Constraints on audience design in bidialectal speech: a case-study of Fareed Zakaria

Auromita Mitra[*]

Draft– compiled on June 12, 2025

1 Introduction

Most people around the world command multiple varieties of a language (Giles and Coupland, 1991), with many acquiring a second variety in adulthood (Eckert, 2017; Rickford and Price, 2013). It is commonly observed, both anecdotally and in sociolinguistic research, that speakers may modify their speech when in an environment where the linguistic norms differ. The motivations might include facilitating communication and fitting in, but also showing allegiance to certain social groups and identities (Trudgill (1999); Foulkes and Docherty (2014); Munro et al. (1999); Evans and Iverson (2007), a.o.). While these already demonstrate that a variety of pressures might interact in a speaker's agentive decisions about their language, another set of facts also governs such variation: automatic (non-agentive) processes that result from the interaction of language systems within a speaker, or between a speaker and their linguistic environment. These have been studied in the context of bilingualism (and to a smaller extent bidialectalism; see Lønes et al. (2023)), which involves the acquisition and processing of multiple recognized linguistic norms (Grosjean and Miller, 1994; Caramazza et al., 1973; Simonet, 2016), and in the context of mechanisms like priming, which involve automatic adjustments to perception and production in response to surrounding linguistic norms (Pickering and Garrod, 2004; Goldinger, 1996; Bock and Griffin, 2000; Garrod and Doherty, 1994). Such non-agentive phenomena might reflect constraints on the representation and processing of linguistic units, as well as pressures to maintain linguistically-relevant contrasts. These agentive and non-agentive processes are usually studied in separate contexts, often with different methodologies, and their interactions are not yet well-understood.

In this study, I examine variation in a corpus of English speech from a single individual— the political commentator and journalist Fareed Zakaria— with different audiences over a period of 21 years. Zakaria was born in Mumbai, India, in 1968 and moved to the United States at the age of 17, where he continues to live. As part of his work, he has appeared on numerous televised interviews and public forum discussions in both India and the US, addressing primarily Indian/American audiences respectively. Zakaria's interactions with these different audiences are characterized by recognizable clusters of features pertaining to two different dialects of English: Indian English (IndE) and American English (AmE). His linguistic history thus makes him an example of a speaker who has acquired a second variety of a language in adulthood, and continues to produce both his first-learned and second-learned dialect (D1 and D2) as a function of social/linguistic context. This suggests that the various pressures on variation discussed in the previous paragraph are likely to shape Zakaria's linguistic behavior (Nycz, 2015). There are multiple ways of understanding how/why a speaker might use a specific set of linguistic characteristics (here, a dialect variety) while interacting with a specific audience or interlocutor. Whatever the exact mechanism, we might characterize the behavior itself as Zakaria 'switching between dialects' (here, his D1 and D2 norms) depending on the audience, akin to a bilingual individual switching between languages based on their audience. This is the approach taken in a previous study of Zakaria's linguistic behavior by Sharma (2018), who conceptualizes it

[*] Department of Linguistics, New York University; auromita.mitra@nyu.edu

as audience design (Bell, 1984, 2006). In this study, I am interested in the exact characteristics of these norms that he 'switches' between. I examine the acoustic properties of his speech towards different audiences at various time-points, and suggest that within this broad behavior of audience design, the characteristics of speech are influenced by both automatic linguistic processes resulting from interaction between sound systems, and strategic choices that reflect Zakaria's conscious identity-creation and social motivations at a given time. Examining multiple features and gradient acoustic data, I suggest that neither strand of explanation, agentive or automatic, is independently sufficient to account for all the observed patterns, and that the most ecologically valid explanations must draw on both. This opens up questions about how these processes interact.

Given that this data is not from a controlled experiment, in the sections below I often outline hypotheses about patterns we would expect to see *if* a particular process is operative. However, given our current state of understanding about how these various processes interact, the question of which exact process would operate on a given feature in a given condition is more difficult to predict based on existing literature. Therefore, in this study I adopt the more exploratory method of identifying patterns and then trying to explain the observations in terms of known and independently-motivated processes (as, e.g., Sankoff (2004)).

1.1 Bilingualism and bidialectalism

As discussed, I follow Sharma (2018) in conceptualizing Zakaria's speech as navigating between two sound systems: his D1, whose structure corresponds to the phonetic norms of Indian English, and D2, whose structure corresponds to the phonetic norms of American English. Neither of these varieties are monolithic entities. Given Zakaria's upper-middle class upbringing in India and his social network in the US largely revolving around elite academic institutions, I take both his D1 and D2 systems to be most closely aligned to the generalized/standard forms of these varieties, devoid of specific regional features. Further, I assume that these systems exert influence over one another. It is largely accepted that both the languages of a bilingual speaker as well as the dialects of a bidialectal speaker can be understood as separate but not autonomous systems (Baker and Trofimovich (2005); Caramazza et al. (1973); de Leeuw (2014); Grosjean (1998); Mayr et al. (2017); Paradis (2001), a.o.). To what extent the degree of separation as well the degree of autonomy are parallel across bilingualism and bidialectalism, and indeed whether a clear-cut distinction between the two is feasible at all, is not obvious (Nycz, 2015; Lønes et al., 2023). However, recent research has attempted to probe this parallel, often with an eye towards using models developed for bilingualism to understand similar phenomena in bidialectal speech (see Lønes et al. (2023) for a review). Since I am interested in interactions between sound systems, I highlight some points of observed similarity between bilingual and bidialectal speech which might suggest parallels in lexical and phonological representation/processing, in order to motivate the reference to models of bilingual speech while discussing the phonetic patterns in Zakaria's speech. I am less concerned with the specific explanations for these phenomena, than that the empirical facts are parallel in bilingual and bidialectal speech.

Lexical representations are thought to include a lemma, which contains grammatical information, and a lexeme, which contains phonological information. Priming studies with bilingual speakers show that exposure to a word in one language does not prime its non-cognate counterpart in the other, which has been taken as evidence for separate lexical representations (lemmas) for words across the two languages. Cross-dialect repetition priming studies reveal similar patterns, suggesting that just like bilingual speakers, bidialectal speakers can also have separate lexical representations for their two dialects (Chen and Zhou, 2022; Woutersen et al., 1994) (but, see Melinger (2018) for differences in the processing of cross-dialect translation equivalents). This potential parallel in the organization of lexical representations is further highlighted by the behavior of cognates. Word-naming tasks with bilingual speakers show that when naming words in one of their languages, pictures that have phonologically similar labels in other language were named more quickly. This is explained as follows: activating a lemma in one language activates the semantically-related lemma in the other. This activation cascades down to the respective lexemes. In the case of cognates, since these lexemes share phonological content, the representation is co-activated, facilitating faster encoding. This has been taken as evidence that (i) lexical representations across languages can activate one another; and (ii) phonologically similar lexical representations share phonemic content across languages. Importantly, Kirk et al. (2018) reports a parallel facilitation effect for cross-dialect cognates, suggesting that lexical representations (and processing) across dialects might be organized similarly. Another observed parallel between bilingual and bidialectal speech is in the posited control mechanisms that speakers use to select between their varieties. Under certain accounts, in order to select elements from one language, a speaker must inhibit the other, which incurs a "switch cost" (Green, 1986, 1998; Blanco-Elorrieta et al., 2018; Olson, 2013). Across various levels of bilingual speech, researchers have reported that for L2 learners, switching from L2 to L1 incurs a greater cost than switching from L1 into L2. In proficient bilinguals, however, the switch costs are symmetrical across languages (Olson, 2013; Tsui et al., 2019).[1] Relevant to our discussion, Kirk et al. (2018, 2022) found the same pattern to obtain when bidialectal speakers switch between dialects. Importantly, the effect of proficiency was identical. This is significant because many existing models of bilingualism make reference to proficiency in the later-learned variety. Such parallels suggest that (aspects of) these models may be successfully applied to understand patterns in bidialectal speech as well. With this assumption, I speak of interaction between 'varieties' to encompass both multiple languages and multiple dialects while discussing prior studies and models of phonetic interaction.

1.2 Sharma (2018) on Fareed Zakaria

The present study builds on a previous investigation of Zakaria's speech when directed at different audiences: Sharma (2018). Sharma identifies 12 different phonological variables that distinguish Indian English and American English, and focuses on system-level shifts (total % of tokens that are realized as 'AmE'; coded categorically by ear). She finds evidence of audience design: the proportion of 'AmE'-coded tokens is higher with American audiences and vice-versa. The other goal of the study was to examine the small-scale temporal dynamics of the variation, i.e. whether certain periods within a given conversation have higher/lower proportions of 'AmE' tokens, to understand real-time cognitive processing. Sharma reports short-term increases in the number of 'IndE'-coded tokens during periods of high cognitive load within a conversation, concluding that higher cognitive load results in a reversion to the first-learned variety regardless of the audience, showing that the extent of audience design at any given time is constrained by cognitive factors. Moreover, she likens this to the cognitive primacy of the first language in bilingual speech, drawing explicit connections between Zakaria's bidialectalism and theories of bilingualism. While my goals differ from Sharma's, two aspects of her proposal inform the current study: (i) within a broad overarching behavior of agentive audience design, other 'automatic' factors influence the exact patterns of shifts in Zakaria's speech; (ii) certain patterns of interaction between Zakaria's sound systems parallel those reported in bilingual speech.

Another point of interest for this study is that the 12 features that Sharma (2018) identifies as salient differences between IndE and AmE differ in various ways: they include both vowel and consonantal features, some apply to specific lexical items, while others are lexically non-selective. The complete set of variables are reproduced in figure 1. Sharma consciously makes the simplifying assumption that all parts of the system (all the phonological variables) are equally available for the

[1] See Kroll and Gollan (2013); Blanco-Elorrieta and Pylkkänen (2017); Blanco-Elorrieta and Caramazza (2021) for discussion and alternative explanations for such patterns.

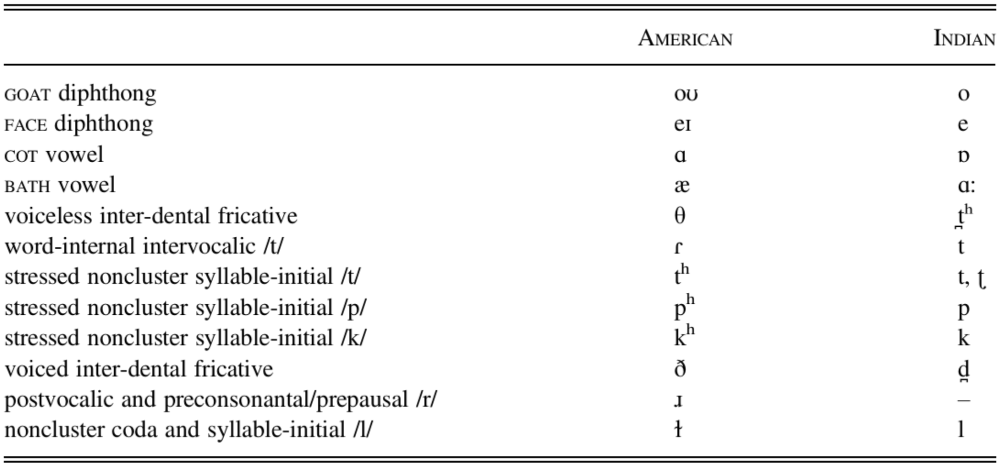
socially-driven shifts, focusing on overall system-wide changes in the number of 'AmE' or 'IndE' coded tokens. Given my interest in linguistic constraints on variation, I focus on a subset of these variables that differ in their linguistic properties and probe this assumption explicitly. Moreover, most of these identified variables are gradient and have the possibility of varying on a continuum between an AmE and an IndE prototype. Since I am specifically interested in the characteristics of Zakaria's D1 and D2 norms, and given the existing literature on gradient phonetic shifts in both agentive and 'automatic' processes of variation, I depart from Sharma (2018) in examining acoustic shifts, rather than categorically coding tokens as 'AmE' or 'IndE'. Finally, since I am concerned with the D1 and D2 systems, I also examine how these features change over the years, without examining short-term dynamics within a single conversation as Sharma does.

1.3 Variables in the current study

I examine five of the variables identified in figure 1: aspiration in the stops /p/, /t/, and /k/ (measured through the acoustic feature of voice onset time; VOT), and vowel realization in the BATH class and LOT class (measured through the acoustic feature of vowel quality— first and second formant frequencies). Three of these variables (aspiration in /p/,/t/, and /k/) vary along the same acoustic feature, whereas the latter two vary along a different feature. Because these call for different methodologies, I will present the methodology, specific hypotheses, and results for these two sets in separate subsections.

Beyond expository ease, however, there is also theoretical reason to doubt that of all of these variables are completely independent of each other. Although speech is highly variable (e.g. due to idiosyncratic speaker characteristics, cross-linguistic differences in how a given category is realized, etc), researchers have observed that within a given speaker, categories with shared phonological features tend to 'vary together', so that variability in a given category can constrain the possible patterns of variability in the other. This has lead to the idea that a general principle of *uniformity* constraints phonetic variation (Chodroff and Wilson, 2017; Johnson, 2021). Another relevant fact is the pressure to maintain linguistically-relevant contrasts within a given sound system. Since VOT is a primary cue for voicing in English, the relative VOTs of /p, t, k/ are directly involved in maintaining the voicing contrast across stops. Lengthening the VOT in one stop to the exclusion of the others would endanger this contrast. On the other hand, the two vowel features are not involved in maintaining contrast in the same way: producing the BATH class as /æ/ (AmE-like) while producing the COT class as /ɔ/ (IndE-like) theoretically wouldn't endanger contrast. Therefore, their changes need not constrain each other due to linguistic pressures. Existing findings confirm that in socially-motivated shifts as well as cross-variety interaction, different vowels can indeed vary in different ways and to different extents (Sankoff, 2004; Babel, 2012). In light of this, I treat VOT as a single feature that varies between IndE and AmE (and applies to the categories /p,t,k/), and think of the two vowel features (realization of the BATH class, realization of the COT class) as potentially independent of each other. Finally, note that these vowel features also differ from VOT in being lexically-specific (i.e., the vowel difference between dialects applies to particular lexical sets). I return to this when I discuss the results for vowels.

The analyses reveal that Zakaria is able to produce AmE-like targets for both VOT and vowel features, adding to a growing body of research suggesting that an individual's linguistic abilities remain plastic though adulthood (de Leeuw and Celata, 2019). Moreover, Zakaria's D1 system evidences shifts towards D2 norms. In VOT, this is manifested as longer VOTs than prototypical IndE values, and in vowels, as fronted realizations of the BATH vowel and lowered realizations of the COT vowel. Thus Zakaria's D1 shows unmistakable influence of his long exposure to AmE, paralleling findings from bilingual speakers. In spite of the acquisition of D2 and changes to D1, the results show that Zakaria



	American	Indian
goat diphthong	oʊ	o
face diphthong	eɪ	e
cot vowel	ɑ	ɒ
bath vowel	æ	ɑː
voiceless inter-dental fricative	θ	t̪^{h}
word-internal intervocalic /t/	ɾ	t
stressed noncluster syllable-initial /t/	t^{h}	t, ʈ
stressed noncluster syllable-initial /p/	p^{h}	p
stressed noncluster syllable-initial /k/	k^{h}	k
voiced inter-dental fricative	ð	d̪
postvocalic and preconsonantal/prepausal /r/	ɹ	–
noncluster coda and syllable-initial /l/	ɫ	l

Figure 1: Phonological variables examined in Sharma (2018)

maintains consistent differences between his varieties across all five features and at every timepoint. Alongside confirming Sharma's findings of audience design, this again mirrors proficient bilingual speakers who are able to maintain language-specific targets for their categories, and employ them in socially appropriate contexts (Flege et al., 2003; Grosjean and Miller, 1994). The patterns of shift in VOT over the years, in conjunction with independent evidence about Zakaria's subjective experience of shifting identity and social goals, provide evidence for specific socially-motivated shifts in both D1 and D2 (Bell, 2006; Giles et al., 1991). In line with many previous findings (e.g. Nielsen (2011); Piccinini and Arvaniti (2015)), this shows that VOT can function as a site for socially-motivated shifts. In contrast, I find that the COT and BATH vowels remain unchanged over the time period examined here, suggesting that Zakaria maintains stable targets for these categories in both his D1 and D2, and that these particular features may not be equally malleable for socially-motivated shifts over shorter time-spans. I propose that their lexically-specific nature is partly responsible for this difference.

2 Materials

The materials for this study come from eight public interviews of Fareed Zakaria, recorded for broadcast, over a span of 21 years, from 2003 to 2024. Four of these feature primarily Indian audiences and four primarily American audiences and interviewers. The interviews are listed in table 1, and cover similar topics (international relations, politics, current affairs). All of these were downloaded as videos (mp4) from the internet and then converted to wav sound files. I used the ASR tool BedWord (Ma et al., 2024)to generate transcripts and then used semi-automatic annotation in DARLA (Reddy and Stanford, 2015) to generate phone-annotated textgrids. The transcripts were checked and manually corrected, and parts of audio with significant background noise, interviewer speech, overlapping speech, and long pauses were removed. The durations in the table reflect the length of the usable audio.[2] As the table shows, the durations are not equal, and consequently some contexts and time-points have a larger number of tokens for analysis than others.

The interviews were chosen to match those reported in Sharma (2018) (with the exception of the 2003 interview with an American audience, which was not available online at the time of this study; I

[2] The 2012 interview for Express Adda was only available as a series of shorter clips on YouTube; these were merged into a single audio. All other recordings came from single video files.

Label	Interview	Duration of usable audio
ame-1	Charlie Rose, 2003	21 min
ame-2	Charlie Rose, 2008	40 min
ame-3	Charlie Rose, 2010	14 min
ame-4	Firing Line, PBS, 2024	20 min
ie-1	Walk the Talk, NDTV, 2003	15.5 min
ie-2	Walk the Talk, NDTV, 2008	14.5 min
ie-3	Express Adda, Indian Express, 2012	13 min
ie-4	Express Adda, Indian Express, 2024	1.7 hrs

Table 1: List of interviews with duration of analyzed audio (interviewer/overlapping speech, noise, long pauses, and disfluencies removed)

instead used an interview on Charlie Rose from the same year). A gap of seven years between Sharma (2018) and this study allowed me to add data from a more recent time-point, and I selected two interviews from 2024 that deal with similar topics. Although the gaps between subsequent timepoints are not even, I opted to match the data from Sharma (2018) as closely as possible, to allow for comparisons of the findings. The difference in audience (Indian vs American) is coded as Context (ie vs ame), and the four time-points as Timepoint (1,2,3,4). For simplicity, I labeled each interview as its unique context-timepoint combination (ie-1, ame-1, ie-3, etc). I also refer to the 8 unique context-timepoint pairs as 'conditions'. I use the lowercase labels *ame* and *ie* to refer to the audience type/ context of Zakaria's speech in a given instance, and the terms *D1* and *D2* to refer to the phonological systems he is drawing on. While talking about prototypical features of the two dialects, independent of Zakaria's speech, I use the terms 'Indian English' IndE and 'American English' AmE. This is to recognize that the idea of his 'switching between dialects' depending on the audience is a theoretical assumption, and part of the question pursued here is what these varieties look like, to what extent they have properties prototypical of AmE and IndE, and to what extent this correlates with the context (ame vs ie) in which he has produced the speech. The eight audio files, along with their phone-annotated textgrids, were used to analyze the acoustic features outlined above: voice onset time or VOT (corresponds the presence/absence of aspiration in voiceless stops), and vowel quality. The following two sections describe the analyses for the voice onset time and vowel quality respectively.

