire English speakers (Hawkins and Midgley 2005, Ferragne and Pellegrino 2010). I then evaluate which of the three intraspeaker variation models may explain Lester's behaviour best: audience design, an attempt to project some kind of persona for performative reasons, or a movement towards increased use of prestige forms based on attention paid to speech.

3 Methodology

3.1 Phil Lester: A Case Study

The subject of this paper is Phil Lester (known as "AmazingPhil" on YouTube). Born in 1987 in Rawtenstall, Lancashire, he lived in London from 2012 to the time of data collection (2015). He completed his undergraduate and postgraduate degrees in York from 2005 to 2009. While he does not top the list of the most subscribed British YouTubers, he is one of its most enduring, amassing more than three million subscribers to his channel, which has featured as many as two hundred videos since his debut in 2006 (VidStatsX 2016). Lester's longevity makes him a particularly useful candidate for this study, remaining prolific since his first video in 2006 with an output of several videos a month in 2016. This is in contrast to other British vloggers who have more subscribers but were established more recently (see Zoella [SocialBlade 2016a], danisnotonfire [SocialBlade 2016b]), or those who saw early popularity but have low video output (e.g., crabstickz [SocialBlade 2016c] or charlieissocoollike [SocialBlade 2016d]), having uploaded only three to four videos as of March 2016. As Lester has been on YouTube almost since its inception, this may mean that any kind of style-shifting or linguistic habits related to vlogging will be far more established in his speech than that of newer vloggers.

Although the AmazingPhil channel heavily features solo vlogs, Lester has ventured into other projects as his popularity has grown. These include a radio show on BBC Radio 1, which he hosts along with fellow YouTuber and housemate Dan Howell (danisnotonfire), a collaborative gaming YouTube channel (DanandPhilGAMES), and live shows/recordings.

3.2 Context Categorisation

Four contexts were used in this study, described and ordered from "most spontaneous" to "least spontaneous". As previously discussed, categorisation of contexts as more or less scripted resulted from the different affordances available to each context. These included whether Lester was spontaneously reacting to an event, whether the video featured post-production (i.e., if it was edited before uploading), and if the opportunity was available for Lester to reshoot the scene. As such, the contexts are as follows, from most to least spontaneous: live video, gaming video, collaborative vlog, and solo vlog.

Lester was living in London during all the videos used in this paper. All videos were chosen by virtue of being the closest upload in each context to the publication of the solo vlog.

3.2.1 Solo Vlog (14th December 2015; 6m 23s)

This vlog features Lester on his own (AmazingPhil 2015a) and was the most recent video at the time of data collection. It fits several of the "traditional" hallmarks of a vlog: Lester constructs a narrative about a past event, speaking to a single camera. It is heavily edited, and it is impossible to tell how many times Lester may have filmed these scenes. As such, it is argued to be the least spontaneous context.

3.2.2 Collaborative Vlog (29 November 2015; 7m 8s)

This video (AmazingPhil 2015b) shows Lester collaborating with Howell, his housemate and fellow vlogger. Similar to a typical solo vlog, both speakers address the camera directly, and the video is edited extensively in post-production. However, in this video Lester must react to things said by his collaborator, so it is a conversation with the other person as much as with the referees. Howell and Lester interact with each other and their referees by answering questions sent to them on Twitter.

3.2.3 Gaming Video (8 February 2015; 11m 29s)

A gaming video, or "Let's Play", requires reactions in real time to gameplay footage. Here, Lester (DanAndPhilGAMES 2015) provides commentary as he plays a video game, meaning he reacts to previously unseen events occurring in the game. Although still edited, this is done far more sparingly to prevent missing game footage.

3.2.4 Live Video (6 December 2015; 10m)

In this video (PhanShows 2015), Lester is recording himself live, reading and responding to messages sent in by his viewers in a chatroom. He has no control over the messages that are sent in and the footage is completely unedited and broadcast in real time. There is therefore no delay between production and reception. Only the first ten minutes of the full 46-minute video were used for analysis for comparability with the other videos.

3.3 Procedure

The list of studied monophthongs included TRAP, BATH, STRUT, and FOOT. For ease of description, each vowel is referred to based on Wells's (1982) lexical sets representing pronunciation of the stressed syllable in RP English, e.g., TRAP for /æ/.

