

THE DIPHTHONGIZATION OBSERVATION: AN ANALYSIS OF THE “INDIE VOICE” PERSONA

SARAH SMITH
University of Rochester

1 Introduction

This paper aims to do the following: 1) describe the phenomenon of “indie voice” through a linguistic lens, touching on its public perception; 2) discuss previous sociolinguistic studies focused on style within the frame of singing and how those have inspired the current paper; 3) analyze data on indie-pop singer Halsey’s diphthongization of certain vowels when singing, and compare these diphthongs to the “pure” vowels she uses in normal speech; and 4) argue that this diphthongization—and by extrapolation, many other indie voice sound changes—is indexical of a specific “indie persona” by calling on and adapting performance- and persona-based theories of style-shifting.

2 Background

2.1 Sociolinguistics and Song

Spoken language and sung music are not isolated forms of language use. There are many overlapping factors that connect the variation observed in speech to that of singing, the most relevant for the purposes of this paper being register. Schilling-Estes defines a register to be a “readily identifiable speech variety that individuals use in specific, well-defined speech situations” (Schilling-Estes, 1998). Registers are subsumed within the realm of style, as they can also be viewed as a set of stylized variables bunched together to form an identifiable variety of speech. Registers also exist in music, and are typically genre-specific, as illustrated by Karen from Ace Linguist’s explanation that “you use very different vowels for classical operatic singing than you do when singing a bluesy rock song” (Ace Linguist, 2021). Singers commonly utilize a different register when singing from their typical spoken vernacular, sometimes picking

up forms in their sung accent that are not present at all in their everyday speech. This is not always a conscious decision, rather is sometimes physiologically necessary due to the mode of vocalization that a particular song requires. Articulations may be nudged around in the vocal tract away from their standards in order to produce a sound with more resonance, or one that matches the aural qualities of the instrumental better (Ugwu, 2015). This is not to say that sound changes in singing are strictly coming from physiological demand. Just as style-shifting in speech mostly happens due to social motivations, this can also be the case for song.

Spoken and sung language are not equivalent, however. Sung speech has an affordance to creativity that spoken speech cannot accommodate in the same way. Karen from *Ace Linguist* claims that “sung speech gives us an opportunity to witness experimentation and playfulness in the low-stakes world of music and art” (*Ace Linguist*, 2021). Music both sets up constraints, such as the need to follow a melody and rhythm, and supports experimentation due to its malleability, which are facets that come together to allow artists to come up with novel productions unheard in speech that work to not just suit the song but enhance it. These novel productions can systematize to create new linguistic repertoires, which over time and across singers may come to denote a particular genre or identity.

The connections between spoken and sung language, along with music’s ability to create novel linguistic identities, means that there have been multiple studies focused on sung speech. There have, of course, been considerations of popular song language from a musicological perspective, such as Blake’s description of timbre and Burkhart’s discussion of paralinguistic musical features, both explaining how these musical aspects index aesthetic values (Blake, 2012; Burkhart, 2017). But Trudgill’s seminal paper on the localized accents of British punk-rock singers, which examined their use of a low-prestige southern English accent in comparison to British pop singers’ “Americanized” features (Trudgill, 1997), served as the catalyst for many further sociolinguists curious about sung dialect. Simpson expanded upon Trudgill’s findings, defining more clearly the “American” features that British pop singers imported and how those differed from the singers’ spoken styles, as well as explaining that the stereotypically “American” features “do not always tally with any specific variety of American English” (Simpson, 1999). Beal connected this work to the concepts of performance and persona (to be discussed in Section 4) through doing a similar examination of the Arctic Monkeys’ use of local features (Beal, 2009). Jansen and Westphal took a different approach and looked at pop singer Rihanna’s use of Caribbean English Creole features in her music and how that highlighted her identity (Jansen & Westphal, 2017). Lastly, Jones et al. hypothesized on the articulatory placement of indie voice, which most closely relates to this study (Jones et al., 2017).