3 Voice onset time (VOT)

A salient difference between Indian English and American English is that syllable-initial stressed voiceless stops are aspirated in the latter, while they are are unaspirated in the former. This has been noted in many existing accounts of Indian English phonology (Pingali, 2009, 2022; Wiltshire, 2020), and examined in greater detail in recent work by Narkar (2021) and Narkar and Staroverov (2022). Across languages and speakers, it has generally been observed that VOT is at least partially determined by the place of articulation of the stop, such that VOT for /p/</t,k/, e.g. Cho and Ladefoged (1999) (although this is subject to individual differences and speech style: in a corpus of spontaneous American English speech, Yao (2009) found only one of the two speakers to reliably vary VOT by place of articulation). Aspiration results in a larger quantity of being air expelled during the release phase of the stop. This has a number of effects on the acoustic properties of the sound wave, one of which is a delay in the start of voicing for the following vowel, leading to a longer VOT compared to unaspirated stops. Accordingly, existing work comparing Indian English to aspirating varieties such

as American or British English have found that the former has shorter VOT in ‘aspirating’ contexts (i.e. initial voiceless stops in stressed syllables) compared to the latter (Narkar and Staroverov, 2022).

Apart from place of articulation and aspiration, a number of factors affect VOT including, importantly, whether speech is spontaneous or elicited/read, connected or isolated words. This makes direct comparison of VOT values across studies tricky, as they are often not elicited in comparable contexts. There are four studies that I am aware of which have directly compared VOTs in Indian English with those in aspirating varieties within the same study (comparable setups and speech type). Awan and Stine (2011) report VOT differences of roughly 30 ms in /p,t/ between IndE and AmE read speech, in both word-initial and word-medial stressed syllables. In all cases, VOTs produced by the AmE speakers in aspirating contexts were roughly twice as long as the corresponding VOTs of IndE speakers. Interestingly, this sample was from a group of IndE speakers with varying L1s who had moved to the US in adulthood (around 23 years), and resided there for an average of 7.5 years. Hansen et al. (2010) report very similar differences (30–40 ms) in isolated productions of word-initial /p,t,k/ from four speakers each of IndE and AmE, the former having moved to the US 3–6 months before the study. A VOT difference between AmE and IndE (L1 Hindi) speakers in wordlist production is also reported in McCullough (2013), although numerical values are not provided (the graph is reproduced in the Appendix; figure 17). In Narkar and Staroverov (2022), a comparison of IndE and British English stops from 130 speakers in a corpus of read speech revealed a difference of 30 ms for /p/ and 28 ms for /k/ in aspirating (word-initial) contexts. Together, these suggest that across isolated and connected speech contexts, prototypical realizations of syllable-initial voiceless stops in the two dialects would differ, with the AmE VOTs being roughly 30 ms longer than corresponding IndE VOTs. Note that there are some known differences in IndE aspiration based on the L1 of the speaker. In particular, speakers whose L1s lack a 4-way laryngeal contrast (e.g. Tamil) have been reported to produce longer VOTs in aspirating contexts (Wiltshire and Harnsberger, 2006; Wiltshire, 2020). The values summarized above focus on data from speakers of languages with a 4-way contrast, as this matches Zakaria’s language background. Another point to note is that with the exception of Narkar and Staroverov (2022), all the IndE values reported here are from speakers who had lived in the US for some period at the time of recording. This might affect their VOTs, as studies of mobile speakers have consistently found VOT ‘drifts’ after even short exposure to an ambient language with a different VOT range (Sancier et al., 1997; Tobin et al., 2017). However, this is also comparable to Zakaria’s linguistic situation, since he is primarily in an AmE environment. Therefore, any persistent differences in VOT between his dialects are expected to exist over and above the effects of drift.

IndE VOTs from speakers residing in India are available in the literature, but these do not directly compare VOTs to aspirating varieties in similar production contexts. In read sentences from 14 speakers (10 residing in India, 4 having been in the US for less than 6 months), Sirsa and Redford (2013) report VOTs of 7–11 ms for /p/ and 15–20 ms for /k/. All of the speakers had been educated in English throughout school and university. Since the study was concerned with the speakers’ L1 backgrounds, the authors don’t discuss any differences in VOT between the speakers residing in India vs the US. Wiltshire and Harnsberger (2006) reports VOTs in isolated word-initial stops for 4 L1-Gujrati speakers living in India: 20ms for /p/, 36ms for /k/. All the speakers had started learning English between the ages of 3-6, were pursuing university degrees taught in English, and used it regularly in a variety of contexts. Finally, in wordlist elicitation from 3 L1-Assamese speakers in India, Wiltshire (2020) reports VOTs of 13 ms for /p/, 15 ms for /t/, and 26 ms for /k/. While age of English acquisition is not reported, the authors report that all three speakers were ‘in their twenties to early thirties, highly educated, and living in urban environments’. On the whole, these values are slightly smaller than those summarized above (although recall the caveat about comparing values across studies). As far as I am aware, there are no reported VOTs from IndE spontaneous speech. Spontaneous speech data is available for AmE, however: Yao (2009) examines AmE VOTs in the spontaneous speech of 19 speakers from the Buckeye Corpus and reports an average VOT of 48 ms for /p/, 51.2 ms for /t/, and 57.9 ms for /k/. From a corpus of conversational speech by 13 speakers of AmE, Piccinini and Arvaniti (2015) report comparable values for /k/, with slightly higher values for /p,t/: an average VOT of 56 ms for /p/, 62 ms for /t/, and 59 ms for /k/.

The data summarized here suggests that Zakaria's targets for prototypical AmE VOTs might approximate those reported in Yao (2009) and Piccinini and Arvaniti (2015), whereas targets for prototypical IndE targets might be around 30 ms (or 0.5 times) smaller than these.

An underlying assumption in this study is that when speaking to an American audience, Zakaria is drawing more heavily from his D2 sound system, and vice-versa for Indian audiences. Studies of speakers who command multiple languages differing in their typical VOT values have shown that speakers can maintain separate VOT ranges for their two varieties in production (Antoniou et al., 2010; Magloire and Green, 1999; Piccinini and Arvaniti, 2015) as well as perception (Elman et al., 1977; Hazan and Boulakia, 1993; Garcia-Sierra et al., 2009; Gonzales and Lotto, 2013). Although I am not aware of similar VOT studies across the dialects of a bidialectal speaker, given the parallels noted in section 1.1 and Zakaria's overall ability to maintain a global perceptible difference across his varieties (Sharma, 2018), I hypothesize that it is in-theory possible to maintain dialect-specific VOT values as well. Therefore, if stops are more likely to be aspirated in Zakaria's D2 (which draws from AmE), then we expect VOTs to be longer in ame contexts, compared to the ie contexts. I hypothesize that this will be the case for each stop.

Even within a single language, it is possible for an individual's VOT values to change over time. This can happen due to age, sociolinguistic factors, or influence from another language/variety whose typical VOT values are different (Ryalls et al., 2004; Sancier et al., 1997; Chang, 2012). Given the expected targets for the two varieties outlined above, lengthening of VOTs over time can be taken as an indiction of movement towards AmE-like norms, and shortening of VOTs as movement towards IndE-like norms. Given the substantial literature on VOT drifts (changes toward ambient values) in response to both short-term and long-term exposure, one plausible direction of change over time is towards AmE norms. Such exposure-driven drifts are not language-selective, and have been reported to affect both languages of the bilingual individual (Sancier et al., 1997; Tobin et al., 2017). However, such changes are not absolute or inevitable since, as noted earlier, individuals are still capable of maintaining language-specific VOT values. A possible outcome of drift in this case, then, might be that Zakaria's VOTs in his D1 are longer than those of IndE speakers who are not in an AmE-speaking environment, without necessarily being indistinguishable from his D2 norms. The literature surveyed above does show longer IndE VOTs in speakers residing in the U.S. compared to speakers recorded in India (with the caveat of cross-study comparisons), supporting such an expectation. One concrete hypothesis therefore is that Zakaria's D1 VOTs will lengthen over time. However, since VOT changes over time are also known to be affected by numerous social, physiological, and cognitive factors, this part of the analysis is largely exploratory.

Finally, another difference between the dialects is that the coronal stops that are realized as alveolar /t,d/ in AmE are most often realized as retroflex /ʈ, ɖ/ in IndE. I have ignored this additional source of variability here and focused on VOT alone. Sharma (2018) observes some instances of retroflex stops with aspiration in Zakaria's speech, showing that the two dimensions of variation (aspiration and place of articulation) as at least somewhat independent, but reports that such tokens are rare. However, since there is some evidence that retroflex stops might have shorter VOTs compared to labial, dental, palatal, and velar stops Hussain (2018) (including in Indian English; Wiltshire and Harnsberger (2006)), follow-up work should consider this dimension of variability in relation to the VOT data reported here.

context/ timepoint	ame				ie			
	total	p	t	k	**total**	p	t	k
1	**149**	53	42	54	**183**	55	56	72
2	**560**	176	178	206	**194**	57	63	74
3	**160**	52	54	54	**192**	60	58	74
4	**175**	56	63	56	**562**	164	192	206

Table 2: VOT token counts in each condition

3.1 Methodology

The analysis focuses on voiceless stops (/p/, /t/, /k/) in the initial position of stressed syllables, since these are exactly the stops that are expected to be aspirated in American English, and thus differ in VOT between the varieties. There were 2175 tokens in total (673 /p/, 706 /t/, 796 /k/); table 2 shows the token counts for each stop and at each context-timepoint pair. Lexical frequency determines the relative numbers of each stop, and the length of the audio clip determines the token count in each condition. As the table shows, the proportion of different stops is comparable within each condition.

The common practice for measuring VOT is to mark the release burst of the stop and the onset of vowel, and measure the distance between the these. To enable examining a large number of data, I used the software package AutoVOT (Keshet et al., 2014) to automate this. Given an audio file and textgrids annotated with the approximate location of each relevant stop, AutoVOT uses a classifier to identify the position of the stop burst and vowel voicing onset. The time between the two can then be measured, to get the VOT for that token.

As described in section 2, I used DARLA to generate phone-level annotations for each interview file. I used this phone-annotated textgrid, the CMU pronunciation dictionary, and a set of praat and python scripts to generate the required input for AutoVOT. This workflow is based on the recipe described by Eleanor Chodroff[3] with some modifications. Some of the scripts were also adapted to account for changes in newer versions of Praat. I first identified unique words in the data with stop consonants in prevocalic position, excluding those with word-initial /s/. I removed instances of *to*, since this was frequent and often reduced. I retained only those stop-vowel sequences that were in the initial position of stressed syllables and used a praat script to create a textgrid that identifies the midpoint of the DARLA-generated interval for each stop consonant and extends it by creating new boundaries 31 ms away from the midpoint on each side. This serves as the input to AutoVOT, which identifies the location of the burst within this extended window. The software offers the option to either use one of the built-in classifiers, or train a new one. I used the former, opting for one that was trained on American English data. While these were not trained on Indian English data, I chose to use the same classifier for the entire dataset to enable comparison. AutoVOT was run separately for each stop, and the output textgrids merged to give a final output like shown in figure 2 for each file. From these, VOT was calculated using a praat script that measured the difference in milliseconds between the burst and the onset of voicing. This measure was used for descriptive visualizations and statistical analyses. Table 3 summarizes the mean VOT in each condition.

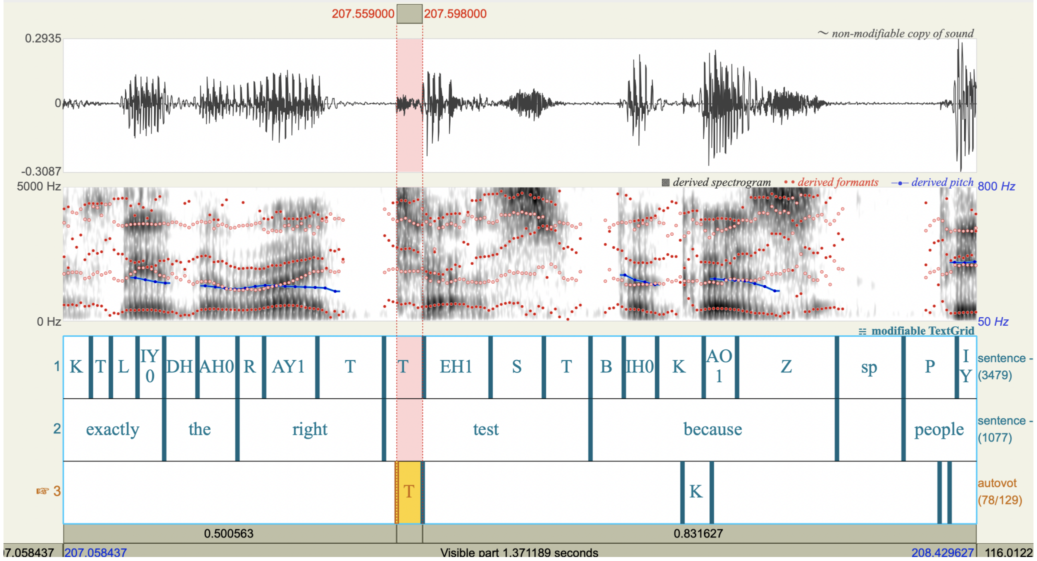


Figure 2: Output textgrid from AutoVOT

Context	Stop	Timepoint-1	Timepoint-2	Timepoint-3	Timepoint-4
ame	/p/	31.91	34.32	17.94	24.48
	/t/	44.79	39.38	33.2	32.84
	/k/	56.02	43.52	34.56	35.7
ie	/p/	15.55	15.88	20.5	18.74
	/t/	25.71	23.17	24.43	26.15
	/k/	24.49	27.43	22.46	30.67

Table 3: Average VOT for stressed syllable-initial /p,t,k/ at each context and timepoint

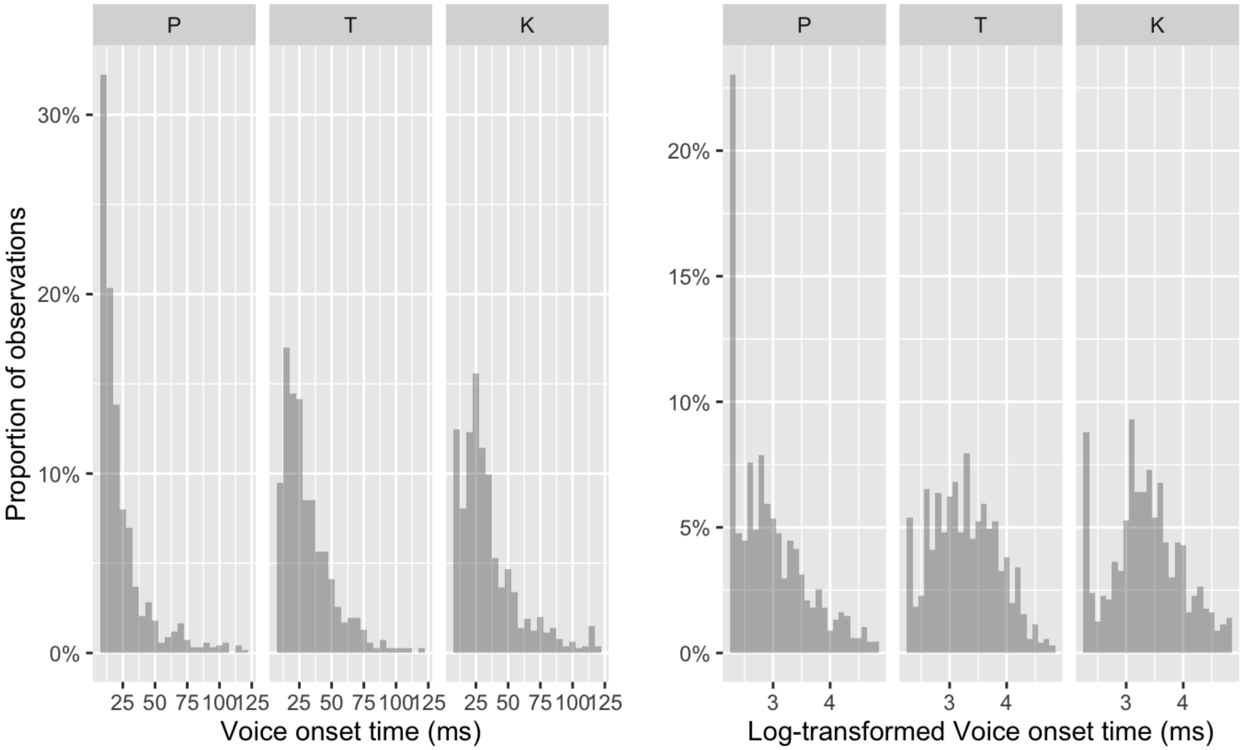


Figure 3: Distribution of VOTs before and after log-transformation

3.2 Statistical analyses

The distribution of VOTs in the data is strongly right-skewed (see figure 3). Since all values are positive, I use log(VOT) as the response variable in all models (Sonderegger, 2012; Stuart-Smith et al., 2015). All statistical analyses were performed in R (R Core Team, 2024). I used linear mixed effects models to test the effect of audience and time on the VOT of /p, t, k/. The dependent variable was log(VOT), a numerical variable with a range of 2.2 to 4.8. The independent variables/ predictors were Stop (levels: p, t, k), Context (levels: ame, ie), and Timepoint (levels: 1, 2, 3, 4). These are all categorical, and were sum coded. Since Zakaria's speech over the years is expected to be influenced by a number of factors, including cognitive and sociolinguistic, we have no reason to expect a linear effect of time. To reflect this, Timepoint was treated as a categorical, as opposed to ordinal or numerical, variable. I fitted models using the lme4 package Bates et al. (2015), and performed a sequence of model comparisons using ANOVAs to determine the best model fit. To more closely examine differences between specific conditions, I then used emmeans (Lenth, 2025) to compute pairwise comparisons among all levels of the predictors in the best-fitting model. For all significance tests, the alpha criterion was set at $|t| > 2$. Word was treated as a random effect in all models.

3.3 Results

Given previous findings, we expect the identity of the stop to predict VOT, such that /p/ has the lowest VOT, followed by /t/ and /k/. This was confirmed for our data at each context-timepoint pair, ruling out the possibility of gross mis-measurements in any of the files. Reflecting this difference,

[3] https://eleanorchodroff.com/tutorial/autovot.html

Stop is a significant predictor of VOT in the model. Since stop-specific differences are not of interest in this study, I will not discuss this effect any further. The first question of interest concerns VOT differences between Zakaria's D1 and D2. We expect longer VOTs for each stop when it is produced before American audiences (Context = ame). If this is a reliable behavior, we expect Context to be a significant predictor of log(VOT) in the model. This is confirmed; adding Context as a predictor improves model fit compared to a null model that only includes *stop* as a predictor ($\chi^2(1) = 300.84, p < 0.001$). As figure 4 shows, VOTs are shorter in the ie context across all the stops. To address the questions about Zakaria's speech over time, I test the effect of timepoint on VOT. Although VOT can change over the lifespan due to purely articulatory factors (Ryalls et al., 2004), this is not of interest in this study. It was confirmed that in the speech examined here, there is no overall lengthening/shortening of VOT over the years, as reflected in the absence of a significant independent effect of Timepoint on VOT. Instead, we are interested in VOT changes due to the interaction between D1 and D2. This will be reflected by context-specific changes over time. Thus, to test the hypothesis that one/both of Zakaria's sound systems are moving towards/away from the other over the years, I tested for the effect of a Context$*$Timepoint interaction. This significantly improved model fit compared to a model that lacks an interaction ($\chi^2(3) = 54.64, p < 0.001$), showing that the time affects VOT differently for the two contexts. Given the research questions, the optimal model was: $VOT \sim Stop + Context * Timepoint + (1|Word)$. Table 4 summarizes the model coefficient estimates (β), standard errors (S.E.), t-values, and p-values for this model. p-values smaller than 0.05 are starred.