In acoustic studies, the first two formants of vowels can be used to describe vowel quality, in particular, the shape and position of the tongue: F1 with height and F2 with backing (Ladefoged and Johnson 2014). To achieve an analysis of F1 and F2 for each lexical set, each video was annotated using ELAN (Sloetjes and Wittenburg 2008) and formant values extracted automatically using FAVE (Rosenfelder et al. 2011). All default settings were used for FAVE-EXTRACT; however, measurements were taken at the vowel's midpoint (50%). Following automatic alignment, the output was hand-checked for mistakes such as errors in vowel attribution; for example, schwas are transcribed the same way as STRUT and need to be excluded from an analysis of the FOOT/STRUT contrast. Additionally, as FAVE does not differentiate between the TRAP and BATH vowels (because it is based on an American English model), these were re-coded manually. In both the longitudinal and cross-contextual comparison, vowel data were left unnormalized: it was deemed unnecessary as normalization is usually used to eliminate variation caused by physiological issues, and Lester is unlikely to have undergone physiological changes during the studied period (Disner 1980).

A mixed-model regression was performed on the resulting formant data using the *lme4* package in R (Bates et al. 2015, R Core Team 2016), testing whether the contexts had a significant effect on F1 and F2, which were

tested separately. The effect of "word" was also tested to see if this impacted the results. The solo video was used as a reference level to which all other contexts were compared.

4 Results

Figures 1 and 2 show F1 and F2 for each vowel across the four video contexts. The descriptive statistics and the results of the mixed-model regression are given in Table 1.