2.2 A Game of Telephone

As described in Trudgill (1997), Simpson (1999), and Beal (2009), Americans have dominated the popular music industry for decades; and in an attempt to replicate and associate themselves with the success of the American market, many artists in other English-speaking countries copied the most salient linguistic features of American accents and used those features in their own work. As the decades went on, with more and more mainstream singers all utilizing the same accent, those features began to index being “mainstream” rather than being “American”. This is when, in the 70’s, British punk-rock bands began to embrace their local Southern English accent and use that in their music instead. Similarly, Australian hip-hop artists tended to stick to their local voice rather than accommodating to the American pronunciation model. In both cases, the

rationale seems to be that using one's own vernacular, rather than changing one's voice to fit the same accent everyone in the corporate mainstream is using, asserts a genuine, "keeping it real" attitude. Jansen and Westphal describe this "going local" as "an active choice against an American-influenced mainstream and towards linguistic and creative independence and authenticity" (Jansen & Westphal, 2017). As these "down to earth" musicians became popular internationally, other vernaculars were introduced into the popular music sphere, and in particular many low-prestige Australian and British features were picked up and imitated by other counterculture artists. Then they too were copied, features found across multiple accents were converged, and these non-American linguistic features were spread across and became marked aspects of various alternative genres. This game of counterculture musical accent telephone may be (part of) how the phenomenon of "indie voice" came about (Ace Linguist, 2021).

2.3 Singing in Cursive

Indie music as a genre is characterized by an anti-mainstream ideology and an association with smaller record labels rather than huge management systems. It is reasonable to consider the indie movement a descendent of the punk values that sparked the telephone game described in the previous section, and therefore is also fair to assume that the accent used in indie music would also follow from the linguistic features derived from those values. "Indie Voice," also referred to as "singing in cursive" or "hip singing," is a particular register of sung speech used by many musicians in the indie genre. This style attracted large-scale attention in 2015, when a BuzzFeed article was written about singer Selena Gomez's use of it in her song "Good for You" (Ugwu, 2015). Not only was this style noticed, but it was also hugely criticized and mocked for sounding excessively ornate and being difficult to understand (Ace Linguist, 2021). The overall goal of indie voice is to share a feeling of intimacy, individuality, and vulnerability. Various linguistic features are characteristic of indie voice, such as a quiet breathy tone, lack of aspiration, extensive vocal fry, and shakiness. There are also distinct vowel choices which are particularly salient to listeners. Certain commonly remarked upon simple vowel changes are as follows:

/aɪ/ → [ɑɪ] /ʌ/ → [a] /eɪ/ → [æɪ] /æʊ/ → [aʊ]

But what's most unique to the register, as well as most ridiculed by listeners, is its distinctive substitution of front-rising diphthongs in the place of most any standard monophthong, examples being:

/ɛ/ → [ɛɪ] /ʌ/ → [ʌɪ] /ɔ/ → [ɔɪ]

In her general guide to the observed sounds of indie voice, Karen from the blog Ace Linguist reports that, "Diphthongization can be very obvious or extremely subtle, but either way the vowels no longer sound pure. These diphthongs are closing diphthongs – they go from a low vowel to a high vowel. The one exception is /ʊ/ → [ʊɪ], where the tongue stays at the same height as it moves forward" (Ace Linguist, 2021). Nearly all diphthongized vowels move specifically towards an /i/ sound, and most occur before a coronal consonant, which lead Jones et al. to believe that "the diphthongs are prolonged audible transitions between the tongue's vocalic position and the articulatory target for the following consonant" (Jones et al., 2017).

This characteristic diphthongization will be the focus for the data analysis portion of this paper. This is for two main reasons. First, it is perceptually salient and easy to pick out. Second, it is a feature that is not already indexical of some other genre, identity, or location. Many of the

individual vowel changes come from British and Australian dialects, and the overall soft voice quality features are taken from R&B and folk music, but diphthongization is purely an indie voice characteristic.