To examine the Context$*$Timepoint interaction more closely, I examined pairwise comparisons in this model using emmeans, which revealed that over the years, Zakaria's VOTs in the ame context is decreasing, while those in the ie context are increasing. Table 3 and 5 summarize this: although there is a dip in VOT at timepoint 3 for both contexts, the overall trend is towards larger VOTs in ie and smaller VOTs in ame over the years. The VOT change in each context is significant, as confirmed by subsetting the data by context and finding a significant effect of timepoint for each, in the expected directions (model outputs for individual stop consonants can be found in the Appendix). This means that the difference in VOT between the two contexts is decreasing over time. Figure 6 shows this pattern: at timepoint 1, there are large VOT differences between ame and ie for each stop, whereas at timepoint 4, the differences across contexts are much smaller. The histogram in figure 5 makes this even clearer: the VOT distributions in the two contexts start out as distinct at timepoint 1, and are nearly overlapping at timepoint 4.

Finally, the interaction of stop with context$*$timepoint was found to improve model fit, suggesting that the effect of the main predictor of interest differs across the stops. In order to interpret the three-way interaction, I subsetted the data by stop, and fitted separate models for /p/, /t/, /k/ to examine the patterns within each stop more closely. Note that this reduces the statistical power of the models due to reduced data points. However, doing this allows for less complex models that can be interpreted more straighforwardly. One consideration that motivates this is that the /p/ has overall lower VOTs than the other stops. This means that for this stop, there is simply less room for movement for either processing constraints or socially motivated change. This principle of room-for-movement is a known linguistic constraint on variability across all levels, and has been observed in existing studies on variation in VOT (e.g. Bullock and Toribio (2009)). Subsetting the data prevents patterns in the other stops from being obscured. The model outputs for each individual stop can be found in the Appendix. The result for each stop mirrors the overall patterns: VOTs are significantly longer in the ame context, and (barring some exceptions) are decreasing over time in ame and increasing over time in ie.

One limitation of this analysis is that I did not account for speech rate in calculating VOT differences. While there is no clear consensus about the effect of speech rate on VOT, at least some

Fixed effect	Estimate (β)
Stop-p	-0.285^{***} (0.025)
Stop-t	0.031 (0.024)
Timepoint-1	0.102^{***} (0.024)
Timepoint-2	0.025 (0.020)
Timepoint-3	-0.150^{***} (0.024)
Context-ame	0.225^{***} (0.013)
Timepoint1:ame	0.122^{***} (0.024)
Timepoint2:ame	0.042^{**} (0.020)
Timepoint3:ame	-0.030 (0.024)
Intercept	3.217^{***} (0.019)
Observations	2,175
Akaike Inf. Crit.	3,439.034
Bayesian Inf. Crit.	3,507.251
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Optimal model: VOT

Context	Contrast	Std. error	t-value	p-value
ame	timepoint1 / timepoint2	0.057	3.221	0.007
ame	timepoint1 / timepoint3	0.091	6.624	< 0.001
ame	timepoint1 / timepoint4	0.083	5.623	< 0.001
ame	timepoint2 / timepoint3	0.062	5.082	< 0.001
ame	timepoint2 / timepoint4	0.055	3.841	0.001
ame	timepoint3 / timepoint4	0.055	-1.171	0.645
ie	timepoint1 / timepoint2	0.054	-0.079	1.000
ie	timepoint1 / timepoint3	0.060	1.822	0.263
ie	timepoint1 / timepoint4	0.038	-3.901	0.001
ie	timepoint2 / timepoint3	0.060	1.926	0.218
ie	timepoint2 / timepoint4	0.038	-3.888	0.001
ie	timepoint3 / timepoint4	0.034	-6.201	< 0.001

Table 5: Pairwise contrasts: VOT across time

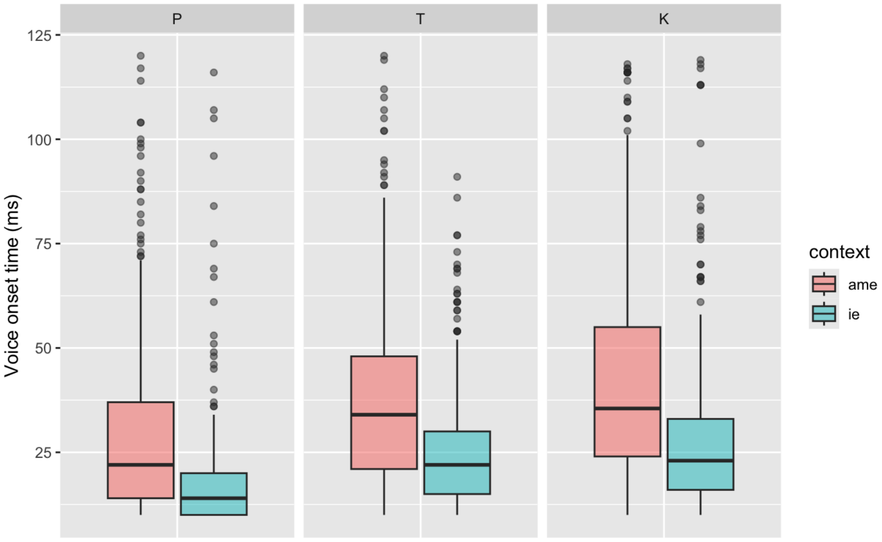


Figure 4: VOTs across contexts for each stop consonant

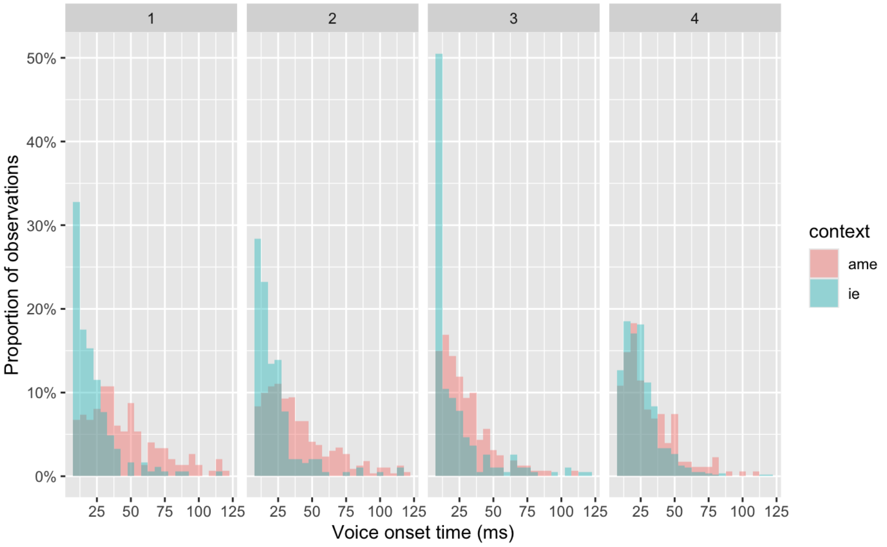


Figure 5: VOT distributions over time

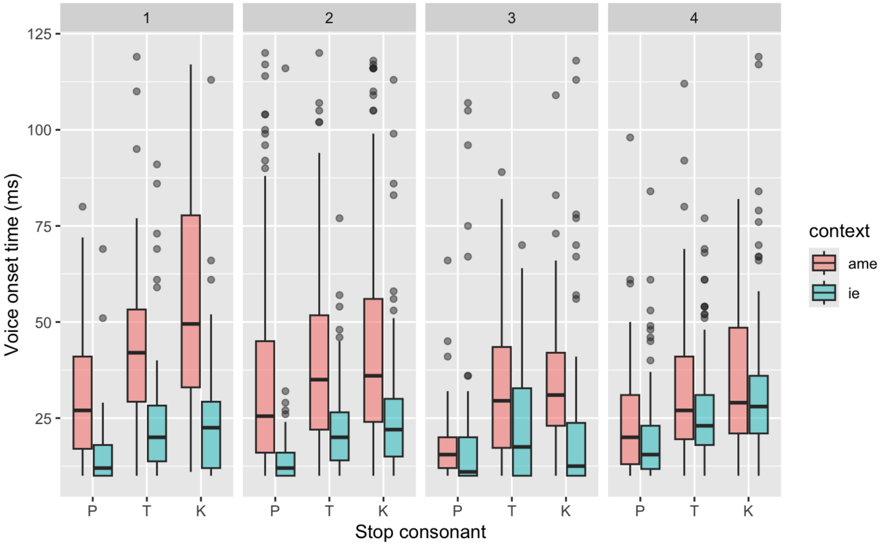


Figure 6: VOT difference between ie and ame contexts over time

studies have reported VOT in stressed syllables to be predicted by (certain measures of) speech rate, e.g. Sonderegger (2012). Here, the measure that was found to be relevant was local speaking rate, a measure of how much the rate of a given phrase deviates from the speaker's own global/mean speaking rate. Since this study analyzed condition-wise VOTs that are averaged over a large number of tokens from different phrases, this effect is not expected to significantly affect the results. However, follow-up work should verify this by using speech rate-normalized VOTs.

3.4 Discussion: VOT

First, we note that Zakaria's VOTs both ie and ame contexts (his 'D1' and 'D2' VOTs) are shorter than those reported in the three comparison studies: Awan and Stine (2011); Hansen et al. (2010), and Narkar and Staroverov (2022). This is not surprising, given that VOTs in spontaneous speech are expected to be shorter than those in isolated word-elicitation or reading tasks. Comparing Zakaria's ame-context productions to spontaneous speech data from Yao (2007) and Piccinini and Arvaniti (2015), we find that at timepoint 1, his VOT for /k/ is nearly identical to both, whereas those for /p,t/ are comparable (<10 ms difference) to Yao (2007) (and thus shorter than Piccinini and Arvaniti (2015), who reported longer VOTs than Yao (2007) for /p,t/). If we take the values from these studies to be indicative of prototypical AmE targets in spontaneous speech, it is reasonable to conclude that at timepoint 1 Zakaria is capable of producing these targets, especially for /k/. At this same time, his VOTs in the ie context are much smaller, by an order of 15 ms for /p/, 20 ms for /t/, and 30 ms for /k/. This difference is consistent with those reported in the available comparison studies, particularly for /k/ (note that the overall lower VOT values for /p,t/ in spontaneous speech likely limits the difference range between AmE and IndE for these stops). His D1 VOTs are comparable to those reported in Wiltshire and Harnsberger (2006) and Sirsa and Redford (2013), in single-word and read speech from speakers living in India. Assuming that these represent reasonable targets for IndE spontaneous speech, Zakaria's speech at timepoint 1 is similar to bilingual speakers who maintain separate VOTs for the different languages they command, e.g. Piccinini and Arvaniti (2015); Antoniou et al. (2010). Since Zakaria had been living in the US for nearly 21 years at this point, and data from earlier timepoints is not available, we are not able to say anything about his acquisition process. Instead, the data allows us to conclude that (i) he has acquired the VOT range of his D2 at this point, and (ii) he has retained the ability to still produce his D1 VOT norms. With this in mind, what explains the patterns over time?

Before discussing this further, I first rule out a potential alternative explanation, namely that his timepoint 1 productions do not actually evidence acquisition of D2 norms. Another theoretically-possible option could be that the time trajectory studied here still reflects his acquisition. Zakaria grew up speaking Hindi/Urdu which, like other Indic languages, has a 4-way laryngeal contrast. Therefore, he had an aspirated category in one of his languages when he moved to the US. We might imagine that at earlier timepoints he was mapping the syllable-initial stops in D2 to the aspirated stops of his L1, and only later does he actually acquire the AmE aspirated stop category. This would explain the shortening of VOT over time, as the prototypical VOT for aspirated stops in Indic languages is longer than those in aspirated varieties in English Narkar (2021); Narkar and Staroverov (2022); Lisker and Abramson (1964). However, the actual VOT values do not support such an interpretation. The D2 VOT values at timepoint 1 are already at the lower end of reported VOTs for AmE spontaneous speech. Aspirated stops in Indic languages have much higher VOTs, making it very unlikely that these are instances of transfer from an L1. Instead, these appear to be AmE-specific values. Thus, we retain the conclusion that at timepoint 1 (2003), Zakaria had already acquired the prototypical VOT targets for his D2.

The patterns over time may be seen as two separate trends: (i) his D1 VOTs are lengthening over time (moving towards D2 norms); (ii) his D2 VOTs are shortening over time (moving towards D1 norms). Both these patterns have been individually noted in existing literature on VOT changes, and we do not have sufficient evidence to say conclusively whether or not these are independent of each other. It is also plausible that no single factor is responsible for all of the observed shifts.

Since Zakaria lives and works primarily in an AmE-speaking environment, the changes to his D1 are on the surface easier to explain. There is a large literature noting changes to individuals' first-learned variety towards the linguistic norms of their place of residence. One pathway for such changes, particularly to VOT, is drift. This is a process where an individual's VOT in either L1 or L2 moves towards ambient norms upon exposure to a different VOT range (Sancier et al., 1997; Chang, 2012; Tobin et al., 2017). Moreover, since such shifts are thought to proceed through low-level perceptually-driven processes, it is plausible that they are cumulative, so that a longer length of residence should correlate with more D2-like norms (here, longer VOTs), as is indeed seen in the data here. Under this explanation, we assume that the changes observed over the four timepoints here are part of a longer process that started earlier than 2003 and is likely to continue. Importantly, drift relies on continued exposure to the D2 norms, and we would expect that if Zakaria were to spend some length of time in India, his D1 VOTs would again drift towards shorter values.

A related pathway for the same pattern of shift is language/dialect attrition. This is defined as long-term changes to an individual's first-learned variety upon acquisition of a second language/dialect in adulthood (De Leeuw, 2019). The main conceptual difference between attrition and drift is that the former is understood as an effect on L1/D1 specifically, and conceived as a long-term change. In practice, as it pertains to the effects on L1/D1, it is not clear that there is a clear dividing line between 'long-term drift' and 'attrition'. Attrition interacts with proficiency and language use: the more accurately a speaker is able to produce their L2 targets, the more the extent of attrition in L1 (Major, 1992), and the more immersed a speaker is in an L2 environment, the greater the attrition in L1 (De Leeuw et al., 2010). The latter two findings give us reason to pause– while drift/attrition is compatible with Zakaria's linguistic situation, the details are difficult to account for on these grounds alone. Specifically, attrition is usually thought to result from (and therefore coincide with) the process of the acquisition of the new variety. By 2003, Zakaria shows clear evidence of having already acquired his D2, being highly proficient in producing it, and had been immersed in a D2-environment for 21 years. In light of this, the fact that he produces VOTs that are at the low end of reported IndE values is notable. If relying solely on attrition/drift to explain his D1 patterns, one would need to explain why his D1 VOTs apparently remain IndE-like up until 2003, and only then start to lengthen.

I propose that a second factor is at play: agentive and socially-motivated accommodation. Specifically, I propose that Zakaria's D1 VOT norms do undergo the expected drift towards longer AmE-like values over the years, but that at the earlier timepoints examined here, he is additionally trying to approximate IndE-like norms as a part of a strategic social move. This explains the particularly short VOTs at timepoint 1. In contrast, at later timepoints he is less motivated to accommodate (for reasons to be discussed soon), and therefore produces his D1 targets which are, as a result of drift, longer than prototypical IndE values. Note that audience design/accommodation by itself is also unable to account for the patterns— if we simply assume that Zakaria is accommodating to IndE norms at the earlier timepoints and not at the later timepoints, we would still need to explain why the later VOTs in the absence of accommodation are longer than prototypical IndE norms. Instead, positing an interaction between the processes provides a more empirically adequate picture. While it is not possible to provide conclusive evidence for this account based on his speech data alone, I will argue that it coheres with the patterns of shift in Zakaria's D2, allowing us to account for the full range of his VOT behavior in terms of known processes. I now turn to these D2 shifts, and motivate why his audience design strategies might follow the pattern that I sketched by discussing available evidence from his individual biography over the years.

The data from Zakaria's speech in ame contexts, his 'D2', shows that while his VOTs approximate AmE-like norms at timepoint 1, they deviate from this in subsequent years, becoming shorter. In terms of linguistic processes arising from cross-variety interactions, this is initially surprising. Given the known effects of ambient language and continued D2 use, and given that he has already acquired his D2 at least by 2003, it is not clear why his D2 VOTs would then shorten (move towards D1 norms) over the years. I propose that this pattern reflects not automatic (psycho)linguistic processes, but rather his changing social motivations, which are expressed and achieved in part through the use of language. The fact that he 'switches between dialects' depending on his audience is already an example of audience design in his linguistic behavior. I will propose that his specific goals in such design have shifted over the years. In a series of informal interviews on podcasts, Zakaria discusses his personal life, career, attitudes towards identity, origin, and language over the years. Below, I provide some quotes from these conversations, and use them to piece together a coherent picture of his motivations as they pertain to his linguistic behavior. The Communication Accommodation model (Giles et al., 1991) posits that language is used as a tool to achieve social ends. While the translation of social attitudes into linguistic behavior need not be conscious (see, e.g. the effect of unconscious bias and power dynamics on accommodation to an interlocutor: Pardo (2010); Babel (2010)), Zakaria reveals metalinguistic awareness of the effect of linguistic choice in public speaking, as well as its intentional use to modulate distance between a public speaker and their audience. In a podcast talking about the use of Hindi/Urdu (which he speaks of as two different languages) by politicians in India, he says:

> If you're speaking one of the languages, there's a way to alternate between both, which a lot of Indian politicians used to do as a way of signaling a broad embrace of both the Hindu and the Muslim communities. Nehru, India's first prime minister, used to often do that. He would say, "I am delighted to be coming here to your home." He'd repeat the word *home*, first in Urdu, then in Hindi so that, in effect, both constituencies were covered.
>
> Modi, by contrast, India's current prime minister, is a great Hindu nationalist. He takes pains almost never to use an Urdu word when he speaks. He speaks in a kind of highly Sanskritized Hindi that most Indians actually find hard to understand because the everyday language, Bollywood Hindi, is a mixture of Hindi words and Urdu words, so there are

> Persian and Sanskrit origins. But for Modi and his ilk...it's very important to "cleanse the language Hindi from foreign influences." That is why they will speak a very Sanskritized Hindi.

This shows not only a keen awareness of the strategic use of language, but also that it informs his understanding of politics and audience dynamics. It is not surprising, therefore, that this kind of strategic language use may be incorporated into his own work, which in large part comprises of convincing audiences of his views. How does this translate to the dynamics of his life and work over the years? Much of Zakaria's work revolves around American politics, and early on, he was keenly aware of being an outsider:

> I remember once being asked when I was a graduate student at Harvard— Tony Lake was then national security adviser, and his office called and said... "Mr. Lake would like you to come to the White House to brief him." I walked in and there were five people around the table: Tony Lake; Deputy National Security Advisor Sandy Berger; George Stephanopoulos, who was then director of communications at the White House; Joe Nye, who was a senior professor at Harvard; one other person; and myself. And I kept thinking to myself, "Are they going to realize at some point that I'm not an American citizen? They're asking me for my advice on what America should do, and I am on a student visa."

This gave him a strong motive (or pressure) to 'fit in', as seen in the quote below from an interview with comedian and television host Hasan Minhaj. In response to being asked whether "Fareed the young man had this desire to assimilate, to fit in, both naturally and due to your circumstances", he says:

> ...when I came to America I knew one human being in America— my brother [who] was also a scholarship student in college...I had no money because, you know, even if my parents were okay by Indian standards, the currency was unconvertible so you're kind of really on your own, right. And I don't think most Americans really understand—native born Americans—what it's like. You're seven thousand miles away from your country, your parents are by American standards poor, you know nobody here. Of course you're gonna try and fit in.