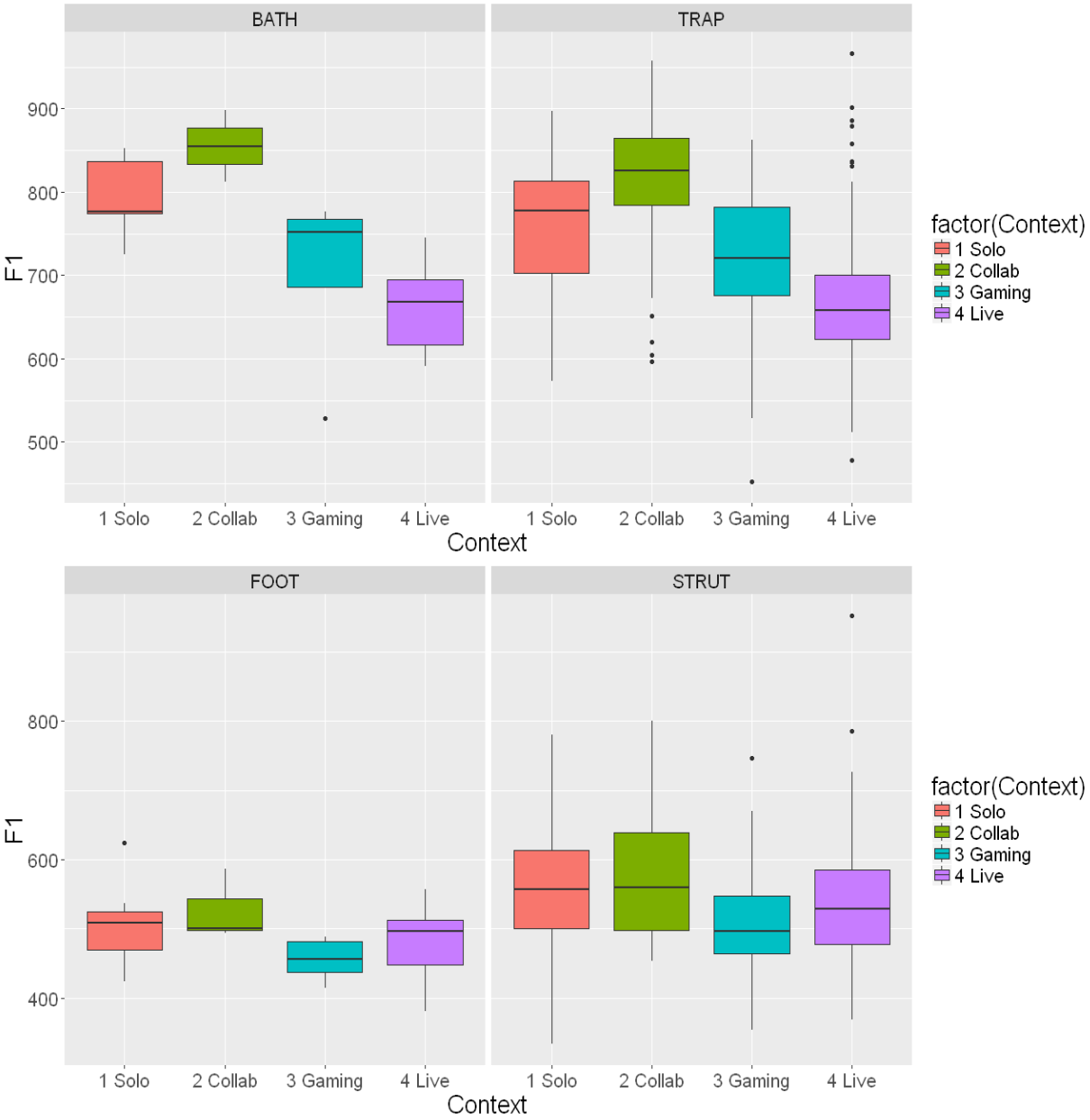

Figure 1: F1 by vowel and video context.

Figures 1 and 2 show the overall differences in vowel quality that are expected for a typical speaker of English. For example, TRAP and BATH are consistently the lowest vowels, as evidenced by their high F1 values. With respect to the effect of video contexts, the solo and collaborative vlogs exhibit the most similar patterns for most vowels in both F1 and F2. Indeed, only the F1 of TRAP showed any significant differences between the solo and collaborative contexts. The gaming and live contexts also appear to have similar distribution to one another, albeit to a lesser extent. For the majority of vowels, significant differences were found for both contexts in comparison to the solo vlog, although this primarily occurred in F1 (Table 1). F1 was lowest in these contexts for FOOT and STRUT, meaning Lester pronounced these higher in the gaming and live

videos than in the vlogs. For F2, vowels were consistently higher (with the exception of BATH) in the live videos, suggesting overall fronting of the vowel space in this least-scripted of contexts. In F2, the range of variation is especially broad in the live context compared to any other, despite similar median values, which suggests some inconsistency in pronunciation within this context. In other words, the F2 for all four monophthongs appears to be very consistent within the collaborative, solo, and gaming vlogs, but is highly variable when Lester is in a live video.

Individual vowels also varied based on context. TRAP showed large amounts of variation for F1 across contexts, displaying both a highly variable median value and also a fairly large range in each context. Its F2, on the other hand, stayed reasonably constant in all but the live context. Significant differences were found in all contexts for F1 compared to the solo vlog. Variation was also present for Lester's BATH vowel in F1 but not F2; the only significant difference found was in the F1 of this vowel in the live context. For STRUT, Lester showed significant raising in both the gaming and live contexts compared to the solo vlog, and significant fronting in the live context. Significant differences for Lester's FOOT vowel, on the other hand, were only shown for F1 in the gaming context (raising) and in the live context for F2 (fronting).

In sum, the overall effect on these differences indicates possible style-shifting between contexts, manifested by vowel raising in both live and gaming contexts, and fronting in the live contexts.