3 Data Analysis

3.1 The Variable and The Question

What we're concerned with is the addition of a short, secondary vowel to a standardly monophthong vowel, henceforth called diphthongization. The variable is the diphthongized quality of a vowel, and the variants are whether that quality is present or not. The envelope of variation encompasses all vowels except for the high-front /i/ and /ɪ/, given those can't slide to "close off" any higher or more forward. Measurement of this variable was done both impressionistically (whether or not a diphthong is present) and quantitatively (using formant values). A visual comparison of a diphthongized /ʌ/ and a monophthong /ʌ/ is in Figure 1.

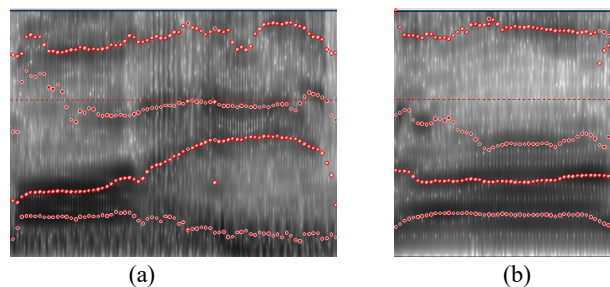


Fig 1. Spectrograms for (a) the diphthongized [ʌ] in “much” and (b) monophthong [ʌ] in “clubs”. The formant shift to the [ɪ] is clearly present in (a) and not in (b).

The question: Will American indie(-pop) singer Halsey, who is commonly quoted as being a user of indie voice, produce this extensive diphthongization in her singing? Will it be in the manner and frequency described (closing diphthongs, all the time)? Will this differ from how she pronounces vowels in everyday speech? I hypothesize that Halsey will indeed often use extensive diphthongization by adding a short /ɪ/ to multiple vowels that are pronounced as monophthongs in her spoken language.

3.2 Method

Two clips were obtained from Australian radio station Triple J's “Like A Version,” which each include audio of Halsey both speaking and singing. These clips were both taken on the same day, with the same audience (two interviewers and two musicians) present. The total length of both clips together comes to 10 minutes 21 seconds, with around 2 minutes of that time devoted to talking and 8 minutes spent singing. The speaking and singing portions were extracted from each other, put into Praat, and hand-transcribed and annotated on the word level. Any audio sections where voices overlapped or there were purely instrumentals with no singing were deleted. The annotated files were put through DARLA's Semi-Automated Alignment and Extraction suite to conduct mean formant analyses to allow visual description of the changing vowels (Fig. 2). DARLA does not automatically provide the second formant values for diphthongs, so every

diphthong's two sets of formants were calculated using Praat one at a time. Files were then auditorily examined to acquire token counts of each type of vowel, with a focus on catching any uses of *i*-diphthongization.

3.3 Results and Analysis

Token counts for each standard type of vowel, as well as how it was actually pronounced in each speech setting, are shown in the chart below. In this and all future analysis, “standardly monophthongs” refers to the vowels {*ɑ*, *æ*, *ʌ*, *ɔ*, *ə*, *ɜ*, *ɛ*, *ɪ*, *i*, *ʊ*, *u*} and “standardly diphthongs” refers to {*aʊ*, *aɪ*, *eɪ*, *oʊ*}.

	Speaking		Singing	
	Pronounced as Monophthong	Pronounced as Diphthong	Pronounced as Monophthong	Pronounced as Diphthong
Standardly Monophthongs	337	0	654	38
Standardly Diphthongs	24	69	102	179

The boxes of interest in the charts are those of a standard monophthong pronounced as a diphthong (diphthongization), for which Halsey had no tokens when speaking, but 38 when singing. 38 is not a particularly large number, but it is far more than zero. To confirm, a two-sample t-test assuming equal variances was run on the standardly monophthong data to determine whether the diphthongization difference between speech settings was significant. Means were acquired by coding all instances of diphthongization as 1 and standard monophthong pronunciations as 0. The resulting p-value was $p = 0.0000108$, which is much less than $\alpha = 0.05$, so we can conclude that the difference is significant, and it is her indie voice singing environment that is inspiring diphthongization.