While this is a familiar dynamic for many immigrants in similar situations, it is particularly salient for Zakaria since, as the following exchange reveals, he perceived that 'fitting in' was essential for gaining legitimacy in his work, which he equates with not 'trading on identity':

> HM: Early in your career did you ever feel like you had to do that [assimilate] as a survival mechanism of like, 'you know what man I'm the only one [person of color] in the room...I really have to make sure my arguments are airtight'?
>
> Z: That's fair—look, I was an immigrant, I was trying to assimilate and...I was aware that not a lot of people are interested in...American people are not that interested where you come from...so yeah, I think...to be totally honest...there was a part of me that just wanted to fit in.
> It's a great question because I've thought about it a lot... when I started out...it was very important to me that people...that I was able to make arguments that anyone would either agree with or disagree with on the on the basis of the the value of the argument, not the identity of the person making the argument. And you know, I felt like–look I'm–I got a

> PhD at Harvard...I want to be seen as a great journalist...I don't want to trade on my identity.
>
> I genuinely felt like...I'm trying to meet everyone—I wanna meet you on neutral ground.

This adds support to the intuition that at the earlier timepoints (when he was early in his career and trying to establish himself as a journalist in a largely non-diverse profession), Zakaria might have had a strong incentive to accommodate to AmE norms in his D2. Specifically, this subjective account of his experiences and desire to 'meet people on neutral ground' would be consistent with a move to reduce linguistic markers of 'identity' that would make him stand out. I argue that similar (although not identical) dynamics also explain his IndE-like VOT norms in D1 at earlier timepoints. While the pressure to accommodate to an Indian audience would certainly differ in its social and professional implications for Zakaria, it is worth noting that his interviewer in the 2003 interview, Shekhar Gupta, was an eminent journalist and one of the prominent figures in political journalism in India. Zakaria was relatively early in his career and did not yet have an established readership in India, possibly creating a power differential. Moreover, it is likely that a young Zakaria was less secure in a stable cross-cultural identity, and therefore would have experienced more social pressure to be relatable. Finally, another aspect of this dynamic is that with increased globalization and mobility, cross-cultural identities (and their representations in media) are arguably more the norm in recent years compared to 2003. I propose that over the years, as Zakaria became more established in his career, presumably more secure in his individual identity, and addressing a more globalized Indian audience, his efforts to accommodate to IndE norms decreased, his VOT productions evidencing the intermediate targets that have resulted from phonetic drift.

In his D2 speech, though, the trend towards shorter VOTs (more D1-like norms) over the years cannot be explained merely by a reduced incentive to accommodate. This is because his default targets for D2, given his proficiency and exposure, would be expected to be AmE-like even in the absence of accommodation. Instead, his subjective experience of his identity and role as a public figure over time suggests that these patterns might stem from a conscious effort to adopt D1-like norms— introducing a linguistic signature of decidedly 'non-neutral' identity. In the same conversation with Hasan Minhaj, Zakaria describes how the rise of anti-Muslim and anti-immigrant sentiments in America in the wake of 9/11 shifted his approach towards his identity in relation to his work:

> After 9/11 I realized that I had some real special knowledge of having actually lived and breathed...in a Muslim community, you know, not just read about the religion but knew it intimately. And so I did occasionally...mention, right, [that he was Muslim]...it was a way of trying to express something that I felt people weren't getting.
>
> One time there was a couple of columns I did do...opening commentaries for the show where I did say very explicitly, 'I am a Muslim and I'm writing this as a Muslim' and it was when Trump was campaigning and was saying the— really the most nasty stuff about Muslims, and talking about a Muslim ban and things like that. I remember...at that point I decided to myself, look, I need to stand up and be counted because I want my viewers to know that...you're hearing all this stuff about Muslims but, you know, I am one...there was this weird feeling of like, I didn't want to not say this because it would seem like cowardice. It would seem like I was somehow ashamed of it...it was a very complicated feeling but I decided it was important to do. I got a lot of hate mail, you know. It was the only time I can think of...there was one other time after 9/11 when I got really nasty stuff that even made me worry...I still felt like I had to do it because it was...a way of making people aware that these vast characterizations of people might include people that they respect and admire...I wanted to stand up and be counted.

As research on phonetic accommodation as well as a large body of sociolinguistic research has shown, self-identification with a group affects an individual's patterns of speech with or without conscious awareness (Labov, 1963), and changes to social identity in adulthood is an important driver for linguistic change (Campbell-Kibler et al., 2014). In the same conversation, Zakaria also reveals his growing awareness of the iconographic significance of his identity as a person of color, an immigrant, and a Muslim in the journalism sector, and a conscious choice to lean into that role:

> I am very aware—and I've tried to be better about being aware—that I stand for something for a lot of people of color people who are, you know, even just immigrants—often Muslims, particularly South Asians... You know, whenever I give a talk what I'll notice is a lot of the people who mill around afterwards will look like me and people will say to me—young kids particularly— 'It is so great to see someone like you doing well'. So I'm very aware that I have a role as a role model and I try to be very conscious of that and honor it and encourage people...but I don't want to become a spokesman for an ethnic point of view

These glimpses into Zakaria's subjective experiences reveal that as Zakaria has grown older, he has consciously chosen to embrace his multicultural identity in his public life. This coincides with a stage of his career where he has already established himself in the field, which might make 'fitting in' a less important professional incentive. I propose that this combination of factors allows Zakaria to use his linguistic resources towards this newer identity creation, and this is reflected in the movement of D2 VOTs towards D1 norms over the years. Sharma (2018) uses the term *biographical indexicality* to describe an individual's use of their own personal linguistic history, and the audience's knowledge of it, as a way to perform a 'stripped-down, real me' stance by reverting to one's own first-learned variety at certain points in the conversation. Zakaria's movement towards D1 norms can be seen as a long-term analogue of this kind of *real me* persona creation. Another way to conceptualize this shift is that it doesn't target *D1* norms per se, but rather an intermediate value that signals a more globalized/multicultural identity, achieved by what Evans and Iverson (2007) term *accent softening.*

Ultimately, of course, it is also impossible to separate Zakaria's evolving goals from the broader cultural change in how personal identity and diversity are perceived in public and political conversation. While talking about his desire to fit in early on, Zakaria at one point reminds Minhaj that "by the way, starting out there there wasn't any advantage to doing it...people forget now but thirty, thirty five years ago...it didn't do you any good to say that, you know, you were Brown or Muslim or anything like that". Thus, not only might his social goals have changed over the 21 years examined in this study, but the indexical meaning and social significance of the varieties themselves have shifted as well, and Zakaria's changing speech reflects these various forces.

4 Vowels

In a subset of the lexicon, Indian English and American English differ in the vowel category that is used in that class of words. Some members of the BATH class from Wells (1982) are realized as /ɑ/ in IndE and /æ/ in AmE. Examples of such words are 'bath', 'class', 'dance', etc. Note that both dialects have both /æ/ and /aː/ as contrastive categories. Words like 'father' (PALM class) are produced with /ɑ/, whereas words like 'ant' (TRAP class) with /æ/, across both dialects. For simplicity, in this work I code the (between-dialect invariant) vowel in words like *father* and *ant* as AA and AE respectively, and code the (between-dialect variable) vowel in words like *bath* and *class* as vbath. The list of words coded as vbath (45 unique words) can be found in the Appendix. The phonology of the two dialects leads to the following expectation: vbath maps on to AE in AmE and to AA in IndE. Thus, if we hypothesize that Zakaria is drawing on his D2 system while in ame contexts, and on his D1 system in

Lexical set	Example words	Label	Notes
TRAP	tap, back, badge, scalp, hand, cancel	AE	
PALM	psalm, father, bra, spa, lager	AA	
BATH	staff, brass, ask, dance, calf	v_bath	Expected: realized as AE in ame and AA in ie
CLOTH	broth, cross, long, Boston	AO	
THOUGHT	taught, sauce, hawk, jaw, broad	AO	Patterns with the LOT class for speakers with a LOT-THOUGHT merger
LOT	stop, sock, dodge, romp, possible	v_lot	Expected: realized as AA in ame and AO in ie

Table 6: Transcription convention for relevant lexical sets

ie contexts, we expect the target realization for a vbath vowel to vary between contexts: the target is AE in the former case and AA in the latter.

In a parallel manner, a subset of the lexicon differs in the realization of the Well's class LOT vowels. THOUGHT, CLOTH, and LOT map to the same vowel in IndE (the rounded low-mid back vowel /ɔ/). In AmE, the THOUGHT and CLOTH classes map to the same rounded vowel, but the LOT class maps to the low back vowel /ɑ/ (the vowel of the PALM class). As before, I code the (between-dialect invariant) PALM vowel as AA, the between-dialect invariant vowel in the CLOTH and THOUGHT class as AO, and the (between-dialect variable) LOT vowels as vcot. The list of words coded as vcot (183 unique words) can be found in the Appendix. The realization of this vowel is expected to vary between dialects, such that it maps on to AO in IndE and to AA in AmE.

In some varieties of American English, the THOUGHT class has merged with the LOT class, both being produced as /ɑ/ (dubbed the *cot-caught merger*; Labov et al. (2006)). The CMU dictionary, which is used by DARLA, only makes a 2-way distinction between the round vowel AO and the low back vowel AA. The Wells classes THOUGHT and CLOTH are both transcribed as AO, whereas LOT (as well as PALM) is transcribed as AA. The CMU thus preserves a distinction between LOT and THOUGHT, but not between THOUGHT and CLOTH, representing the phonological system of speakers who maintain the THOUGHT-LOT distinction. But for speakers with merged THOUGHT-LOT, this transcription does not allow us to distinguish the class of words that they would produce as /ɑ/ (THOUGHT) from the class that they would still produce as /ɔ/ (CLOTH). If Zakaria's D2 system includes the merger, then another class of words (THOUGHT) is expected to behave exactly like the LOT class, varying between AA and AO based on the dialect system Zakaria is using. For the present analysis, I have retained the CMU convention and grouped THOUGHT and CLOTH vowels as one category, assuming that both of these map onto the rounded vowel /ɔ/. As mentioned, this category is labeled AO, whereas the LOT class is labeled as vcot. Table 6 shows the original lexical sets from Wells (1982), the transcription convention used in this work, and their expected realization in IndE and AmE. Before describing the methodology, I note that if Zakaria's D2 system is indeed a merged one, the coding decision for vcot in this study might have the following effects on the results: (i) the tokens labeled as between-dialect variable vcot here (LOT class) are a subset of the tokens that are actually treated as vcot by Zakaria (LOT + THOUGHT class); (ii) if that is the case, then the THOUGHT tokens will be coded as instances of AO, even though Zakaria actually treats them as vcot (between-dialect variable). This would add variability to the so-labelled AO category, making it potentially extend to a lower part of the vowel space. As the results will show, this does seem to be the case, suggesting that Zakaria's D2 is a merged system. I return to this in the discussion.

4.1 Methodology and hypotheses

I first extracted all tokens of AE, AA, and AO from phone-annotated textgrids, and manually relabeled instances of the variable tokens as vbath or vcot as described above. I then used the first and second formant frequencies (F1 and F2) from the vowel midpoint, as measured by DARLA, for statistical analyses. There were 192 tokens of vbath and 524 tokens of vcot. Table 7 shows the condition-wise

Context/ Vowel	ame					ie				
	total	1	2	3	4	**total**	1	2	3	4
vbath	**83**	11	60	6	6	**109**	27	26	15	41
vcot	**261**	53	120	42	46	**263**	44	51	28	140

Table 7: Vowel token counts in each context and timepoint (columns 1-4)

token counts of these variable vowels. As is immediately noticeable, the token counts are uneven, with some cells having very low counts. Since the counts for the variable tokens are constrained lexically, there is no way to address this aside from using much larger samples of speech. To avoid spurious results in the present study, I limit the statistical analyses to comparisons between the ame and ie contexts, and only provide a descriptive account of the changes over time. Follow-up work should both verify these findings and carry out statistical analyses for the timepoint effects with a larger amount of data. Since the vowel space is multi-dimensional and the possible patterns of change more complex than those for VOT, I use converging evidence from a variety of visualizations, measurements, and statistical analyses to interpret the findings and inform the discussion.

4.1.1 Reference categories and difference measures

The vowel categories of interest in this study are vbath and vcot. We expect that the former varies along the AE—AA continuum, and the latter varies along the AA—AO continuum. The phonological features that distinguish these vowel categories are height and backness. In their phonetic realization, this corresponds to a difference in vowel quality, whose main acoustic correlates are the first and second formant frequencies (F1 and F2). The higher a vowel, the lower its F1, and the more backed a vowel, the lower its F2. While both vowel under vary along both height and backness dimensions, the AE—AA contrast is primarily one of backness, whereas the AA—AO contrast is primarily one of height. To simplify the analysis and reduce noise, I focus only on the primary dimension of contrast for each category in the statistical analyses, reporting F2 variation in vbath and F1 variation in vcot. For the purpose of this variation, I refer to the categories AA, AE, and AO as 'reference categories', reflecting the assumption that in prototypical AmE and IndE, these are the targets for vbath and vcot, as discussed above.

AA is more backed compared to AE, and therefore will have lower F2 values. Thus, we expect that a vbath token produced in an ie context will have lower F2 (closer to AA) than a token produced in an ame context (closer to AE). AO is higher than AA, and will therefore have lower F1 values. We expect a vcot token produced in an ie context to have lower F1 values (closer to AO) compares to a token produced in an ame context (closer to AA). I elaborate on these measurements in the next section.

A non-trivial question at this point is to determine what counts as the 'reference' F1 and F2 values, i.e., the prototypical positions of the reference vowels. One option might be to use expected values corresponding to both varieties from existing literature. However, it is well-known that vowel realizations vary across individual speakers. Therefore, using a speaker's typical/average realization of a category as an estimate of their target is a more ecologically valid option. Here, the working hypothesis is that Zakaria is navigating between two vowel systems, D1 (which corresponds to IndE norms) and D2 (which corresponds to AmE norms). Moreover, his speech is sampled at different times in his life. There is a considerable body of research showing that phonological systems change over time, both for individuals and speech communities. Since the aim in this study is to examine

Zakaria's speech in eight 'conditions', I work with the hypothesis that at each point, he is accessing a certain vowel space, which is characterized by typical realizations for each category, including AA, AE, and AO, our reference vowels of interest. Thus, each time he produces one of the variable tokens, e.g. vbath, there is potential to produce it as though it were an instance of AA, or an instance of AE. The phonetic realization of these would be whatever is the prototypical/target realization of AA or AE *in the system that he is currently operating in.* To reflect this, I take the relevant reference categories for a given vbath token to be the average F2s of AA and AE in that condition, and the reference categories for a given vcot token to be the average F1s of AA and AO realizations in that condition. Using these condition-specific reference categories, I calculated where each token falls on the continuum between its two possible realizations by calculating the F2 difference (for vbath tokens) or F1 difference (for vcot tokens) from each of its two references. This led to two difference measures for each token: for vbath tokens, these are diff.f2.ae (F2 distance form reference AE) and diff.f2.aa (F2 distance from reference AA), whereas for vcot these are diff.f1.aa (F1 distance form reference AA) and diff.f1.ao (F1 distance form reference AO). The statistical analyses were based on these difference measures. In the rest of the section, I describe how these measures relate to the position of the token in the vowel space, and the specific hypotheses regarding each measure.

4.1.2 Diff measures and position in the vowel space

Since I only compare tokens in a given condition to the reference vowels in that same condition, I will simply use the term 'reference vowel(s)', leaving implicit that it is the reference for the specific condition in which the token was produced. There are three reference categories: AA, AE, AO. For both F1 and F2 measures, I always subtract the reference F1/F2 value from the token's F1/F2 value. Recall that front vowels have higher F2 values than back vowels. Thus, if a token is more fronted than the reference, the diff.f2 measure will be positive, whereas if the token is farther behind, the diff.f2 measure will be negative. Comparing between two tokens, the one with the higher diff.f2 value will be more fronted in the vowel space. Moreover, the closer a token is to the reference, the smaller the absolute value of its difference measure. The higher a vowel, the lower its F1. Thus in a parallel fashion, if a vcot token is lower than a reference category, the corresponding diff.f1 value will be positive. If the token is higher than the reference, the resulting diff.f1 will be negative. Comparing between tokens, the one with a higher diff.f1 value will be lower in the vowel space. As with F2 measures, the closer a vowel is to the reference, the smaller the absolute numerical value of the difference measure. Figure 7 shows a simplified schematic of how the magnitude and sign of the diff values corresponds to the position of a vbath token relative to its two references.

4.1.3 Calculating distance from reference categories

The aim of the analysis is to compare diff values across conditions, with the hypothesis that the distance from AE should reduce in the ame context, whereas the distance from AA should reduce in an ie context. One option is to fit separate models for each of the measures diff.f2.ae and diff.f2.aa, expecting the context to predict these values in the way outlined above. However, this leads to a potential confound: recall that the references for each token are taken from the condition in which the token was produced. Therefore, the total distance between the two references, e.g. AA and AE, is not identical across all eight conditions. In effect, this means that the possible range of diff.f2.ae values could differ across conditions, and a condition where the two references are farther apart would have a larger absolute value for both diff measures. Modeling these two diff measures separately would therefore be misleading. Figure 7 schematizes an example where two tokens vbath1 and vbath2 that are both midway between AA and AE with respect to their respective references artificially look

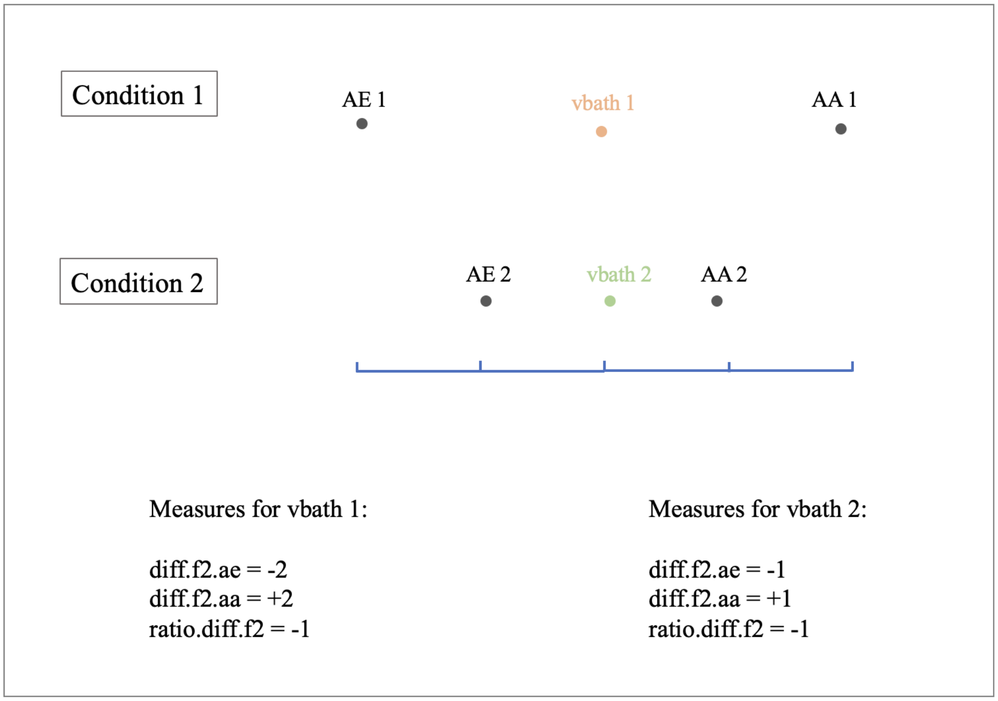


Figure 7: Schematic diff measures for vbath tokens in two conditions (1 and 2), which differ in the distance between the reference vowels AE and AA

different when we consider just one of the diff.f2 measures. We would mistakenly conclude that the former is more backed compared to the latter.