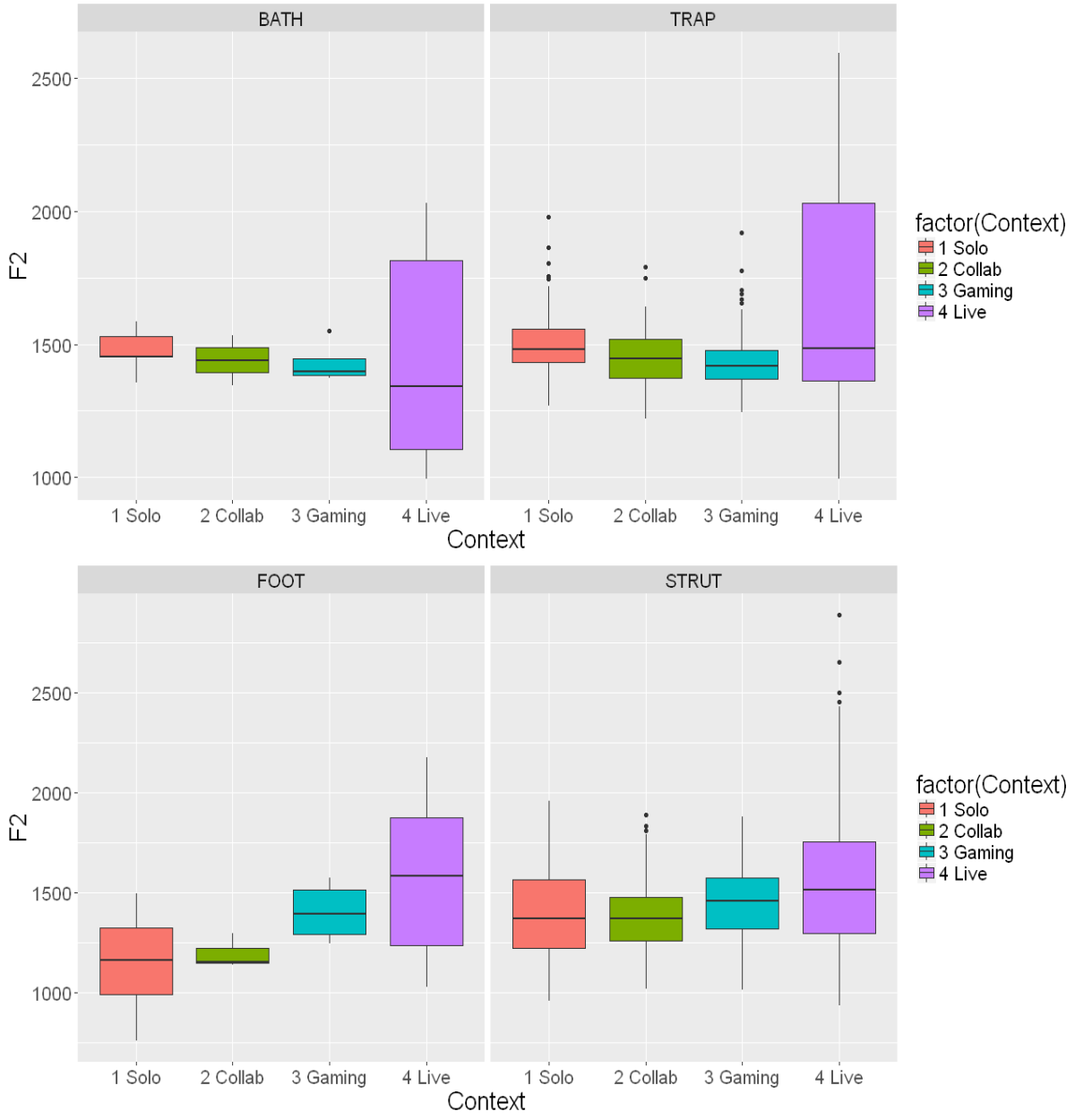

Figure 2: F2 by vowel and video context.

Table 1: Mean F1 and F2 values and results of mixed linear regression comparing each video context to the Solo vlog context

Vowel	Solo vlog (Hz; *N*)	Collab vlog (Hz; *N*)	Gaming video (Hz; *N*)	Live video (Hz; *N*)	df	*N*
F1						
TRAP	755 (58)	801 (29)	725 (63)	672 (116)		266
p =		0.016	0.001	< 0.001	3	
BATH	793 (5)	855 (2)	702 (4)	663 (6)		17
p =		0.14	0.07	0.013	3	
FOOT	501 (7)	527 (3)	456 (7)	503 (8)		25
p =		0.41	0.02	0.19	3	
STRUT	609 (33)	627 (20)	548 (30)	569 (74)		157
p =		0.41	<0.001	<0.002	3	
F2						
TRAP	1507	1467	1449	1689		266
p =		0.7	0.4	<0.001	3	
BATH	1475	1440 (2)	1431	1452		17
p =		0.8	0.8	0.9	3	
FOOT	1303	1197	1404	1259		25
p =		0.5	0.18	0.002	3	
STRUT	1256	1272	1328	1322		157
p =		0.9	1.88	<0.001	3	

Using formant data from Hawkins and Midgley (2005) and Ferragne and Pellegrino (2010), I compare Lester's data to those of RP speakers (Table 2) and SSBE (Table 3), taken from speakers in London, as well as to those of Lancashire (Northern) English speakers (Table 4). Notably, Ferragne and Pellegrino's (2010) research conflates RP and SSBE, but enough differences emerge between their data and Hawkins and Midgley (2005) to warrant a separate comparison. These include more fronting in FOOT and STRUT, and a lower TRAP compared to RP accents. For Lancashire English accents, of note are the extremely similar values for FOOT/STRUT and also TRAP/BATH in F1, representing the merging of these vowels, respectively.