Not every vowel underwent diphthongization. Only /*ɑ*/, /*æ*/, /*ʌ*/, /*ɔ*/, /*ə*/, and /*ɛ*/ had diphthongized tokens, meaning that /*ɜ*/, /*ʊ*/, and /*u*/ were excluded. This makes sense, as /*ɜ*/ is sometimes considered to be a diphthong in the first place, so it shouldn't undergo extra diphthongization. Both *ʊ* and *u* are already high vowels. They do not have the ability to be raised further, only fronted. As mentioned in Section 2.3, diphthongization should be possible on /*ʊ*/, so either it was just unfortunate that Halsey's audio didn't have any tokens of that, or diphthongization is only available for vowels that can be both fronted and raised, but exploring that theory further is outside the scope of this paper.

For a visual representation of the sound changes happening across different vowel conditions, mean formant values for all spoken vowels, all sung vowels, and all diphthongized vowels were put into the NORM Vowel Normalization and Plotting Suite to create the vowel space maps in Figure 2. NORM requires Arpabet vowel notation rather than IPA, so that is what is shown here. Conversions between the two systems are available online.

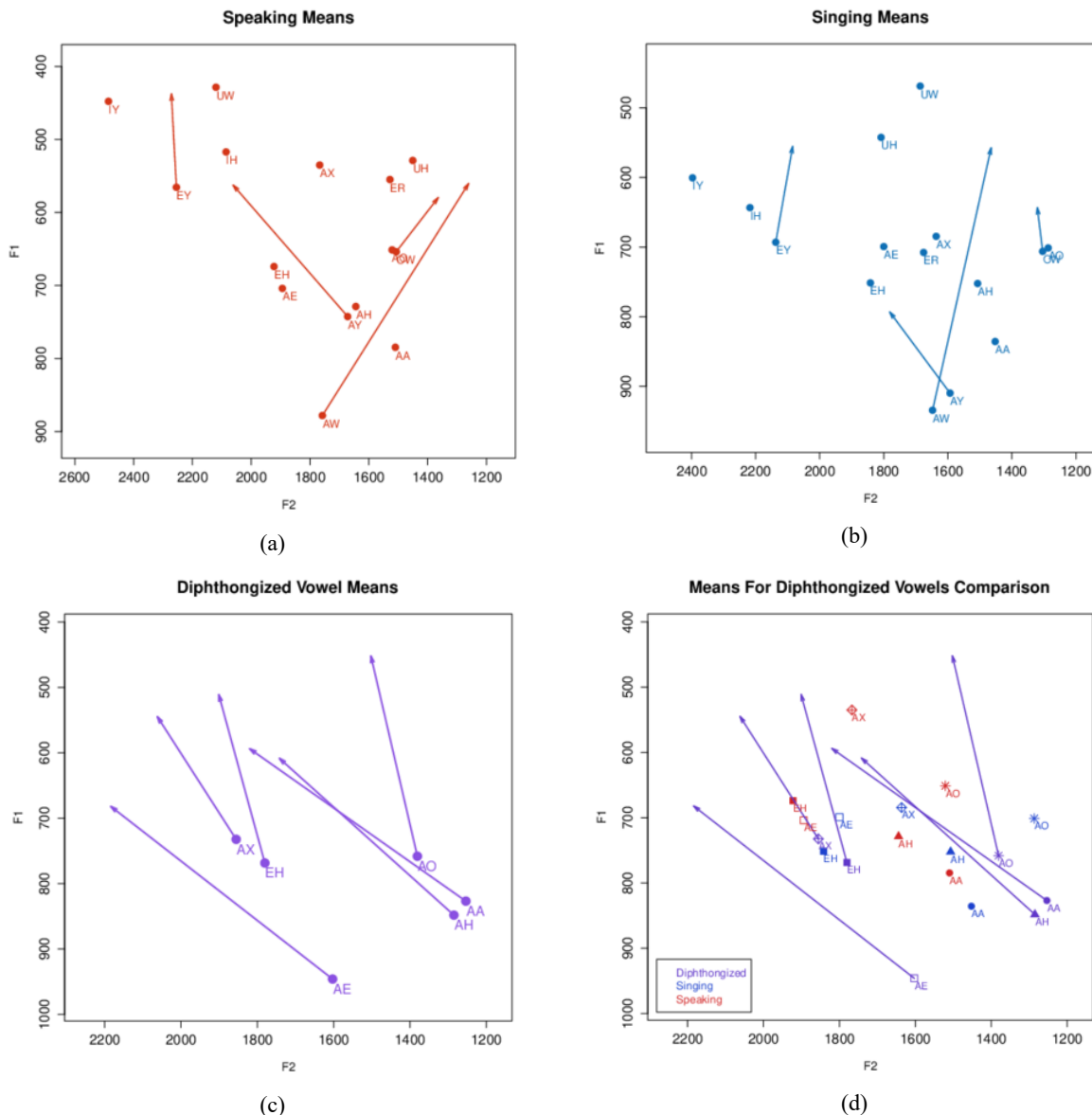


Fig 2. Vowel spaces for (a) all vowel means from Halsey's speaking audio, (b) all vowel means (excluding diphthongized ones) for her singing audio, (c) means for the diphthongized vowels from her singing audio, and (d) comparison between the placement of the diphthongized vowels and their standard monophthong counterparts from speech and song.

It is clear from these plots that vowels are generally being somewhat lowered and backed in song from their positions in speech. What's more interesting is how the diphthongized vowels are nearly all starting lower than both their spoken and sung counterparts. Their moving higher and to a more fronted position is a given considering the definition of i-diphthongization, but this initial lowering is an effect I did not foresee, and it would be interesting to explore the reasons behind it in further research. My tentative explanation is that this happens in order to make the glide from the initial vowel to /i/ more distinct and audible. Now that it has been established that diphthongization and other indie voice sound changes are indeed happening in the manner described in Section 2.3, it's important to discuss why they might be happening.

4 Connections to Theory

4.1 Previous Theories

