Studia Linguistica Universitatis Iagellonicae Cracoviensis 138 (2021): 227–245
doi:10.4467/20834624SL.21.018.14861
www.ejournals.eu/Studia-Linguistica

MATEUSZ URBAN
Jagiellonian University in Kraków
mateusz.urban@uj.edu.pl
ORCID: 0000-0002-1155-7889

FRANKLIN D. ROOSEVELT AND THE AMERICAN THEATRE STANDARD: THE LOW VOWELS

Keywords: acoustic phonetics, north-eastern American standard, American Theatre Standard, Franklin D. Roosevelt

Abstract

Franklin D. Roosevelt's accent is often used as an illustration of the elite pronunciation heard among the north-eastern US upper classes until roughly the mid 20th century. Known under several names and often considered a mixture of British and American features, this variety is frequently identified with the American Theatre Standard, a norm popularized by acting schools in the early 20th century. Working on the assumption that Roosevelt is an exemplar of the north-eastern standard, the aim of the current study is a preliminary acoustic exploration of his accent with the aim of assessing the plausibility of such comparisons, focusing on the DRESS, TRAP, BATH, START and LOT vowels. Density plots created based on F1 and F2 measured in eight radio speeches were used to examine the relative position of these vowels in the vowel space. Linear mixed-effects regression was then used to model F1 and F2 in selected pairs of vowels to determine whether the differences along the two formant dimensions are significant. The results confirm a conclusion reached in an earlier auditory study (Brandenburg, Braden 1952) according to which Roosevelt's BATH was variable between [æ] and a lower and retracted [a], a vowel quality found in Eastern New England and in American Theatre Pronunciation. At the same time, a merged START/LOT vowel in Roosevelt's speech makes it unjustified to fully identify his accent with the latter variety.

1. Background and motivation of the study

In view of his social standing and place of origin, Franklin D. Roosevelt's pronunciation is frequently given by popular sources (e.g. MacNeil, Cran 2005: 50; Aschmann 2018) as an example of the upper-class north-eastern American English standard of

pronunciation.[1] The accent, notable for its non-rhoticity, was the prestigious norm until the mid 20th century when it was gradually supplanted in this role by its *r*-ful contender from the (Mid)west (Bonfiglio 2002; White 2020). Furthermore, the variety has quite often been identified or linked with the pronunciation standard for stage and screen actors laid out most famously in Edith Skinner's handbook (1942/1990).

The current study aims to evaluate the plausibility of comparing the north-eastern standard to the theatre standard. Taking Roosevelt (henceforth, FDR) as an exemplar of the former, an acoustic overview of his low monophthongs will be carried out to determine the number of vowel categories and their arrangement in the vowel space and to evaluate the likelihood that he could qualify as a speaker of the stage accent in the sense of Skinner (1942/1990). In particular, the relative position of the DRESS, TRAP, BATH, START and LOT vowels will be estimated. The focus on these vowels is justified by their exceptional configuration in stage pronunciation compared to other American varieties.

On a more general level, the article is meant as a modest contribution to an increasingly popular line of phonetic research focusing on archival data (Hickey 2016). FDR's case is exceptional in this context, because the sheer number of recordings of his voice makes it possible to select enough good-quality material for a reliable analysis. At the same time, by focusing on the individual, we follow in the footsteps of Jonathan Harrington and his colleagues' (Harrington et al. 2000, 2005; Harrington 2006, 2007), Przedlacka, Ashby (2019), or the authors of the studies collected in Beaman, Buchstaller (2021). Unlike most of these researchers, however, we do not adopt a longitudinal perspective, although in principle such a study of FDR's speech would be possible.

The data and R code needed to replicate the analyses in this paper as well as full model summaries are archived as an Open Science Framework project at https://osf.io/jy2fe/.

2. General information

2.1 "Mid-Atlantic" English in the popular media, in theatre studies and in linguistics

This section offers a review of popular and academic sources that discuss the north-eastern upper-class standard and other varieties that are (allegedly) related to it. The overall picture is rather confusing, part of which is because the authors use a variety of names to describe the accent they are discussing, and it is not always clear if they mean the same entity.

In the popular media (see e.g. Safire 1987; Smith 2012; Taylor 2013; Fallows 2015a, 2015b; Abad-Santos 2017; Rathe 2020), the accent is most commonly referred to as

[1] I would like to thank sincerely the three anonymous reviewers for their helpful comments and valuable suggestions, which made me revise significant portions of my analysis. All the remaining errors are my own.