Table 2: Average formant values for RP speakers in the 20–25 age group; Hawkins and Midgley (2005)

	TRAP	**BATH**	**FOOT**	**STRUT**
F1	971	604	413	658
F2	1473	1040	1285	1208

Table 3: Average formant values for SSBE speakers; Ferragne and Pellegrino (2010)

	TRAP	**BATH**	**FOOT**	**STRUT**
F1	751	655	397	623
F2	1558	1044	1550	1370

Table 4: Average formant values for Lancashire; Ferragne and Pellegrino (2010)

	TRAP	**BATH**	**FOOT**	**STRUT**
F1	697	689	483	485
F2	1454	1112	1144	1130

Comparing the above formant values to those of Lester shows that he does not converge on any particular variety in a given context (Table 5). For example, the F1 of FOOT was closest to that of RP speakers in the live context, but the F1 of STRUT was closest to Lancashire speakers in this same context. In other words, Lester does not overwhelmingly use Lancashire English variants in the gaming and live contexts and SSBE/RP in the vlogging contexts. Many vowels were also closer to one variety in F1 and another for F2, such as the F1 of TRAP being closest to Lancashire but closest to SSBE for F2 in the live context.

Given that the FOOT-STRUT merger is a hallmark of Northern accents and can be clearly seen in the data from Lancashire accents, we would likewise expect more coalescence in more spontaneous/casual contexts. This prediction appears to have been upheld to some degree, as these two vowels are far closer in F1 in the live context compared to any other. This is not completely consistent, given that the F2s of FOOT and STRUT do not show as much coalescence. As mentioned, however, the range of values for FOOT and STRUT is largest in the live context compared to all others in F2, potentially explaining this discrepancy. In sum, it seems reasonable to cautiously conclude that Lester shows more evidence of the FOOT/STRUT merger in more spontaneous contexts. This is reversed in less spontaneous contexts, manifested through more SSBE-like F1 values for STRUT.

TRAP/BATH, on the other hand, is less consistent. Lester appears to have a significantly more fronted BATH vowel than any of these three varieties, with F2 values ranging from 1431Hz to 1475Hz across contexts. Indeed, these values are considerably close to those of his TRAP vowel, especially for F1—another feature noted as distinctive for Lancashire according to Ferragne and Pellegrino's data based on the lack of split between these two vowels. Overall, Lester shows far less variability in these two vowels across contexts than FOOT/STRUT.

Table 5: Comparison of Lester's formant values to RP, SSBE, and Lancashire English varieties

	TRAP (F1/F2)		BATH (F1/F2)	FOOT (F1/F2)		STRUT (F1/F2)	
Solo	SSBE		Lanc	Lanc	RP	SSBE	RP
Collab	SSBE	RP/Lanc	Lanc	Lanc		SSBE	RP
Gaming	SSBE/Lanc	RP/Lanc	Lanc	Lanc	SSBE	Lanc	SSBE
Live	Lanc	SSBE	Lanc	RP		Lanc	SSBE

5 Discussion

In this analysis of Phil Lester's speech across four different video contexts, we have seen that the context of the video is more predictive than the number of interlocutors in the video. For example, the collaborative vlog contexts only showed significant differences from the solo vlog contexts for the F1 of TRAP, while far more vowels showed significant differences between the solo vlog and the gaming contexts. Nearly all were significantly different between the solo vlog and the live contexts. This mainly manifested in an increase in F1 and F2, indicating that Lester lowered and fronted many of his vowels in the live contexts.

A comparison of Lester's data to those of different varieties of English from Hawkins and Midgley (2005) and Ferragne and Pellegrino (2010) revealed that Lester did not show a consistent movement towards RP, SSBE, or Lancashire English-like formant values in any context. Specifically, we might have expected Lester to consistently use RP or SSBE-like variants in the less spontaneous contexts; this was not the case, as Lester did not converge on any particular variety in a particular context.

On the other hand, it was demonstrated that FOOT and STRUT vowels showed most coalescence in F1 for the live contexts, and this trend was reversed in other contexts. TRAP and BATH, however, did not show the expected pattern of being more distinguished in the less spontaneous contexts, which would indicate alignment with RP/SSBE. It therefore appears that Lester is indeed showing evidence of style-shifting, but as far as single-point formant values are concerned, this primarily manifests through FOOT and STRUT merging.