A condition-wise plot of the distance between the reference AE and AA, and between AA and AO, shows that these are not identical across the eight conditions; the plots are included in the Appendix (figures 18, 19). Recall that for each variable token the two diff measures (distance from each reference) are not independent of each other– as a token moves farther back from AE, it must also become less and less fronted compared to AA. To avoid the confound just discussed, I ran analyses that combined these measures, first by modeling them together, and then by combining them into a single ratio measure. I ran two statistical analyses, described below along with their specific hypotheses:

i) Simple diff measures: I treated each diff measure as an independent datapoint, and treated the reference vowel as a predictor in the model. That is, instead of two models of the following forms: $diff.f2.aa \sim context + timepoint$ and $diff.f2.ae \sim context + timepoint$, I fitted a single model of the form $diff.f2 \sim context + timepoint + reference$. Since reference is sum-coded, the model computes the effect of context (and timepoint) on the grand mean of the diff values from both references. Since a larger/smaller initial distance between AA and AE would make both diff.f2.ae and diff.f2.aa proportionately larger/smaller, the average is not affected. Therefore, using this measure neutralizes the effect of reference-vowel differences between the conditions. Moreover, this approach leads to increased statistical power due to more datapoints, and avoids the need to fit multiple models per variable vowel.

Hypothesis: As a vbath token moves farther back from AE (and thus closer to AA), its diff.f2 value relative to AE becomes more negative, whereas the diff.f2 value relative to AA becomes less positive. Thus, the numerical value of both measures decreases. Thus, we expect the context to predict diff.f2, such that these values are lower in the ie context compared to the ame context. As a vcot token moves away from AA (and thus closer to AO), its diff.f1 value relative to AA becomes more negative, whereas the diff.f1 value relative to AO becomes less positive, in both cases decreasing the numerical value. Thus, we expect diff.f1 values to be lower in the ie context compared to the ame context.

ii) Ratio diff measure: One way of restating our informal hypothesis is in terms of the relative value of the two distances. That is, when a vbath token moves away from AE (and therefore closer to AA), it's distance from AE *relative to its distance from AA* increases. It should be clear that representing distance in this way immediately removes any effect of differences in the total distance between AE and AA, since each distance is now relativized to its distance from the other reference. This can be verified from figure 7: X and Y have the same ratio.diff value even though the individual distances from AE and AA differ. I calculated the ratio.diff measures using the numbers diff.f2.aa/diff.f2.ae for vbath, and diff.f1.ao/diff.f1.aa for vcot. The farther back a vbath token moves from AE, and the farther up a vcot token moves from AA, the larger the absolute value of the respective ratios. Moreover, the sign of the ratio is informative about the position of the token: a negative ratio means that the two measures have opposite signs, and therefore the token is positioned between the two references. A positive ratio means that the measures are either both positive or both negative, showing that, e.g., the token is even more fronted than AE, even more backed than AA, etc. In keeping with the idea that the references serve as underlying targets for the variable tokens, I call these realizations instances of 'overshooting' the target. Since the ratio measures are less straightforward to interpret numerically, I refrain from using these in statistical models, and instead use them for plots and descriptive statistics.

Hypothesis: For both vbath and vcot, the absolute values of the ratios are expected to be larger in the ame context compared to the ie context.

4.2 Results

Figure 8 depicts Zakaria's vowel space in the ame and ie contexts (pooled across timepoints), with the dots representing the realization of individual tokens, and the labels representing the average position of the category in the F1-F2 space. The overall shape and size of the vowel space is fairly similar across the contexts as expected. However, the average positions of his variable tokens differ: visually, the vbath category in the ame context is closer to AE, compared to the vbath category in the ie context. Moreover, while the vcot category in both contexts is very close to AA, in the ame context it appears to be even lower than non-variable AA. In the ie context, by contrast, it is higher than the non-variable AA category. To look more closely at these categories of interest, figures 9 and 10 depict vowel ellipses for the variable vbath and vcot tokens respectively, along with their respective reference categories. This allows us to observe not just the mean position of the category, but also the extent of spread, or variability, in its realizations. We see that for the vbath category, the reference AA and AE categories are fairly 'tight' (less variable), whereas the vbath realizations are more variable, as reflected in the larger ellipse size. Moreover, whereas the ellipse for vbath is nearly overlapping the AE category in the ame context, in the ie context it overlaps equally with both references, suggesting that the canonical realization is midway between the two references. Figure 10 shows that the ellipses for vcot, AA, and AO look fairly similar across contexts, the main difference being in the height of the variable vcot category: it is higher in the ie context compared to the ame context. In both contexts, however, vcot overlaps more with the AA category. Moreover, note the the AO category shows high variability (large ellipse sizes) and extensive overlap with AA, in both contexts. As mentioned earlier, this likely results from the labeling conventions of the CMU dictionary which was used for the annotation of the data here. Specifically, the observed mean position of the AO category is likely lowered by the presence of some 'vcot' tokens that were (mis)labeled here as AO. Two points are important for interpreting our

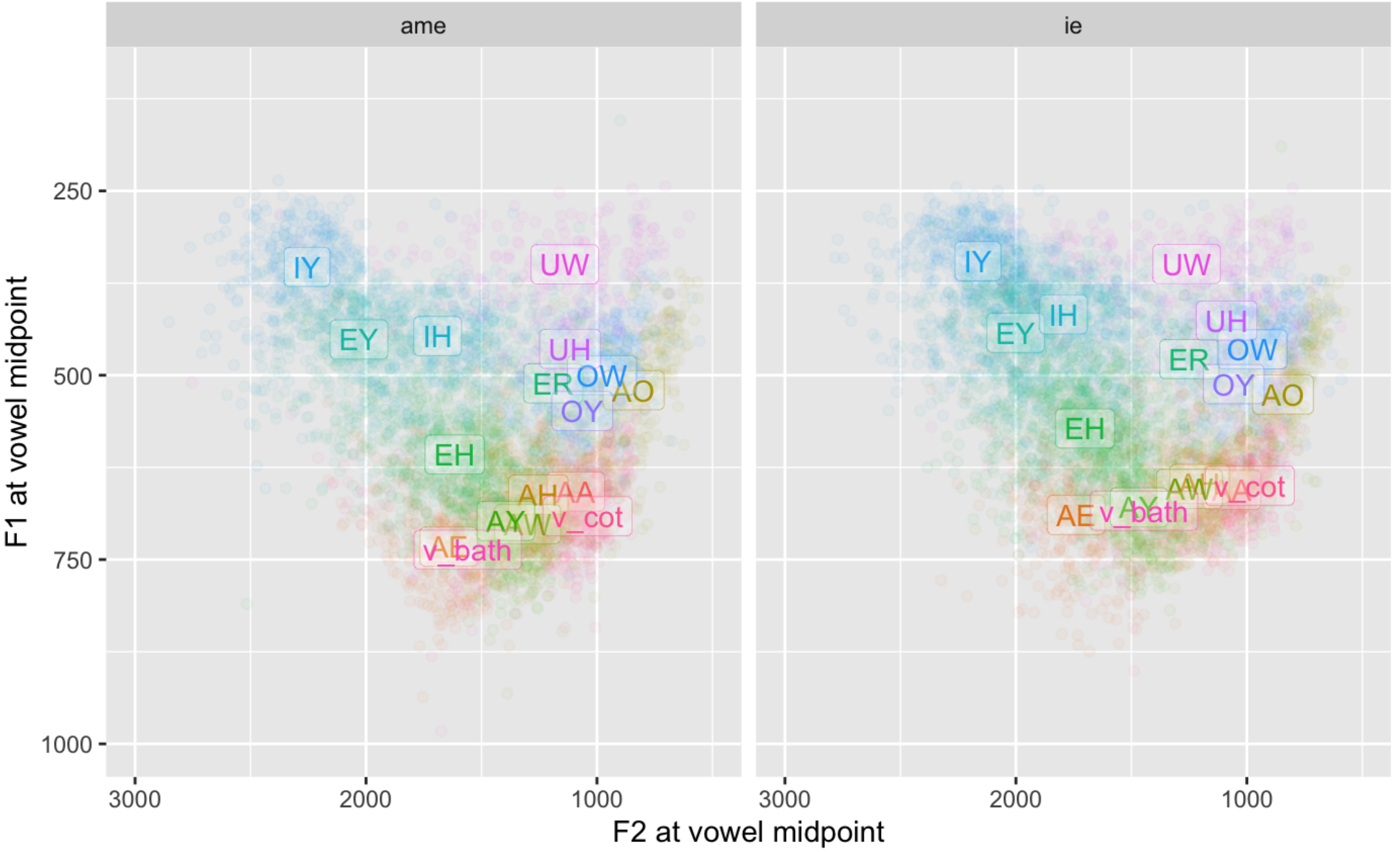


Figure 8: Zakaria's vowel space with American vs Indian audiences

results: (i) the figures show that Zakaria's vcot vowel is closer to AA than to AO. This means that if we were able to remove the effect of this coding artifice, making the AO category higher, it would only serve to exaggerate the distance between vcot and AO. The main finding, therefore, would not change. (ii) Moreover, we see that this coding artifice appears to have affected the AO category *in both ame and ie contexts*. Since prototypical IndE makes no distinction between the THOUGHT and CLOTH classes (both are rounded), this emphasizes the finding that Zakaria's vcot category behaves in an AmE-like manner in both contexts.

To see how these visualized patterns are reflected in the primary dimension of contrast, we turn to the F1 and F2 distance measures. I present the findings for vbath (F2 distance) first, followed by vcot (F1 distance). For all statistical models, the independent variables and their levels, as well as the random effects structure, are identical to the models for VOT reported in section 3.3, and are not repeated here. As mentioned earlier, I do not include timepoint in the statistical analyses due to low token counts in several cells.

4.2.1 vbath

Simple diff measure: Figure 11 shows the F2 distance from AE and AA for the variable vbath tokens. A value of 0 indicates a realization identical to the reference, negative values indicate a more backed realization, and positive values a more fronted realization. As expected, all tokens are more fronted when compared to AA (a back vowel) than when compared to AE (a front vowel), reflected by higher diff.f2 values when the reference is AA. In spite of this uniformity, there is a clear effect of

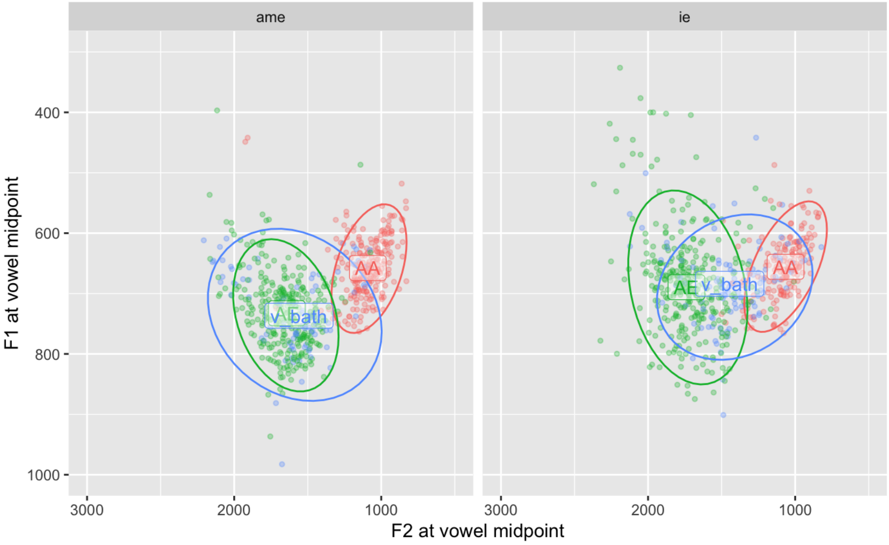


Figure 9: Vowel ellipses: vbath, AE, AA

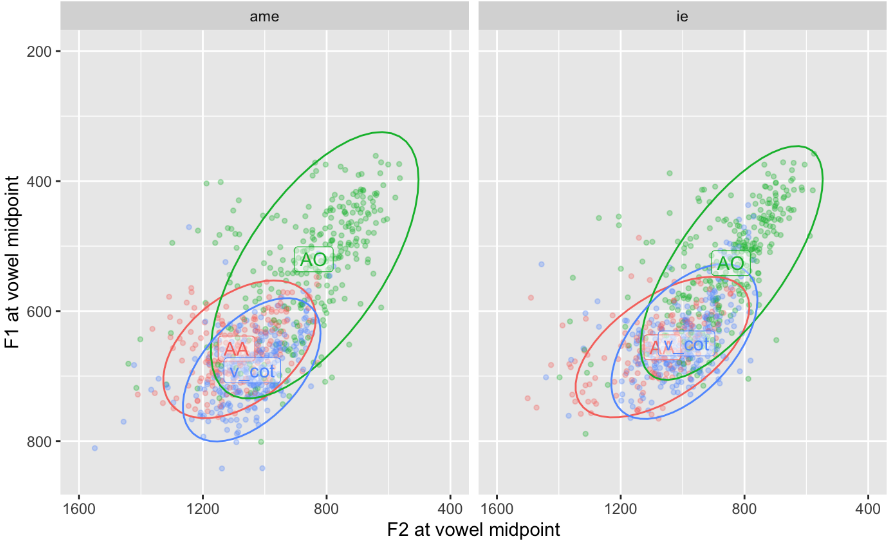


Figure 10: Vowel ellipses: vcot, AA, AO

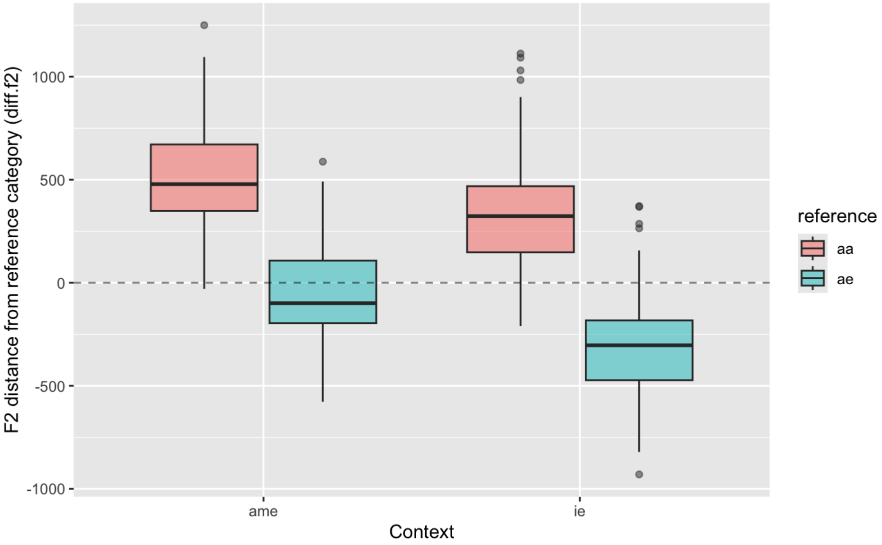


Figure 11: vbath realizations: F2 distance from reference categories

Fixed effect	Estimate (β) and Std. Error
Reference-AA	306.400*** (6.954)
Context-ame	88.794*** (8.991)
Intercept	123.323*** (33.675)
Observations	384
Akaike Inf. Crit.	4,991.752
Bayesian Inf. Crit.	5,011.505
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Model output: F2 distance for vbath tokens

context: while tokens in both contexts are fronted compared to AA, this effect is larger in the ame context. Conversely, while tokens are generally backed compared to AE, this effect is smaller in the ame context, with several tokens overlapping, and some even more fronted than, AE. Overall, this shows that vbath tokens are more fronted (and therefore closer to AE) in the ame context, as hypothesized. If this effect is statistically significant, we expect Context to be a significant predictor of diff.f2, with diff.f2 decreasing as we move from the level ame to the level ie. This is confirmed: compared to a null model that predicts diff.f2 from the reference vowel ($diff.f2 \sim Reference + (1|Word)$), adding Context as a predictor significantly improves model fit ($\chi^2(1) = 84.92, p < 0.001$). The model summary (table 8) confirms that diff.f2 is around 88.8 Hz higher than average in the ame context.

Diff ratio measure: Figure 12 shows, for each vbath token, the ratio of its distance from AA to its distance from AE (diff.f2.aa/diff.f2.ae). An absolute value of 1 indicates that the token is equidistant from both references, an absolute value smaller than 1 indicates that it is closer to AA, and larger than 1, that it is closer to AE. Recall that the sign of the ratio indicates whether the token lies between the two references (negative sign), or in one of the extremes (positive sign).

For ease of visualization, I removed tokens whose ratios were outside the range of -20 to +20

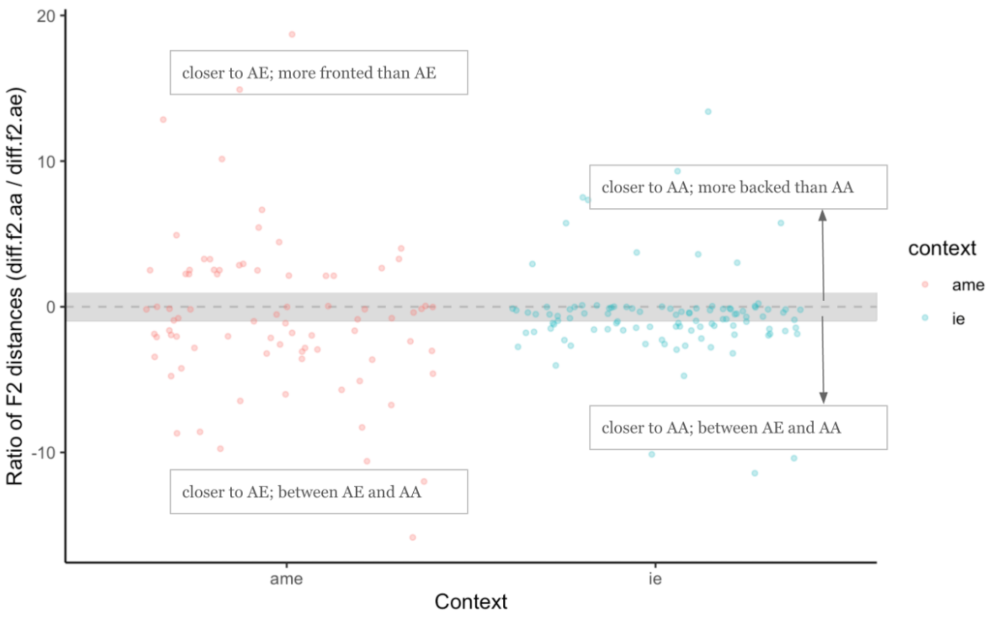


Figure 12: vbath realizations: F2 distance from AA relative to AE (diff.f2.aa/diff.f2.ae); the gray box shows values between -1 and 1

(4.7% of total tokens).[4] The grey shaded box indicates the area between -1 and +1; tokens inside this box have absolute values <1 (are closer to AA). The plot shows that (i) both contexts have tokens that are closer to AA as well as tokens that are closer to AE; (ii) however, tokens in the ie context are clustered around the grey box, whereas many tokens in the ame context are more spread out, indicating more realizations that were closer to AE; (iii) the ame context has a greater number of tokens with positive ratios outside the box, indicating tokens that were realized as even more fronted than the reference AE; (iv) there are virtually no tokens that have positive ratios inside the box, indicating that no vbath tokens were produced as even more backed than reference AA, regardless of the context. These observations confirm that the patterns in vbath observed from the vowel-space plots are straightforwardly reflected in the primary dimension of contrast, F2.