Various scholars have proposed theories as to why style shifting happens in music. Beal considered the possibility of Giles and Smith's accommodation theory, which states that "speakers modify their accents to either converge to or diverge from those of their interlocutors" (Beal, 2009). The issue that arises here is that in the performance of a song, a singer has no interlocutors. It is not a conversation that is taking place. Bell's audience design theory attempts to remedy this problem by placing the audience in the position of interlocutors, so singers would be accommodating to the accent of their audience. This theory is also problematic, since artists tend to stick to a particular style across all of their music but perform to both many different populations (tours, concerts) and sometimes no audience (recording sessions). Thus, they cannot be accommodating towards or against both every possible audience and no audience in songs that are sung in the same style across performances.

Another possible theory is that of "appropriateness". Trudgill explains this concept by saying that "Different situations, different topics, different genres require different linguistic styles and registers. The singing of pop music in this way, it could be argued, is no different from vicars preaching in the register appropriate to Church of England sermons, or BBC newsreaders employing the variety appropriate for the reading of the news" (Trudgill, 1997). He then deprecates this theory, however, by claiming that it is not enough to simply say a certain register is appropriate for a genre, but it must be explained why the specific features used are considered appropriate.

There is also consideration of the articulatory settings used by singers of different genres. Jones et al. illustrates that "The articulatory settings associated with a given singing style are known to have noticeable, systemic effects on the acoustics of the singer's pronunciation" (Jones et al., 2017). This is the prospect that they investigated in their 2017 study, defending the hypothesis that indie voice is the result of a pharyngealized articulatory setting. Even they admit that the phenomenon is not just physiologically driven, and that "while indie-pop's distinctive diphthongs may originally have been a by-product of an articulatory setting, they have since been adopted as part of a musical style" (Jones et al., 2017), that is, the sound changes became an overt aspect of the way users pick up and identify the style, rather than just being a symptom of a particular articulation.

None of these theories have provided us with a complete picture of the reasoning behind the shift to indie voice. I propose that a mixture of Jones et al.'s articulation theory, various scholars' concept of persona-driven style, and Schilling-Estes' and Coupland's concept of performed language all work together to, respectively, target what Simpson says are the three interrelated factors that influence singing style, "the nature of the perceptual linguistic model aspired to, the nature of the pop and rock bands who adopt it and the nature of the bands' targeted audience" (Simpson, 1999).