Nonetheless, although Lester's mean formant values do not show an overall tendency towards more RP or SSBE-like TRAP/BATH vowels, these values are still much more variable in the live context than any other. This may be an indication of these vowels being in flux within this context. Furthermore, it is intriguing that Lester clearly possesses the TRAP/BATH split, but occasionally does not use it in contexts expected from a Southern British or RP accent. For example, Lester demonstrates the split in *after* and *path* but not *glass*, even though all three are members of the BATH lexical set. Wells (2010), who is also from Lancashire, makes anecdotal reference to a tendency for high-status families to be more aware of the existence of this split, which may explain Lester's usage despite being from the North. The inconsistencies may be the result of the difficulty in obtaining the TRAP/BATH split, which features considerable individual variation in which words may be pronounced with /æ/ instead of /a:/, such as in *photograph* (Wells 1982). Although at this juncture it is impossible to observe whether Lester uses the split more or less frequently in certain contexts, future research may be able to study this.

More generally, Lester also seems to show vowel movements similar to changes occurring among London speakers. For example, TRAP lowering is part of an ongoing change in said vowel among present-day London speakers; this vowel is lower in the solo and collaborative contexts than in the live context (Bauer 1994). Several studies have also indicated backing of TRAP, likewise observed in the same contexts (Hawkins and Midgley 2005). Research on the STRUT vowel in Southern British accents includes Hughes et al. (1979), who describes it as fronting, and Fabricius (2002), who describes it as also possibly raising. Lester's STRUT is indeed more fronted in the solo context, but it is also lower. Instead of participation in local sound change, this may indicate a centralising effect, which is also a feature of London accents (Kamata 2006). Given the fact that Lester moved to London in the years preceding analysis, London vowels may be serving as phonetic targets for these style-shifting changes.

As for why these changes are occurring between contexts, Bell's (1992) theory of audience design has already been argued against based on the general similarities between audiences in each context. I argue that Schilling-Estes's (1998) speaker design model is also insufficient. Recall how Coupland (1984:65) stated that

speaker design may lead to the construction of a "persona similar to that of [his] interlocutors". This can be used to "draw on (or carefully avoid) the 'voices' of others" in order to "differentiate situations and display attitudes" (Irvine 2001:31). However, it is unclear exactly what kind of persona Lester would be attempting to project by levelling the Northern aspects of his speech as his audience is primarily international. Furthermore, attitudinal studies on RP accents suggest they are considered less likeable than regional accents and "attract labels like 'posh' and 'snob'" (Garrett 2010:627). It seems far more plausible to suggest that the style-shifting results from Lester's increased awareness of his speech when filming more scripted videos. Although speaker design does not seem to be the primary motivation for the data presented here, however, it may require further investigation before being fully dismissed. For example, the variety of content in a single "vlog" may lead to style-shifting within a particular context: we may see Lester moving between more and less Northern-like vowels based on the topic of the vlog or the type of activity he is engaged in within that vlog.

I argue that Labov's (1966) attention paid to speech model best explains the results found in this particular paper. Under this model, we would expect higher usage of prestigious variants by Lester—particularly those of RP—in the solo vlog context. Lester did not show a completely consistent movement towards RP or SSBE overall; nonetheless, the increased distinction between FOOT/STRUT may be the result of Lester paying more attention to speech, and thus levelling marked aspects of his Northern accent in his vlogs for a more international audience.

6 Conclusion

This paper has shown style-shifting in one YouTube vlogger by demonstrating significant changes in his vowel quality between different YouTube recording contexts that differ in level of spontaneity: a solo vlog, a collaborative vlog, a gaming video, and a live video. While Phil Lester did not show consistent vocalic shifts between his Northern English variety and RP/SSBE, several individual vowel productions in the less scripted contexts mirrored ongoing linguistic changes in London accents, such as TRAP lowering. His FOOT and STRUT vowels also showed greater merging in this context, indicating a greater use of a well-known Northern English feature. TRAP and BATH displayed high variability in F2 in the live context, indicating an instability of this contrast in his speech. Overall, the evidence presented suggests that Lester style-shifts based on the context of the video—not because of audience design or a speaker designed performative register, but because of increased attention paid to speech in more scripted contexts.