4.2.2 vcot

Simple diff measure: Figure 13 shows the F1 distance from AO and AA for the variable vcot tokens. A value of 0 indicates a realization identical to the reference, negative values indicate a higher realization, and positive values a lower realization. As expected, all tokens are higher when compared to AA (a low vowel) than when compared to AO (a mid vowel), reflected by lower diff.f1 values when the reference is AA. In spite of this uniformity, we again see an effect of context: while tokens in both contexts are lowered compared to AO, this effect is larger in the ame context. Conversely, while tokens are generally very close to the AA reference in both contexts, they are generally higher than AA in the ie context. In the ame context, however, almost all the tokens are produced as even lower than AA.

[4] There are some tokens that nearly perfectly overlapped with AE, so that the distance from AE is close to 0. Since the ratio is calculated by dividing the distance from AA by the distance from AE, the ratio value for these tokens is close to infinity. Including these makes it difficult to see the patterns, since the y-axis now has a large range. I therefore removed tokens with ratio values outside -20 and 20 from the graph (i.e., tokens that are more than 20 times closer to AE than to AA). This does not affect the interpretation. The same applies to the graph for vcot (figure 14), where a sizable number of tokens were very close to AA, giving high ratio values.

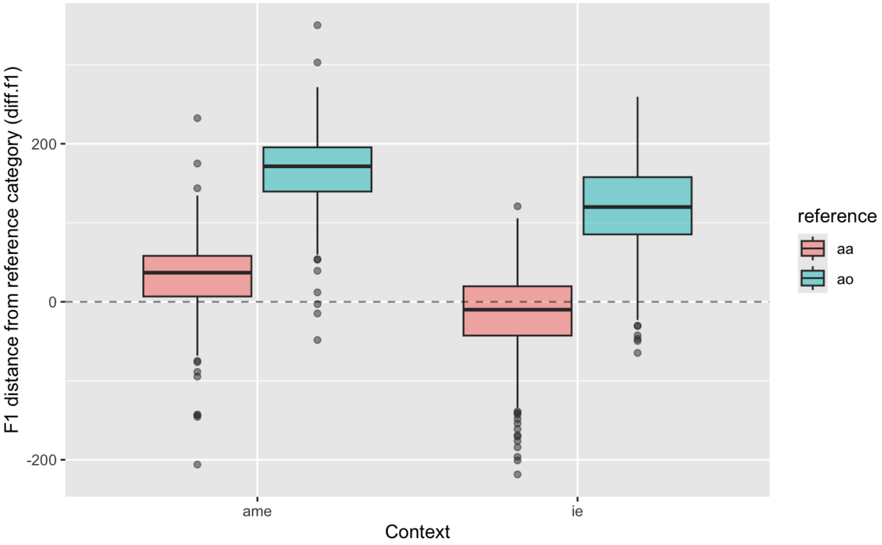


Figure 13: vcot realizations: F1 distance from reference categories

Fixed effect	Estimate (β) and Std. Error
Reference-AA	-66.758^{***} (1.167)
Context-ame	23.948^{***} (1.492)
Intercept	72.694^{***} (3.986)
Observations	1,048
Akaike Inf. Crit.	10,958.280
Bayesian Inf. Crit.	10,983.050
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Model output: F1 distance for vcot tokens

Overall, this shows that vcot tokens are tokens are lower (and therefore farther away from AO) in the ame context. If this effect is statistically significant, we expect Context to be a significant predictor of diff.f1, with diff.f1 decreasing as we move from the level ame to the level ie. This is confirmed: compared to a null model that predicts diff.f1 from the reference vowel ($diff.f1 \sim Reference + (1|Word)$), adding Context as a predictor significantly improves model fit ($\chi^2(1) = 228.9, p < 0.001$). The model summary (table 9) confirms that diff.f1 is around 24 Hz higher than average in the ame context.

Diff ratio measure: Figure 14 shows, for each vcot token, the ratio of its distance from AO to its distance from AA (diff.f1.ao/diff.f1.aa). As before, an absolute value of 1 indicates that the token is equidistant from both references, an absolute value smaller than 1 indicates that it is closer to AO, and larger than 1, that it is closer to AA. Recall that the sign of the ratio indicates whether the token lies between the two references (negative sign), or in one of the extremes (positive sign).

For ease of visualization, I removed tokens whose ratios were outside the range of -20 to +20 (9.2% of total tokens). The grey shaded box indicates the area between -1 and +1; tokens inside this box have absolute values <1 (are closer to AO). The plot shows that (i) most tokens across both contexts are closer to AA than to AO; (ii) however, among the tokens that are closer to AO (in the grey box),

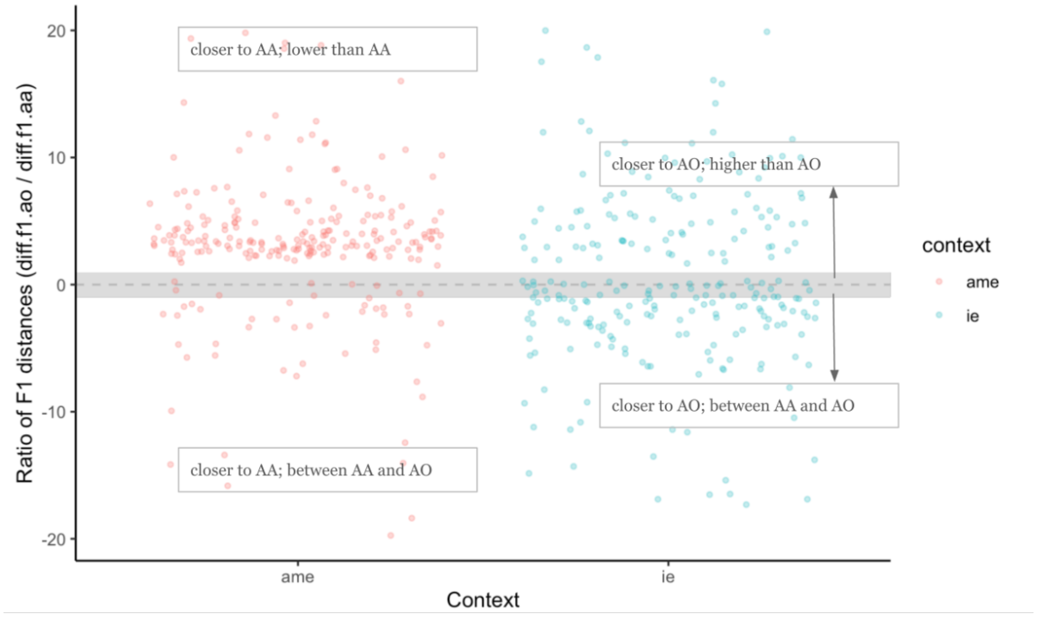


Figure 14: vcot realizations: F1 distance from AO relative to AA (diff.f1.ao/diff.f1.aa); the gray box shows values between -1 and 1

the majority were produced in an ie context; (iii) focusing on the tokens that are closer to AA, the tokens produced in the ie context are evenly spread in the positive and negative directions, indicating that they cluster around the reference AA category. In the ame context, however, there is a strong skew: most tokens have positive ratios, indicating that they are produced as even lower than AA. This pattern is statistically confirmed: a model comparing F1 at vowel midpoint of vcot and AA tokens shows that in the ame context, vcot realizations are significantly lower than AA (by 25.7 Hz), whereas in the ie context they are higher (by 13.3 Hz). Table 10 reports the model output summary. Once again, these observations confirm the patterns seen in the vowel-space plots. Moreover, they show consistent acoustic differences in Zakaria's speech between the two contexts, even when these might not result in categorical differences (e.g. because the variable token 'sounds more like AA than like AO' in both cases).

4.3 Vowel realizations over time

Figure 15 and 16 show vowel ellipses of Zakaria's variable vowels (along with the reference categories) in the ame and ie contexts across time. As noted earlier, I have not fitted statistical models for these timepoint-wise distributions. The ellipses for vbath show changes in both contexts, mainly in the variability of realization (size of the ellipse). The mean position of the category in the F1-F2 space remains the same, suggesting a stable target. This target realization appears to be overlapping AE in the ame context, and midway between AE and AA in the ie context.

For the vcot tokens, the patterns are similar in that the main changes are in the ellipse size rather than the mean position. Again, this suggests that his target realization is stable over time— almost overlapping AA in the ie context, and lower than AA in the ame context. An interesting pattern in this set of vowels is that his AO category which, as noted earlier, was is highly variable in both contexts, also appears to be changing in shape and size over time, especially in the ie context. In spite of this, the AA and vcot categories remains stable, both in their spread and in their relative positions, in both contexts. Finally, the ie system at timepoint 4 is visually nearly identical to his ame system

	F1
Vowel-vcot	25.711^{***} (6.923)
Context-ie	-7.590 (5.293)
Vowel-vcot:Context-ie	-39.017^{***} (6.899)
Intercept	660.860^{***} (5.332)
Observations	899
Akaike Inf. Crit.	9,610.315
Bayesian Inf. Crit.	9,639.122
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Height (F1) of the vcot category relative to AA across contexts

at that timepoint.

4.4 Discussion: vowels

The analysis of Zakaria's vowels shows that: (i) just like his VOT, Zakaria appears to maintain dialect-specific targets for his vbath and vcot vowels— his target realization for vbath is consistently more fronted, and that of vcot lower, while in an ame context; (ii) across both contexts, however, the overall positions of the categories adhere more closely to an AmE prototype than to an IndE prototype— vbath is closer to AE than to AA, vcot is closer to AA than to AO. Specifically, vbath overlaps AE in ame contexts and is midway between AE and AA in ie contexts, whereas vcot is significantly lower than AA in ame contexts and marginally higher than AA in ie contexts; (iii) unlike VOT, these dialect-specific vowel realizations remain stable over time for the time-period examined here.

We can understand differences between any two kinds of varieties (languages, dialects, socially meaningful styles) in two ways: (i) system-internal: the position of a category relative to other categories in the system. E.g., the vbath category in IndE is expected to overlap with the AA category, but the vbath category in AmE is expected to overlap with the AE category. (ii) across systems: when the same individual commands multiple varieties, we can ask how the realization of corresponding categories will compare across their two systems. E.g., for a given individual who commands both IndE and AmE, we expect that their vbath realizations while speaking AmE will be more fronted than their vbath realizations while speaking IndE. (i) automatically results in (ii): since AE is a front vowel and AA is a back vowel, an individual who produces prototypical realizations of vbath in each variety will automatically have a more fronted realization in their AmE. However, the converse does not hold: realizing vbath as more fronted while speaking AmE does not guarantee that it overlaps with AE. In Zakaria's vowel patterns these two aspects of cross-dialect difference come apart: we see (ii) but not (i). Specifically, his vowel realizations overall adhere more closely to a prototypical AmE system than to a prototypical IndE system regardless of the context. However, he maintains a stable difference between contexts, which mirrors the expected differences between the dialects.

An explanation of these patterns, therefore, must address the following:

1. The existence of separate categories across dialect-contexts

2. The within-system position of each category (vbath overlaps AE in ame contexts and is midway between AE and AA in ie contexts; vcot is lower than AA in ame contexts and marginally higher than AA in ie contexts): why do the realizations overall resemble the prototypical AmE vowel system more, in both dialect contexts?

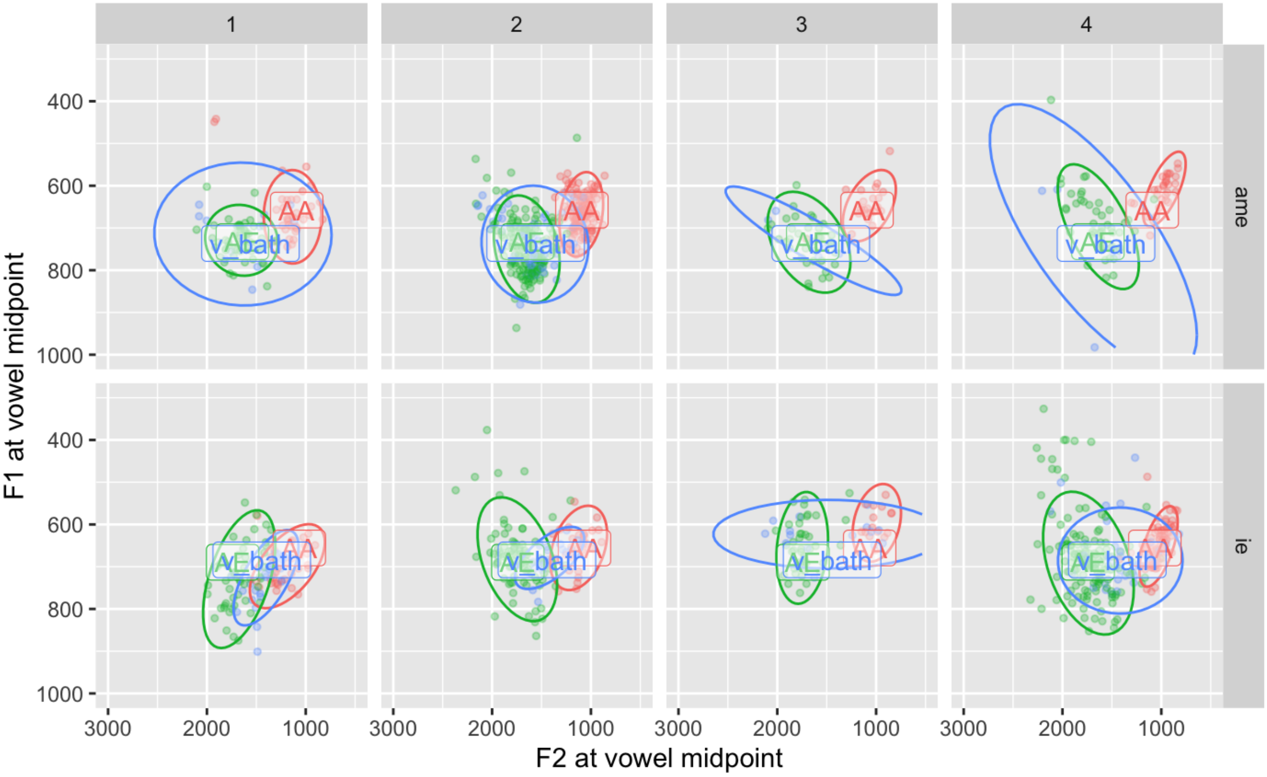


Figure 15: vbath realizations in ame and ie contexts across time

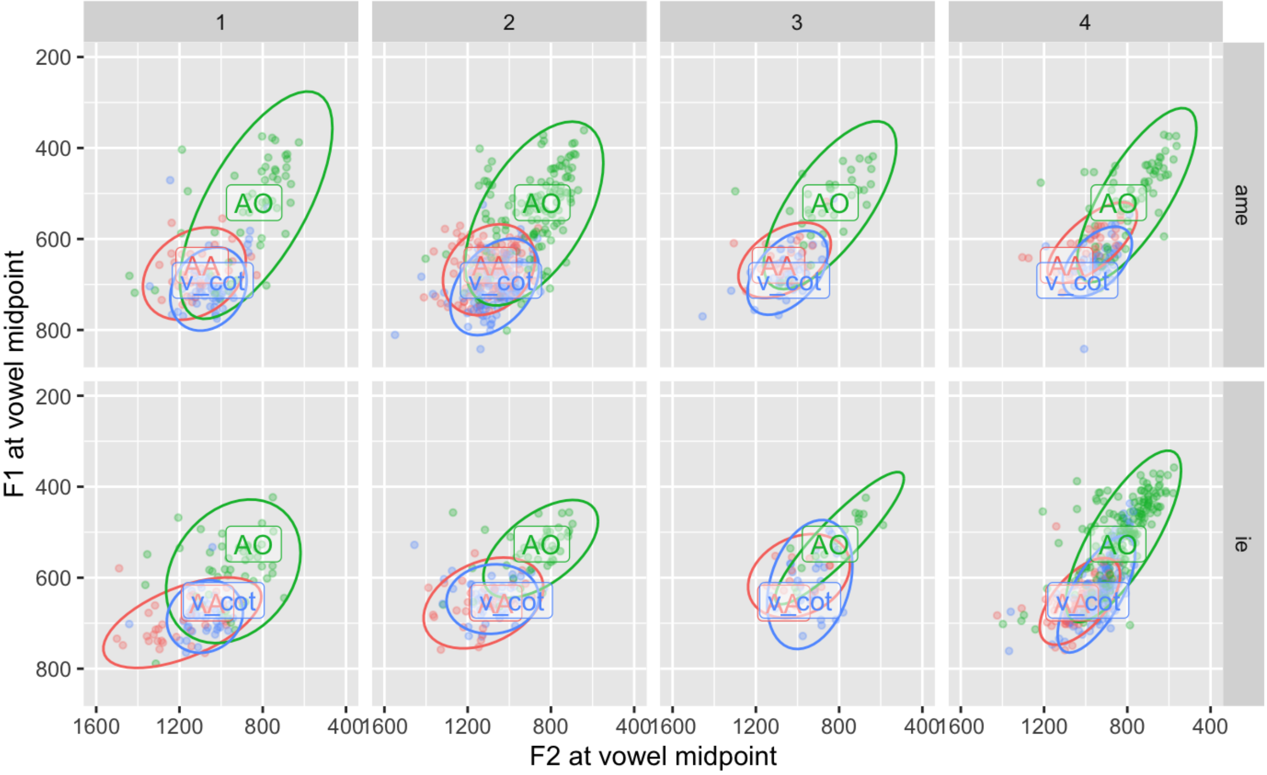


Figure 16: vcot realizations in ame and ie contexts across time

3. The across-system position of corresponding categories (vbath in ame contexts is more fronted than vbath in ie contexts; vcot is lower in ame contexts than in ie contexts): why do the differences between contexts always mirror the expected differences between IndE and AmE?

4. Why these target realizations appear to remain stable over time, unlike the targets for VOT

Consistently distinct realizations across contexts suggests that Zakaria maintains separate, dialect-specific targets for the vbath and vcot vowels. This mirrors well-established findings from bilingual speakers whose languages differ in their vowel categories. Researchers have consistently reported that proficient bilingual speakers are able to create and maintain separate vowel targets across languages (Simonet, 2016). Categories across varieties can differ in two ways: in the category boundary (whether a particular token would be judged as an instance of category X or category Y), and category goodness (how good/ideal an example of category X a particular token is). Zakaria's differences in production might be attributed to either kind of difference, and I refer to them jointly as 'category differences'.

Zakaria's vowel realizations in ame contexts show that at least for these vowels, he has acquired his D2 norms and is able to produce AmE-like targets for the appropriate lexical items. Given that his VOT realizations also evidence proficiency in D2, this is not surprising. This also mirrors existing findings that speakers can acquire lexically-specific vowel features in a second dialect even in adulthood, e.g. Sankoff (2004). However, what is unexpected in light of the VOT data is that his vowel productions in the ie contexts also evidence AmE-like norms. Recall that Zakaria is able to maintain IndE-like norms in his VOT, and that particularly at the earlier timepoints, he uses these in ie contexts. Why do his vowel realizations differ? I will propose that the role of the lexicon in mediating phonetic influence across varieties is at play. Specifically, the relevant difference between these vowel features and the VOT feature is that the between-dialect vowel changes apply only to certain lexical items, whereas for VOT the between-dialect changes apply across-board. I suggest that acquiring such lexically-specific D2 phonetic features affects the lexical representations of these words, in a way that acquiring the D2 VOT values does not. I elaborate on this below.