4.2 Performance

In her paper on Ocracoke English, Schilling-Estes introduces the register of Performance Speech, which is "associated with speakers' attempting to display for others a certain language or

language variety, whether their own or that of another speech community” (Schilling-Estes, 1998). Performance speech targets specific features that are perceptible from the target dialect. She explains this by saying that “When speakers attempt to “put on” a dialect for an audience, they have available to them only those features they can perceive; further, there is evidence that the greater perceptual awareness speakers have of a given language feature (whether this awareness is at the conscious level or not), the greater the extent to which the feature will figure in their demonstrations and discussions of the language variety in question” (Schilling-Estes, 1998). This tendency to lean into particularly noticeable features lines up with the passing down/imitation chain of features described in Section 2.2, justifying why perceptible changes (like diphthongization) become characteristic among users of the indie voice register.

Singing comes under the umbrella of what Coupland calls “high performance”. It is planned (not impromptu like standard speech), intentionally stylized, and aimed for entertainment rather than conversation. Singers are performing the speech of a character (that may or may not be themselves), that affords a “crucial relational dynamic between a performer and the individuated audience member, who may align with or even feel transcendently drawn into the performer’s own identity” (Coupland, 2011). This relational dynamic inspires “‘higher order indexicalities’ – awareness that a certain stylistic variant operates as an index for a certain social meaning” (Bell & Gibson, 2011). Through these performer-audience relations, the stylized form of the performance becomes enregistered (“the process by which a style becomes engraved in the public mind as indexing certain social personas or ‘characterological figures’” (Bell & Gibson, 2011) as indexical of an indie singer. Indie artists are not converging with anyone, since there is no dialect in the spoken world that uses all of the features of indie voice. Rather, during performances, they are differentiating themselves from everyone, to try to push their special hippie image.

4.3 Persona and Indexicality

Le Page’s “Acts of Identity” theory, where linguistic behavior is motivated by attempts to “resemble as closely as possible those of the group or groups with which from time to time we [speakers] wish to identify” (Beal, 2009), seems more encompassing than those discussed in Section 4.1. In terms of this theory, indie singers are modifying their pronunciation when singing towards that of previous users of indie voice, a group with which they wish to identify. Beal claims that “Within the language-ideological framework, linguistic features are seen to become associated with social values so that they acquire indexical meanings. These meanings can change through time, and different groups in a community may attribute different values to the same linguistic feature” (Beal, 2009). What style shifting to indie voice appears to be doing is pushing forward a speaker’s “projected social role and persona” (Simpson, 1999) that, in this genre, aims to project a raw, uncensored image with a “principle of transparency in performance that suggests real and deep experience” and a social “premium on individual character over established convention” (Coupland, 2011; Ugwu, 2015).

How do indie singers achieve this image? In Zhang’s account of persona-driven language, he shows that “To create a distinctive style, individuals draw upon resources that are accessible to them. However, they do not pick up stylistic material in a free-wheeling fashion. The creation of style can be seen as a process of “bricolage” (Levi- Strauss 1966) in which the “bricoleur” selects from a limited and preexisting set of materials at hand and arranges them into a meaningful ensemble” (Zhang, 2005). Following this framework, indie voice is a bricolage of many sound

changes from no one clear source. The use of some of these features (vocal fry, breathy voice) is already culturally indexical of a personal, authentic tone. The indie singers taking those known aspects, as well as copying the noticeable vowel features of previous indie singers and other counterculture musicians and putting everything together allows them to portray a down-to-earth (as opposed to popular, celebrity), “individualistic” persona that associates them with the indie community. Although it may seem strange for such a hodgepodge bundle of features to imply the indie persona in particular, Bell and Gibson clarify that “Indexes do not directly resemble the referent but reference it through association and co-occurrence—smoke is an index of fire because the two co-occur. Linguistic variation comes to have social meaning through the co-occurrence of certain variants with perceived categories of speaker, or with certain genres of speech event” (Bell & Gibson, 2011). As demonstrated in the data above, the co-occurrence of diphthongization and other indie voice features with Halsey’s performance, as opposed to the lack of co-occurrence in her speaking, allows her singing to be indexical of the indie voice persona.