Future research might investigate whether other YouTubers, particularly non-British ones, show similar types of style-shifting. One crucial drawback of the data analysed in this paper is the lack of non-YouTube related audio—it may be argued that the live contexts are not sufficiently different from normal vlogging to represent a full context change, for example. Nevertheless, the dramatic changes observed in Lester's vowel formants suggest that video context does seem to affect speech in a very marked fashion. As YouTube's usefulness as a corpus becomes more recognised, it is clear that the context of video will have to be considered when studying intraspeaker variation in this domain.

References

AmazingPhil. 2006. Phil's video blog—27th March 2006 [YouTube video]. Accessed 16 June 2016, https://www.youtube.com/watch?v=L0dsyXzmHFM

AmazingPhil. 2015a. Phil reacts to his old videos [YouTube video]. Accessed 16 June 2016, https://www.youtube.com/watch?v=mPa2HqBvtjc

AmazingPhil. 2015b. Phil is not on fire 7 [YouTube video]. Accessed 16 June 2016, https://www.youtube.com/watch?v=3YIPSjTmPmo

Ault, Susanne. 2014. Survey: YouTube stars more popular than mainstream celebs among U.S. teens [Online article]. Accessed 16 June 2016, http://variety.com/2014/digital/news/survey-youtube-stars-more-popular-than-mainstream-celebs-among-u-s-teens-1201275245/

Barbu, Stéphanie, Aurélie Nardy, Jean-Pierre Chevrot, and Jacques Juhel. 2013. Language evaluation and use during early childhood: Adhesion to social norms or integration of environmental regularities? *Linguistics* 51(2):381–411. https://doi.org/10.1515/ling-2013-0015

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, Laurie. 1994. *Watching English Change*. London: Longman.

Beck, Julie. 2015. The linguistics of "YouTube voice" [Online article]. Accessed 16 June 2016, http://www.theatlantic.com/technology/archive/2015/12/the-linguistics-of-youtube-voice/418962/

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2):145–204. https://doi.org/10.1017/s004740450001037x

Bell, Allan. 1992. Hit and miss: Referee design in the dialects of New Zealand television advertisements. *Language and*