The concept of the phoneme as a contrastive unit is tied to the lexicon— to acquire a phoneme in a second language, for example, is to learn some phonetic information as well as the list of words in which that category occurs (Simonet, 2016). In the context of a single language, lexical representations are thought to comprise of grammatical (lemma) and morpho-phonological (lexeme) information. Representations of phonologically similar words 'share' phonemic content, evidenced by the fact that word naming is faster when a speaker sees/hears phonologically similar words (Glaser and Düngelhoff, 1984; Lupker, 1979), and greater phonological overlap leads to a larger facilitation effect (Abdel Rahman and Melinger, 2008). Studies with bilingual speakers show that just like lexical items within a single language, phonologically similar words *across* languages also compete with each other for selection during word-recognition tasks (Dijkstra, 2009; Spivey and Marian, 1999; Marian and Spivey, 2003a,b). This suggests that lexical representations from both varieties can share phonological information. Because of this, cross-language sound interactions are mediated by the structure of the lexicon (Simonet, 2016).

What makes words across languages phonologically 'similar'? This has been the subject of a large body of research in bilingualism. Across different models, the two main ideas are: (i) for some categories, there is a 1-1 correspondence between varieties, e.g. Spanish /e/ and Catalan /e/; (ii) categories that are not phonetically identical across varieties, but similar enough (acoustically/gesturally/perceptually) are associated/linked across languages (Flege, 1995; Best and Tyler, 2007), e.g. Spanish /e/ and Catalan /ɛ/. Regardless of language, then, lexical items with the 'same' or associated phonemes are related through phonological similarity in the lexicon.

The most interesting aspect of this research for our discussion is that lexical representations appear to involve not only phonemic (category) information but also fine phonetic detail. This is evidenced by a fairly extensive body of work showing that category updates, i.e. changes to the phonetic details of a category, affect subsequent lexical recognition (McQueen et al., 2006). In other words, exemplars can influence the category boundary (Norris et al., 2003), as well as the category goodness (Xie et al., 2017), of phonological categories, and this kind of exposure-driven shift affects the phonetic information in lexical representations. Moreover, such perceptual retuning affects subsequent productions, and those changes persist over time (Bradlow et al., 1997, 1999; Rvachew and Jamieson, 1989; Rochet, 1995).

Recall from section 1.1 that word-recognition tasks in bidialectal speakers suggest that dialects have separate lemmas, even for phonologically similar words. While using a word from one dialect, the corresponding lemma in the other dialect gets activated (due to shared semantic content), which in turn activates the phonological representations. Thus, the corresponding lexemes across dialects are co-activated. As just discussed, phonological representations are affected by exemplars. Since Zakaria lives in the US, he is exposed to many exemplars of front/low vbath/vcot vowel realizations. Due to this imbalanced input, his D1 lexemes for vbath and vcot words are co-activated upon hearing an AmE production much more often than the converse (co-activation of D2 lexemes upon hearing IndE productions). Eventually, this frequent association with AmE-like productions might affect the representation of the D1 lexeme, resulting in phonetic targets that are more AmE-like.

A reasonable question at this point is why this process should not apply equally to VOT— as Zakaria is likely exposed to many exemplars of long-lag VOTs, why do lexical associations not affect the VOTs of words in his D1 system, making the targets more AmE-like? I propose that this is because the VOT difference across the dialects does not make reference to specific lexical items. While there is evidence that exemplars affect VOT (Amengual, 2012), there is at least a way to conceive of this feature without reference to lexical information. For example, a certain VOT range might be associated with a dialect as a whole, or derivable by a simple rule (e.g. 'lengthen VOTs for AmE'). Thus, the co-activation of a D1 lexeme upon hearing the corresponding AmE word with a long VOT need not result in a change to the D1 lexical representation. By contrast, the vowel differences between the dialects cannot be derived by a parallel rule such as 'realize instances of AA as AE for AmE', because not every instance of AA in IndE is realized as AE in AmE— this only applies to a certain set of words. Thus, the feature itself is *defined* by reference to an arbitrary lexical set. I propose that this makes the vowel realizations more vulnerable to lexically-mediated cross-dialect influence, resulting in AmE-like targets in both Zakaria's D1 and D2.

At this point, we have a situation where two categories (D1-vbath) and (D2-vbath) are becoming phonetically similar due to external pressures. This might trigger another mechanism— dissimilation. The Speech Learning Model (SLM) (Flege, 1987) posits that just like a monolingual system, a bilingual speaker's sound system faces pressure to maintain phonetic distinctions between *all* of their categories, pooled across languages. Therefore, when a speaker establishes a separate category for an L2 sound, this triggers an acoustic dissimilation between that sound and its phonetically-close L1 category, a process akin to phonetic category dispersion. Crucially, this is only operative when a new category has been established— if a speaker simply identifies a new L2 category with an existing L1 category, there is no pressure to dissimilate. Flege et al. (2003) report on the production of the category /e/, whose prototypical realization varies between English and Italian in the extent of formant movement (diphthongization; high in English, low in Italian). Proficient Italian-English bilinguals who had established separate categories for each language produced their English /e/ with *more* formant movement than monolingual English counterparts. The authors propose that this serves to enhance phonetic distance from the monophthongal Italian /e/ category. This pattern is almost exactly mirrored in Zakaria's vcot vowel: while prototypical AmE realizes this vowel as AA, Zakaria produces it with even more lowering than the prototype. This, I propose, serves to enhance phonetic distance from its D1 counterpart which, while canonically distinct from AmE vcot, has become lowered in Zakaria's system due to the (lexically-mediated) influence of exemplars. If this is on the right track and phonetic dissimilation is the goal, then apart from exaggerating the AmE-ness of this D2 system, the other logical option for Zakaria is to increase the IndE-ness of his D1 system by countering the effect of the exemplars. This appears to be the case for his vbath realization: while much more fronted than the canonical IndE realization, the vowel in Zakaria's D1 has resisted complete fronting to AE, and is instead produced at an intermediate position between AA and AE, while in his D2, it overlaps AE as expected. Thus the characteristics of Zakaria's D1 and D2 vowel systems after his long residence in the US can be seen as a push-pull between different and sometimes opposing forces. The stability over time indicates that at least for the timepoints studied here, these have culminated in an equilibrium that he is able to consistently maintain.

This section has provided an account for the characteristics of Zakaria's vowel system as it relates to points 1-3 listed above. Finally, question 4 remains: why these target realizations appear to remain stable over time, unlike the targets for VOT. Recall my proposal that the VOT changes over time reflect Zakaria's changing social goals. Thus, the shifts implicated are socially-driven. In the vast literature on the topic, many researchers have reported differences between phonetic features in their amenability to socially-driven shifts (de Leeuw and Celata, 2019). These have been attributed to various sources. Given that this study examines consonantal features and vowel features, one obvious hypothesis is that consonantal features (or VOT specifically) are overall more flexible than vowel quality. However, this is unlikely. Studies have shown that cross-linguistic interactions in both VOT and vowels can be subject to short-term adjustment in response to linguistic context (e.g. Olson (2013) for VOT; Simonet (2014) for vowels) as well as social situations/goals (e.g. Nielsen (2011) for VOT, Babel (2010) for vowels).

Another possibility, related to the preceding paragraphs, is that lexically-specific features (differences that apply to specific words) are less malleable than lexically non-specific features (such as VOT). Being less tied to specific lexical representations might allow a feature to be associated with a dialect as a whole, or derivable by a simple rule (e.g. 'lengthen VOTs for AmE') and thus easier to manipulate in response to changing social goals. By contrast, as discussed above, the vowel differences between the dialects cannot be derived by a parallel rule such as 'realize instances of AA as AE', which might make them less malleable for short-term changes. Examining more features, including consonantal differences that target specific lexical items, or lexically non-specific vowel features (such as the diphthongization of the /e/ and /o/ vowels in IndE and AmE) can test this hypothesis.

Alternatively, if the changes over time are truly socially-driven, then the explanation for the asymmetry could likewise lie in sociolinguistic facts. Specifically, it is possible that these features differ in social salience. Existing literature supports that idea that social salience of a feature affects the degree to which it is altered for socially-driven shifts. E.g. Babel (2012) reported differences in the extent to which speakers accommodated to different vowels of an interlocutor, attributing this to some vowels being more salient as dialect features than others. Examining the use of Southern British English vowel features over the course of 28 years in two speakers who moved post-adolescence from the North, Sankoff (2004) reports an asymmetry between two vowel features: broad A (using /aː/ rather than /æ/ in words like DANCE), and short U (using /ʌ/ rather than /ʊ/ in words like CUT). While both speakers produced short U in a D2-like manner, only one speaker used broad A. This was attributed to the higher salience of the latter as a Southern feature. Sankoff (2004) speculate that the lower salience of broad U made it 'available for adjustment at no social cost'. Regardless of the motivation, this demonstrates that social salience may lead to differences in the degree of shift even between two vowel features. A similar difference between the two features is reported in Evans and Iverson (2007), over a shorter timespan— speakers adopted the short U, but evidenced only small changes to their CUT vowel so that, although produced lower, it was not category-changing shift. In a parallel vein, in a study of phonetic alignment to the Shetland variety of English Smith and Durham (2012) found that speakers' treatment of certain phonetic features (e.g. th-stopping) did not match their language attitudes and social motivations. They propose that this is because these features are outside of metalinguistic awareness, and therefore less available for socially-mediated realization.

This is plausible for the variety under study here. Given the large variability in the L1s of its speakers, vowels in IndE are often very variable (and one of the main sources of regional differences). In contrast, short VOTs are characteristic of a majority of IndE speakers, and therefore, a deviation is more likely to be noticeable to an Indian audience. Therefore, it's possible that in the context of IndE specifically, changing the vowels in specific lexical items is less effective/essential for asserting identity and relatability (as Zakaria is presumably trying to do early on) than maintaining a short VOT. Anecdotally, in spite of the AmE-like vowel realizations, Zakaria's ie-context speech at timepoint 1 sounds globally very IndE-like, to the extent that the atypical vowel realizations are barely noticeable. Therefore, even if we assume that both features are equally available for manipulation, it is possible that salience makes one of them, VOT, more useful for agentive, socially-driven manipulation.

5 General discussion and conclusion

This study examined five phonetic features that differ between Generalized Indian English (IndE) and Standard American English (AmE): voice onset time in the stops /p/, /t/, and /k/, and vowel realization in the BATH class and the LOT class. I used acoustic data from public interviews of a single individual, the Indian-American political analyst Fareed Zakaria, produced over a period of 21 years (2003-2024) to study gradient changes in the realization of these features while interacting with different audiences (Indian vs American), at four points in time. Zakaria moved to the U.S. in at the age of seventeen and has been living there ever since. His speech from 2003 and later suggests that he was able to acquire AmE-like targets for both VOT and vowel features. This adds to a growing body of research suggesting that an individual's linguistic abilities remain plastic though adulthood (de Leeuw and Celata, 2019), questioning earlier views that posited a 'critical period' in adolescence, at which the first-learned system becomes fixed, and another variety cannot be acquired fully.

In Zakaria's D1, both VOT and vowel features show evidence of attrition/drift towards D2 norms. In VOT, this is manifested as longer VOTs than prototypical values reported in studies from speakers residing in India (Wiltshire, 2020; Sirsa and Redford, 2013). In vowels, this is manifested as fronted realizations of the BATH vowel and lowered realizations of the COT vowel. I proposed that this latter influence proceeded through changes to linked lexical representations. Thus Zakaria's D1, especially at the later timepoints, shows unmistakable influence of his long exposure to AmE.

In spite of this, Zakaria's speech evidences consistent acoustic differences between ie and ame contexts, showing that he maintains distinctions between his dialect varieties, and employs these in a socially meaningful way: drawing on his AmE-like D2 while interacting with primarily American audiences, and his IndE-like D1 while interacting with primarily Indian audiences. This kind of overall attunement to the perceived phonetic norms of the audience can be understood as an example of audience design (Bell, 1984). Audience design at a global level has been reported in an existing study of Zakaria's speech by Sharma (2018) based on the frequency of AmE-like and IndE-like variants of 12 variables while interacting with different audiences. The current results show that this pattern persists even at the level of acoustic patterns in individual variables.

Moreover, patterns of shift in Zakaria's VOT over the years, in conjunction with self-expressed changes in social goals and identity, provide evidence for more specific socially-motivated shifts in both D1 and D2 (Bell, 2006; Giles et al., 1991). At the earlier timepoints examined here, this led him to produce prototypical IndE and AmE values in his D1 and D2 respectively (although the latter could simply reflect his default D2 targets), whereas at later timepoints, his changing goals induced more D1-like norms (shorter VOTs) in his D2, resulting in the two systems becoming more similar. In concert with many previous studies (Nielsen (2011); Piccinini and Arvaniti (2015), a.o.), this showed that VOT can function as a site for socially-motivated shifts. In contrast, his vbath and vcot vowels remain unchanged over time, suggesting that the realizations reflect stable targets, and that these particular features may not be available for the same socially-motivated shifts that affect his VOTs over the years. I proposed that this might result either from their lexically-specific nature (compared to the VOT difference which applies across words), or lower salience, making them less available/useful for socially-motivated manipulation.

The data in this study is an extended version of the corpus examined by Sharma (2018), and I focused on a subset of the variables identified there. As mentioned above, the results cohere with Sharma's observation of robust audience-design in Zakaria's linguistic behavior. Moreover, these findings reinforce Sharma's proposal that agentive social motivations alone are not sufficient to understand all aspects of Zakaria's productions, even within a broad motivation of audience design. While Sharma examines the small-scale temporal dynamics of speech to highlight the role of cognitive processing in constraining audience design, I examine how different features behave within the broad umbrella of 'dialect variety', and how these change over longer periods. In doing so, the results highlight another source of constraints on audience-design: linguistic factors, resulting from the interaction between sound systems. I methodologically depart from Sharma (2018) in three ways: (i) while Sharma consciously assumes that all phonological features are equally amenable to shift, I explicitly probe this by comparing the features in their extent and patterns of shift; (ii) instead of categorically coding tokens as 'AmE' or 'IndE', I used acoustic measures (VOT and formant frequency) to study shifts in the selected variables, allowing me to quantify, for each token, the 'extent of AmE-ness or IndE-ness'; (iii) I did not examine the localized temporal dynamics within each recording. Instead, to understand how Zakaria's D1 and D2 systems might be changing over time, I compared productions across across four timepoints. While Sharma comments on some trends over time (particularly, that the total number of 'AmE' tokens appear to be decreasing), she does not explicitly include these in her analysis. Based on token counts across variables, Sharma (2018) reports 'a slight skewing, such that his style with AmE audiences includes more IndE admixture than vice versa, likely reflecting his later acquisition of an AmE'. I find that for the five features examined here, the converse pattern holds: the realizations overall are more skewed towards AmE norms than IndE norms. This could reflect a difference between different dialectal features, such that the other variables are more resistant to the effect of ambient productions, or an artifact of the coding strategies (categorical vs gradient) used in the two studies.

A distinction brought out by this study is between the within-system organization of the categories in a system, vs the phonetic realization of a given category across the two systems commanded by an individual. Categorical coding of variables in studies of socially-driven variation usually capture the former: each variable token is classified, for example, as being an AmE-realization or an IndE-realization, based on whether it is closer to AA or AO (for vcot words). This is the method employed by Sharma (2018) in analyzing Zakaria's speech. Such an approach would capture the fact that Zakaria's vcot productions are largely AmE-like (closer to AA). However, it fails to articulate how these categories are realized across the two contexts. Thus, it misses the generalization that Zakaria maintains dialect-specific realizations in this two varieties, that these differences are not random but rather reflect the expected direction of difference between the corresponding dialects, and that Zakaria's D2 productions of vcot are more lowered than the canonical AmE target. These patterns, I propose, reveal important information about how the sound systems interact within the individual. Of course, even with categorical coding, phonetically-trained researchers often perceive fine-grained differences that the coding system cannot capture. E.g. Sharma (2018) mentions the occurrence of intermediate forms, Sankoff (2004) explicitly includes a score of 0.5 in their coding system for tokens that 'sound intermediate'. Quantifying these patterns by measuring gradient acoustic features allows us to include them in our analysis, and draw on the sizable psycholinguistic literature on how the sound systems interact within an individual. Such facts can be seen as linguistic constraints on the agentive use of language. Ultimately, they also raise questions about how these production patterns relate to perception— are these persistent phonetic differences between an individual's varieties perceptible to listeners, or do they reflect pressures that are not audience-oriented at all?

To understand both the within-system and across-system phonetic patterns in Zakaria's speech, I drew on models of bilingual speech to explain, for example, how exemplars in the speaker's environment affect their production targets, how such interactions are mediated by the lexicon, and particularly, why there might be a pressure to phonetically distinguish corresponding categories across the two systems. This follows a growing interest in using existing research on bilingualism to understand aspects of bidialectal language use, and more broadly to probe the extent to which these are parallel or separate processes (Lønes et al., 2023).

This study has several limitations. In the analysis of VOT shifts, I did not account for differences in speaking rate, which some studies (although not all, see Stuart-Smith et al. (2015)) have identified as having an independent effect on VOT. Follow-up work should compare these results to a replication analysis with different speaking rate-normalized VOT measures. The analysis of vowels was limited by low token counts, which precluded statistical analyses for the timepoint differences. Although on visual examination vowel positions in the F1-F2 space do not appear to show any changes over time, this should be confirmed statistically with a larger amount of data. Finally, one limitation of this study design is that it focuses on a single individual's speech. While this approach allows for examining a large amount of speech over a long period of time, it is not clear to what extent the fine-grained patterns generalize. On the other hand, it has the advantage of allowing us to consider the individual's biographical information, which in this case appears crucial for understanding speech patterns. By examining a public figure's speech, we can contextualize it against information about their life and language history, providing insights into the various possible influences on their speech. This nuanced understanding may be lost in studies with multiple participants, where findings are often attributed solely to narrow hypotheses. Ideally, we would use a corpus with extensive metadata on its speakers to analyze comparably large amounts of data from different speakers. While it is challenging to match biographies precisely, it is essential to recognize that even in controlled experiments, such complex individual biographies still exist, and are likely to significantly influence speech patterns.

Abdel Rahman, R. and Melinger, A. (2008). Enhanced phonological facilitation and traces of concurrent word form activation in speech production: An object-naming study with multiple distractors. *Quarterly Journal of Experimental Psychology*, 61(9):1410–1440.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15(3):517–530.

Antoniou, M., Best, C. T., Tyler, M. D., and Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both l1 and l2. *Journal of phonetics*, 38(4):640–653.

Awan, S. N. and Stine, C. L. (2011). Voice onset time in indian english-accented speech. *Clinical linguistics phonetics*, 25(11-12):998–1003.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, pages 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1):177–189.

Baker, W. and Trofimovich, P. (2005). Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and speech*, 48(1):1–27.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bell, A. (1984). Language style as audience design. *Language in society*, 13(2):145–204.

Bell, A. (2006). Speech accommodation theory and audience design.

Best, C. T. and Tyler, M. (2007). Nonnative and second-language speech perception. In Bohn, O.-S. and Munro, M. J., editors, *Language experience in second language speech learning: In honour of James Emil Flege*, pages 13–34.

Blanco-Elorrieta, E. and Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, 213:104625.

Blanco-Elorrieta, E., Emmorey, K., and Pylkkänen, L. (2018). Language switching decomposed through meg and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences*, 115(39):9708–9713.

Blanco-Elorrieta, E. and Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, 37(37):9022–9036.

Bock, K. and Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of experimental psychology: General*, 129(2):177.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). Training japanese listeners to identify english/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, 61:977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training japanese listeners to identify english/r/and/l: Iv. some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4):2299–2310.

Bullock, B. E. and Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In Ludmila Isurin, Donald Winford, K. D. B., editor, *Multidisciplinary approaches to code switching*, volume 41, pages 189–206. John Benjamins Publishing Amsterdam, The Netherlands.

Campbell-Kibler, K., Walker, A., Elward, S., and Carmichael, K. (2014). Apparent time and network effects on long-term cross-dialect accommodation among college students. *University of Pennsylvania Working Papers in Linguistics*, 20(2):21–29.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, 54(2):421–428.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2):249–268.