5 Conclusion

This paper has covered the history of sociolinguistics in popular song, the description and analysis of indie voice both generally and specifically looking at diphthongization, and sociolinguistic theories as to why indie voice is used. Goals for further research in formally describing indie voice may be determining whether measuring formants within a song, given the amount of intervening noise provided by instrumentals, is a reliable methodology in order to not rely on impressionistic classification, utilizing more data (longer speech, more songs), searching if there is correlation between indie voice and topic of communication, and looking at multiple users of indie voice and see what features they share with each other, and which they don’t.

References

- Ace Linguist (2021, May 3). *Dialect Dissection: Indie Voice/Cursive Singing - The Definitive Post*. Ace Linguist; Blogger. <https://www.acelinguist.com/2021/05/dialect-dissection-indie-voicecursive.html>
- Beal, Joan (2009). “You’re Not from New York City, You’re from Rotherham.” *Journal of English Linguistics*, 37(3), 223–240. <https://doi.org/10.1177/0075424209340014>
- Bell, Allan and Andy Gibson (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics*, 15(5), 555–572. <https://doi.org/10.1111/j.1467-9841.2011.00517.x>
- Blake, David K. (2012). Timbre as Differentiation in Indie Music. *Music Theory Online*, 18(2). <https://mtosmt.org/issues/mto.12.18.2/mto.12.18.2.blake.php>
- Burkhart, Benjamin (2017) Popular Music Analysis and Social Semiotics: The Case of the Reggae Voice. In: Merrill J. (eds) *Popular Music Studies Today*. Systematische Musikwissenschaft. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-17740-9_4
- Coupland, Nikolas (2011). Voice, place and genre in popular song performance. *Journal of Sociolinguistics*, 15(5), 573–602. <https://doi.org/10.1111/j.1467-9841.2011.00514.x>

- Jansen, Lisa and Michael Westphal (2017). Rihanna Works Her Multivocal Pop Persona: A Morpho-syntactic and Accent Analysis of Rihanna's Singing Style: Pop culture provides rich data that demonstrate the complex interplay of World Englishes. *English Today*, 33(2), 46-55. <http://dx.doi.org/10.1017/S0266078416000651>
- Jones, Colin, Schellenberg, Murray and Bryan Gick. (2017). Indie-Pop Voice: How a Pharyngeal/Retracted Articulatory Setting May Be Driving a New Singing Style. *Canadian Acoustics*, 45, 180-181.
- Schilling-Estes, Natalie (1998). Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society*, 27(1), 53–83.
- Simpson, Paul (1999). Language, Culture and Identity: with (another) Look at Accents in Pop and Rock Singing. *Multilingua*, 18(4), 343-367. <https://doi.org/10.1515/mult.1999.18.4.343>
- triple j (2016a). Halsey - "Hold Me Down" (live for Like A Version) [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=8R_drHL6o9E
- triple j (2016b). Halsey covers Justin Bieber "Love Yourself" for Like A Version [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=_1pUqmaPsFA&t=16s
- Trudgill, Peter (1997). Acts of Conflicting Identity: The Sociolinguistics of British Pop-song Pronunciation. In *Sociolinguistics* (pp. 251–265). Macmillan. <https://link.springer.com/book/10.1007%2F978-1-349-25582-5>
- Ugwu, Reggie (2015, October 6). *Selena Gomez's "Good For You" And The Rise Of "Indie Pop Voice."* BuzzFeed News; BuzzFeed News. <https://www.buzzfeednews.com/article/reggieugwu/what-is-indie-pop-voice>
- Zhang, Qing (2005). A Chinese yuppie in Beijing: phonological variation and the construction of a new professional identity. *Language in Society*, 34(3), 431-466.