Communication 12(3–4):327–340. https://doi.org/10.1016/0271-5309(92)90020-a
communitychannel. 2011. (We Broke Up) [YouTube video]. Accessed 28 September 2016, https://www.youtube.com/watch?v=hq91hmSGUcE
Coupland, Nikolas. 1984. Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language* 46:49–70. https://doi.org/10.1515/ijsl.1984.46.49
Coupland, Nikolas. 2001. Dialect stylization in radio talk. *Language in Society* 30(3):345–375.
Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
DanAndPhilGAMES. 2015. Forest fire!!!—Phil plays Shelter final episode [YouTube video]. Accessed 16 June 2016, https://www.youtube.com/watch?v=ph54okW5Tok
Dan and Phil Survey. 2015. [Tumblr]. Accessed 16 June 2016. http://danandphilsurvey.tumblr.com/results
Disner, Sandra Ferrari. 1980. Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America* 67(1):253–261. https://doi.org/10.1121/1.383734
Fabricius, Anne. 2002. Weak vowels in modern RP: An acoustic study of happY-tensing and KIT/schwa shift. *Language Variation and Change* 14(2):211–237. https://doi.org/10.1017/s0954394502142037
Ferragne, Emmanuel, and François Pellegrino. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association* 40(1):1–34. https://doi.org/10.1017/s0025100309990247
Frobenius, Maximiliane. 2011. Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics* 43(3):814–827. https://doi.org/10.1016/j.pragma.2010.09.018
Frobenius, Maximiliane. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics* 72:59–72. https://doi.org/10.1016/j.pragma.2014.02.008
Garrett, Peter. 2010. *Attitudes to Language*. Cambridge: Cambridge University Press.
Garrett, Peter, Nikolas Coupland, and Angie Williams (Eds.). 2003. *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. University of Wales Press.
Hawkins, Sarah, and Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35(2):183–199. https://doi.org/10.1017/s0025100305002124
Hughes, Arthur, Peter Trudgill, and Dominic Watt. 1979. *English Accents and Dialects*. London: Edward Arnold.
Hutchby, Ian. 2014. Communicative affordances and participation frameworks in mediated interaction. *Journal of Pragmatics* 72:86–89. https://doi.org/10.1016/j.pragma.2014.08.012
Irvine, Judith T. 2001. "Style" as distinctiveness: The culture and ideology of linguistic differentiation. In *Style and Sociolinguistic Variation*, ed. P. Eckert and J. Rickford, 21–44. New York: Cambridge University Press.
Kamata, Miho. 2006. A socio-phonetic study of the DRESS, TRAP and STRUT vowels in London English. *Leeds Working Papers in Linguistics and Phonetics* 11. Accessed 6 February 2017, http://www.leeds.ac.uk/arts/download/1360/kamata2006
Labov, William. 1966. The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.
Labov, William. 1984. Field methods of the project on linguistic change and variation. In *Language in Use*, ed. J. Baugh and J. Sherzer, 28–53. Englewood Cliffs, NJ: Prentice-Hall.
Ladefoged, Peter, and Keith Johnson. 2014. *A Course In Phonetics*. Boston: Cengage Learning.
Lange, Patricia G. 2007. The vulnerable video blogger: Promoting social change through intimacy. *The Scholar and Feminist Online* 5(2). Accessed 28 September 2016, http://sfonline.barnard.edu/blogs/lange_01.htm
LessAmazingPhil. 2017. Trying gummy candy sushi! (Live) [YouTube Video]. Accessed 6 February 2017. https://www.youtube.com/watch?v=wrcDNb9Xmdc
Montgomery, Martin. 1988. D-J talk. *Media, Culture and Society* 8(4):421–440.
Moor, Peter, Ard Heuvelman, and Ria Verleur. 2010. Flaming on YouTube. *Computers in Human Behavior* 26(6):1536–1546. https://doi.org/10.1016/j.chb.2010.05.023
PhanShows. 2015. Phil's younow—December 6th, 2015 [YouTube video]. Accessed 16 June 2016, https://www.youtube.com/watch?v=B2qB-7uvlrg
Philip DeFranco. 2015, February 5. Man cuts off nose to look like red skull... [YouTube video]. Accessed 10 November 2016, https://www.youtube.com/watch?v=2KMqyJTqsmk
R Core Team. 2016. R: A language and environment for statistical computing. Accessed 6 February 2017, http://www.R-project.org
Rheingold, Howard. 2000. *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, MA: MIT press.
Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. 2011. FAVE (Forced Alignment & Vowel Extraction) [Programme suite]. Accessed 28 September 2016, http://fave.ling.upenn.edu/
Schilling-Estes, Natalie. 1998. Investigating "self-conscious" speech: The performance register in Ocracoke English. *Language in Society* 27(1):53–83. https://doi.org/10.1017/s0047404598001031
Shields, Mike. 2013. PewDiePie has 12 million YouTube bros and no advertisers [Online Article]. Accessed 16 June 2016, http://www.adweek.com/videowatch/pewdiepie-has-12-million-youtube-bros-and-no-advertisers-151955
Sloetjes, Han, and Peter Wittenburg. 2008. Elan [Annotation tool, Max Planck Institute for Psycholinguistics, The Language Archive]. Accessed 28 September 2016, http://tla.mpi.nl/tools/tla-tools/elan/
Socialblade. 2016a. Zoella [User analytics website]. Accessed 10 November 2016, http://socialblade.com/youtube/user/zoella280390
Socialblade. 2016b. Danisnotonfire [User analytics website]. Accessed 10 November 2016, http://socialblade.com/youtube/user/danisnotonfire
Socialblade. 2016c. Crabstickz [User analytics website]. Accessed 10 November 2016, http://socialblade.com/youtube/user/crabstickz
Socialblade. 2016d. Charlieissocoollike [User analytics website]. Accessed 10 November 2016, http://socialblade.com/

youtube/user/charlieissocoollike
VidstatsX. 2016. VidStatsX—Phil Lester stats [Website]. Accessed 10 November 2016, http://vidstatsx.com/AmazingPhil/youtube-channel
VlogBrothers. 2008. Nerdfighting in 2008, Pakistan, and Hillary Clinton [YouTube video]. Accessed 28 September 2016, https://www.youtube.com/watch?v=oOHqknrPQZgCite
Wells, John C. 1982. *Accents of English, Vol. 1*. Cambridge: Cambridge University Press.
Wells, John C. 2010. English Places [Blog]. Accessed 6 February 2017, http://phonetic-blog.blogspot.co.uk/2012/03/english-places.html

sarah.l.lee0@gmail.com