Chen, Y. and Zhou, R. (2022). The mental lexicon features of the hakka-mandarin dialect bilingual. *Brain Sciences*, 12(12):1629.

Cho, T. and Ladefoged, P. (1999). Variation and universals in vot: evidence from 18 languages. *Journal of phonetics*, 27(2):207–229.

Chodroff, E. and Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant vot in american english. *Journal of Phonetics*, 61:30–47.

de Leeuw, E. (2014). Reassessing maturational constraints through evidence of l1 attrition in the domain of phonetics. *E. Thomas & I. Mennen, Unravelling bilingualism: A cross-disciplinary perspective. Bristol: Multilingual Matters.*

De Leeuw, E. (2019). Phonetic attrition. In *The Oxford handbook of language attrition.*

de Leeuw, E. and Celata, C. (2019). Plasticity of native phonetic and phonological domains in the context of bilingualism. *Journal of Phonetics*, 75:88–93.

De Leeuw, E., Schmid, M. S., and Mennen, I. (2010). The effects of contact on native language pronunciation in an l2 migrant setting. *Bilingualism: Language and Cognition*, 13(1):33–40.

Dijkstra, T. (2009). The multilingual lexicon. In *Cognition and pragmatics*, pages 369–388. John Benjamins Publishing Company.

Eckert, P. (2017). Age as a sociolinguistic variable. *The handbook of sociolinguistics*, pages 151–167.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal* *of the acoustical Society of America*, 62(4):971–974.

Evans, B. G. and Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6):3814–3826.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1):47–65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92:233–277.

Flege, J. E., Schirru, C., and MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication*, 40(4):467–491.

Foulkes, P. and Docherty, G. (2014). *Urban voices: Accent studies in the British Isles.* Routledge.

Garcia-Sierra, A., Diehl, R. L., and Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech communication*, 51(4):369–378.

Garrod, S. and Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3):181–215.

Giles, H. and Coupland, N. (1991). *Language: Contexts and consequences.* Thomson Brooks/Cole Publishing Co.

Giles, H., Coupland, N., and Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, 1:1–68.

Glaser, W. R. and Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of experimental Psychology: Human perception and performance*, 10(5):640.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*, 22(5):1166.

Gonzales, K. and Lotto, A. J. (2013). A bafri, un pafri: Bilinguals’ pseudoword identifications support language-specific phonetic systems. *Psychological science*, 24(11):2135–2142.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, 27(2):210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2):67–81.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2):131–149.

Grosjean, F. and Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, 5(4):201–206.

Hansen, J. H., Gray, S. S., and Kim, W. (2010). Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification. *Speech Communication*, 52(10):777–789.

Hazan, V. L. and Boulakia, G. (1993). Perception and production of a voicing contrast by french-english bilinguals. *Language and Speech*, 36(1):17–38.

Hussain, Q. (2018). A typological study of voice onset time (vot) in indo-iranian languages. *Journal of Phonetics*, 71:284–305.

Johnson, K. A. (2021). Leveraging the uniformity framework to examine crosslinguistic similarity for long-lag stops in spontaneous cantonese-english bilingual speech. In *Interspeech*, pages 2671–2675.

Keshet, J., Sonderegger, M., and Knowles, T. (2014). Autovot: A tool for automatic measurement of voice onset time using discriminative structured prediction [computer program]. *Version 0.91, retrieved August.*

Kirk, N. W., Declerck, M., Kemp, R. J., and Kempe, V. (2022). Language control in regional dialect speakers–monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, 25(3):511–520.

Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A., and Declerck, M. (2018). Can monolinguals be like bilinguals? evidence from dialect switching. *Cognition*, 170:164–178.

Kroll, J. F. and Gollan, T. H. (2013). Speech planning in two languages. In *The Oxford handbook of language production*.

Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.

Labov, W., Ash, S., and Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter.

Lenth, R. V. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.7.

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.

Lønes, E. H., Kamide, Y., and Melinger, A. (2023). Speaking in dialects: How dialect words are represented and selected for production. *Psychology of Learning and Motivation*, 78:119–159.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7(6):485–495.

Ma, M., Glass, L., and Stanford, J. (2024). Introducing bed word: A new automated speech recognition tool for sociolinguistic interview transcription. *Linguistics Vanguard*, 10(1):641–653.

Magloire, J. and Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in english and spanish. *Phonetica*, 56(3-4):158–185.

Major, R. C. (1992). Losing english as a first language. *The Modern Language Journal*, 76(2):190–208.

Marian, V. and Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24(2):173–193.

Marian, V. and Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and cognition*, 6(2):97–115.

Mayr, R., Morris, J., Mennen, I., and Williams, D. (2017). Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in welsh and english. *International journal of bilingualism*, 21(3):245–267.

McCullough, E. A. (2013). Perceived foreign accent in three varieties of non-native english.

McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive science*, 30(6):1113–1126.

Melinger, A. (2018). Distinguishing languages from dialects: a litmus test using the picture-word interference task. *Cognition*, 172:73–88.

Munro, M. J., Derwing, T. M., and Flege, J. E. (1999). Canadians in alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, 27(4):385–403.

Narkar, J. (2021). The curious absence of aspiration in indian english: The role of phonetics in adaptation. In *37th West Coast Conference on Formal Linguistics*, pages 189–196. Cascadilla Proceedings Project.

Narkar, J. and Staroverov, P. (2022). An acoustic study of voiceless stops in indian english. *Journal of South Asian Linguistics*, 12:53–68.

Nielsen, K. (2011). Specificity and abstractness of vot imitation. *Journal of Phonetics*, 39(2):132–142.

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2):204–238.

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, 9(11):469–482.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41(6):407–420.

Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International journal of bilingualism*, 5(1):19–38.

Pardo, J. S. (2010). Expressing oneself in conversational interaction. In Morsella, E., editor, *Expressing* *oneself/expressing one's self: Communication, cognition, language, and identity*, pages 183–196. London: Taylor & Francis.

Piccinini, P. and Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, 52:121–137.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

Pingali, S. (2009). *Indian English.* Edinburgh University Press.

Pingali, S. (2022). Indian english: Features and development. In *English in East and South Asia*, volume 1, pages 153–167. Routledge, 1 edition.

R Core Team (2024). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reddy, S. and Stanford, J. N. (2015). Toward completely automated vowel extraction: Introducing darla. *Linguistics Vanguard*, 1(1):15–28.

Rickford, J. and Price, M. (2013). Girlz ii women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics*, 17(2):143–179.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. *Speech perception and linguistic experience: Issues in cross-language research*, 379:410.

Rvachew, S. and Jamieson, D. (1989). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, 16:201–208.

Ryalls, J., Simon, M., and Thomason, J. (2004). Voice onset time production in older caucasian-and african-americans. *Journal of Multilingual Communication Disorders*, 2(1):61–67.

Sancier, M. L., Fowler, C. A., et al. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4):421–436.

Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from 'seven up'. *Sociolinguistic variation: Critical reflections*, pages 121–139.

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

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, 43:26–37.

Simonet, M. (2016). The Phonetics and Phonology of Bilingualism. In *Oxford Handbook Topics in Linguistics.* Oxford University Press.

Sirsa, H. and Redford, M. A. (2013). The effects of native language on Indian English sounds and timing patterns. *Journal of Phonetics*, 41(6):393–406.

Smith, J. and Durham, M. (2012). Bidialectalism or dialect death? explaining generational change in the shetland islands, scotland. *American Speech*, 87(1):57–88.

Sonderegger, M. (2012). *Phonetic and phonological dynamics on reality television.* University of Chicago.

Spivey, M. J. and Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science*, 10(3):281–284.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., and Macdonald, R. (2015). The private life of stops: Vot in a real-time corpus of spontaneous glaswegian. *Laboratory Phonology*, 6(3-4):505–549.

Tobin, S. J., Nam, H., and Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics*, 65:45–59.

Trudgill, P. (1999). Dialect contact, dialectology and socionlinguistics. *Cuadernos de Filología Inglesa*, 8.

Tsui, R. K.-Y., Tong, X., and Chan, C. S. K. (2019). Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. *Applied Psycholinguistics*, 40(1):29–58.

Wells, J. C. (1982). *Accents of English: Volume 1*, volume 1. Cambridge University Press.

Wiltshire, C. R. (2020). *Uniformity and variability in the Indian English accent*. Cambridge University Press.

Wiltshire, C. R. and Harnsberger, J. D. (2006). The influence of gujarati and tamil l1s on indian english: A preliminary study. *World Englishes*, 25(1):91–104.

Woutersen, M., Cox, A., Weltens, B., and De Bot, K. (1994). Lexical aspects of standard dialect bilingualism. *Applied psycholinguistics*, 15(4):447–473.

Xie, X., Theodore, R. M., and Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1):206.

Yao, Y. (2007). Closure duration and vot of word-initial voiceless plosives in english in spontaneous connected speech. *UC Berkeley PhonLab Annual Report*, 3(3).

Yao, Y. (2009). Understanding vot variation in spontaneous speech. *UC Berkeley PhonLab Annual Report*, 5(5).

Appendix

1. List of words coded with variable v_bath vowel (the vowel in the primary-stress syllable): ABU, AFGHAN, AFGHANISTAN, AFTER, ANSWER, ASK, ASKED, ASKING, ASKS, BAGHDAD, BANGLADESH, BANGLADESHIS, BASKET, BASQUES, CAN'T, CASTE, CHANCE, CLASS, COMMAND, FAST, FASTER, FRANCE, GLASS, HALF, IRAQI, IRAQIS, LAST, MASKED, MASKS, MASS, MASTER, PAKISTAN, PAKISTANI, PASS, PASSED, PASSING, PAST, PATH, PATHWAY, PLANTS, RATHER, TALIBAN, TASK, TRANS, VAST

2. List of words coded with variable v_cot vowel (the vowel in the primary-stress syllable): ABOLISH, ADOPT, ADOPTED, ADOPTION, ANTIBIOTICS, ASTRONOMICAL, AUDIENCE, AUDIO, AUTONOMY, AWE, BEYOND, BLOCK, BLOCKING, BOMBING, BORROW, BOSNIA, BOTTOM, BOX, BOXES, BUREAUCRACIES, BUREAUCRACY, CAUSED, CAUSES, CAUSING, CAUTION, COLLEGE, COMBAT, COMMENT, COMMENTARY, COMMENTS, COMMON, COMMUNIST, COMPLEX, COMPLICATED, COMPROMISE, CONCENTRATE, CONCENTRATED, CONFERENCE, CONFIDENCE, CONFLICT, CONGRESS, CONQUER, CONSEQUENCE, CONTENT, CONTINENT, CONTRAST, COOPERATIVE, COPIED, COPY, COPYING, CORRESPONDENCE, COST, COSTS, COVID, DEMOCRACIES, DEMOCRACY, DEMOCRATIZING, DEPOSIT, DOLLAR, DOLLARS, DOMINANT, DOMINATE, DONALD, DROP, ECONOMIC, ECONOMICALLY, ECONOMICS, ECONOMIES, ECONOMY, EMBODIES, FOLLOW, FOLLOWED, FOLLOWING, FORGOTTEN, FOSSIL, GENEROSITY, GOD'S, GODDESS, GOSH, HONEST, HONESTLY, HYPOCRISY, IDEOLOGICALLY, IDEOLOGY, IMPOSSIBLE, INCOMPETENT, INVOLVE, INVOLVED, INVOLVES, JOB, JOBS, JOHN, JOHNSON, LOBBYING, LOCKDOWNS, LOCKED, LOGIC, LOGICAL, LOT, LOTS, MODEL, MODELS, MODERATE, MODERN, MODERNIZE, MODERNIZED, MODERNIZING, MODEST, MODESTY, MOLLIFYING, MONARCHY, MONITORS, MOSCOW, NON, NONSENSE, NONWORKING, OBJECT, OBVIOUS, OBVIOUSLY, OCCUPIED, ODD, ONSET, ONTO, OPTIMISM, OPTIMIST, OXFORD, OXLEY, PHENOMENON, POCKET, POLICIES, POLICY, POLITICS, POPULAR, POSITIVE, POSSIBLE, POSSIBLY, POSTURE, POVERTY, PROBABLY, PROBLEM, PROBLEMS, PROCESS, PROCESSING, PROGRESS, PROJECT, PROJECTS, PROMISE, PROMISES, PROMISING, PROPER, PROPERTY, PROPPING, PROSECUTE, PROSECUTING, PROSPEROUS, PSYCHOLOGICAL, RESOLVE, RESPONSE, RESPONSIBLE, ROBERT, ROCK, ROD, SCHOLARS, SHOCK, SHOT, SOLVE, SOLVED, STOCK, STOP, STOPPED, SUBCONTINENT, SYMBOLIC, SYMBOLICALLY, TECHNOLOGICAL, TECHNOLOGICALLY, TECHNOLOGIES, TECHNOLOGY, TECTONIC, THEOCRACY, THOMAS, TOP, TOPPLE, VOLATILE

3. Model outputs for VOT, subsetted by stop (table 11, 12, 13)

Fixed effects	Estimate (β)
Timepoint-1	0.066 (0.045)
Timepoint-2	0.049 (0.038)
Timepoint-3	-0.132^{***} (0.044)
Context-ame	0.192^{***} (0.024)
Timepoint1:ame	0.114^{**} (0.045)
Timepoint2:ame	0.133^{***} (0.039)
Timepoint3:ame	-0.148^{***} (0.044)
Intercept	2.913^{***} (0.031)
Observations	673
Akaike Inf. Crit.	1,133.121
Bayesian Inf. Crit.	1,178.238
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Optimal model: VOT for /p/

Fixed effects	Estimate (β)
Timepoint-1	0.105^{**} (0.043)
Timepoint-2	-0.013 (0.034)
Timepoint-3	-0.097^{**} (0.041)
Context-ame	0.227^{***} (0.022)
Timepoint1:ame	0.082^{*} (0.042)
Timepoint2:ame	0.019 (0.034)
Timepoint3:ame	0.003 (0.041)
Intercept	3.272^{***} (0.029)
Observations	706
Akaike Inf. Crit.	1,063.059
Bayesian Inf. Crit.	1,108.655
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Optimal model: VOT for /t/

Fixed effects	Estimate (β)
Timepoint-1	0.128^{***} (0.039)
Timepoint-2	0.032 (0.032)
Timepoint-3	-0.202^{***} (0.039)
Context-ame	0.257^{***} (0.021)
Timepoint1:ame	0.166^{***} (0.039)
Timepoint2:ame	-0.016 (0.032)
Timepoint3:ame	0.041 (0.039)
Intercept	3.475^{***} (0.035)
Observations	796
Akaike Inf. Crit.	1,236.501
Bayesian Inf. Crit.	1,283.297
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 13: Optimal model: VOT for /k/

4. English VOTs from speakers of L1 AmE (top row) and Hindi (second row) from McCullough (2013) (figure 17)

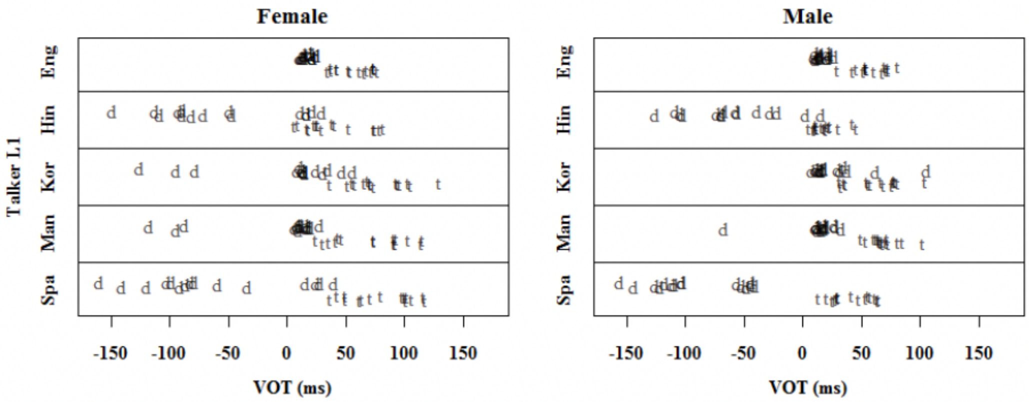


Figure 17: English VOTs from speakers of L1 AmE and Hindi from McCullough (2013). The labels ‘t’ and ‘d’ represent all the voiceless and voiced categories respectively.

5. Distance between reference categories (figures 18, 19)

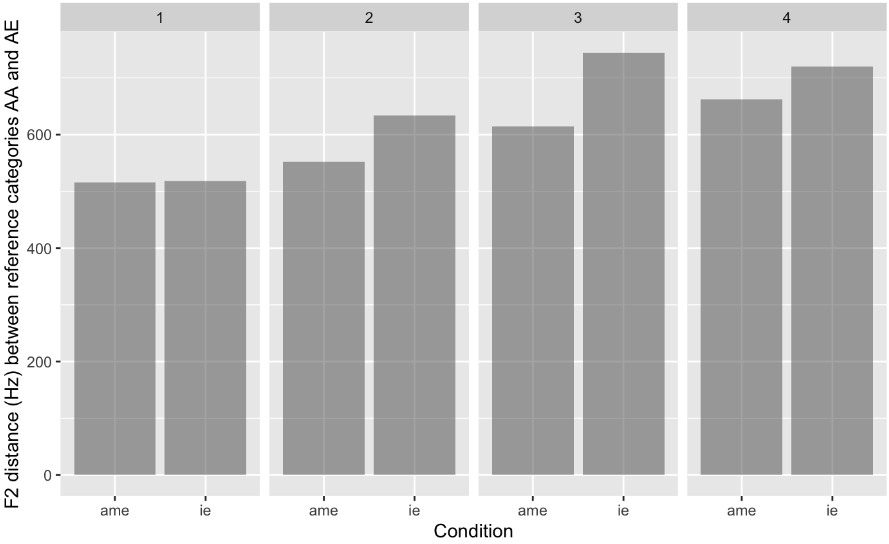


Figure 18: Condition-wise F2 distance between reference categories AA and AE

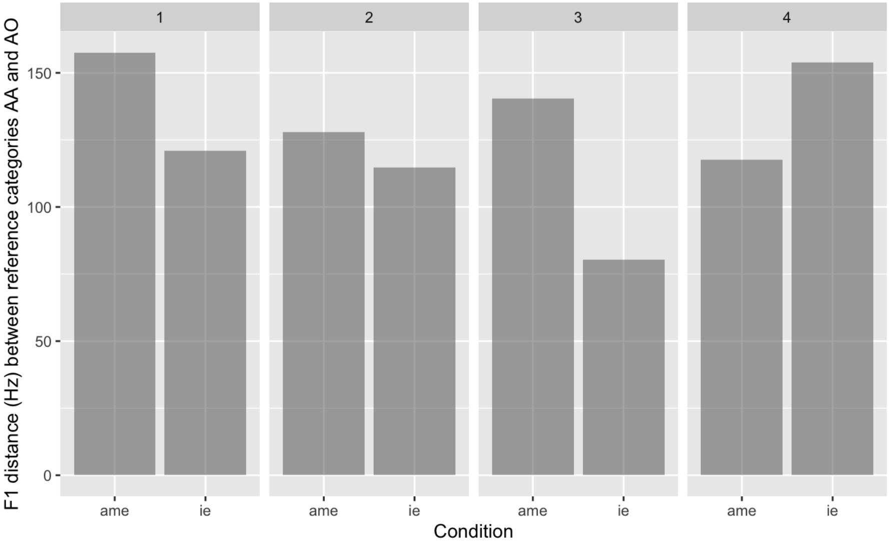


Figure 19: Condition-wise F1 distance between reference categories AA and AO