"Mid-Atlantic/Transatlantic English", to emphasize its mixed Anglo-British character, but among many other epithets cited or used by the authors of these publications we find "faux-British elocution", "(Katharine Hepburn's) fake accent", "an American-British hybrid", "phony-British announcer voice", "that weirdo announcer-voice accent", "the Locust Valley/Larchmont/boarding school, etc. lockjaw", "the Boston Brahmin accent", or "American aristocratic speech". These expressions give us some idea as to the current attitudes towards "Mid-Atlantic" English, and to some extent about its geographical scope. This diversity is matched by the variety of names of alleged speakers – prominent political figures, actors, or radio and TV announcers – cited by popular authors. Beside FDR himself, oft-repeated examples include his wife Eleanor (Smith 2012), William F. Buckley (Safire 1987; Fallows 2015a; Rathe 2020), Jacqueline Kennedy Onassis (Fallows 2015a; Abad-Santos 2017), Katharine Hepburn (Safire 1987; Taylor 2013; Fallows 2015a) and Cary Grant (Taylor 2013).[2] While the style appears to be "instantly recognizable" (Fallows 2015b), the authors usually gloss over the phonetic details (apart from non-rhoticity),[3] and only vaguely discuss internal variation. This is not surprising in texts written by non-linguists for the general public, but the take-home message is that the accent can be defined by the list of its speakers (ignoring the fact that their backgrounds vary) and *r*-lessness.[4]

In theatre studies "Mid-Atlantic/Transatlantic English" may denote the target accent in speech and voice training for actors and singers (e.g. LaBouff 2008: 241–254), which is more commonly known as "Good (American) Speech" (Skinner 1942/1990; McLean 1952), or "the Theatre Standard" (Withers-Wilson 1993). It was introduced in the early 20th century by William Tilly under the name "World English" to be taught in schools, and soon adopted as a standard in acting instruction (Knight 2000; Deacon 2007). Unlike the sources discussed in the preceding paragraph, the manuals that propagate the theatre standard include fine phonetic detail (to be discussed in the next section), but because of their prescriptive nature, they paint a very uniform picture. Interestingly, however, according to its proponents, the accent was modelled on "the speech of the intelligent, cultivated classes" (McLean cited in Knight 2000: 35), and consequently the north-eastern upper-class standard and the theatre standard are occasionally equated (e.g. White 2020: 37–38).

Internal variation is more eagerly discussed by linguists, who refer to the north-eastern standard as the "Boston Brahmin accent" (Wolfram, Schilling 2015: 174), "New York Dainty" (White 2020), or, again, "Mid-Atlantic/Transatlantic English" (Queen 2015: 241–242).[5] In her account of the linguistic history of New York City,

[2] A good representation of popular views on "Mid-Atlantic" English as used by the north-eastern elite on the one hand, and on the other, in Hollywood, may be found in Wikipedia (2021), along with many more examples.

[3] Exceptions do exist, e.g. Abad-Santos' (2017) attempt at a linguistically-informed overview of Jacqueline Kennedy's accent.

[4] This overreliance on non-rhoticity in defining the accent is discussed from a linguistic viewpoint by McWhorter (2015).

[5] There are at least two other senses in which "Mid-Atlantic English" has been used by linguists. In Labov et al. (2006) it refers to the varieties spoken in the Mid-Atlantic cities (Philadelphia, Reading, Wilmington, and Baltimore), which differ from (traditional) NYC English (and

White (2020: 37–38) uses "Good Speech" as proxy for "New York Dainty", when comparing the latter to Received Pronunciation (RP) and (contemporary) Standard American English. At the same time she (2020: 35–36) observes that there were many regional versions of the standard, so that Katharine Hepburn's Connecticut Dainty was markedly different from the New York variant used by most actors. Furthermore, while discussing the use of the accent in the film industry, Queen (2015: 241) writes that "[t]he transatlantic accent was specifically taught to actors and involved stylistic variation of its own, primarily to distinguish higher-class and lower-class characters or to distinguish regional origins of characters". Aschmann (2018), who is technically not a linguist but is nevertheless very attentive to phonetic detail, uses a very narrow definition of Mid-Atlantic, so that for him both Theodore and Franklin D. Roosevelt qualify as speakers of this accent, but Eleanor Roosevelt "spoke almost pure Standard British English", whereas Katharine Hepburn is "clearly local, not Mid-Atlantic", and Jacqueline Kennedy "certainly does not have a Mid-Atlantic accent." Variation is also reported in Boberg's (2018, 2020, forthcoming) comprehensive studies of the developments in American film pronunciation that will be discussed in the next section, addressing some of these issues through the analysis of low vowels.

2.2 Low vowels in the American Theatre Standard and other varieties

In what follows, the term American Theatre Standard (ATS) is used exclusively to denote the accent codified by authors such as Skinner (1942/1990) and McLean (1952). Knight (2000: 41), who calls it "a homogenous accent contrived by speech teachers and actually spoken by no one", characterizes it as a mixture of RP with selected regional features of New England (Knight 2000: 34). This description echoes Elliott's (2000) conclusions:

> It would appear that Good Speech, the stage speech taught to American actors during the early- to mid-twentieth century, partly follows a New England accent as its norm and partly follows the British standard. The editors of the book claim that their stage accent differs significantly from RP (…) and insist that it is American. It is also conservative, maintaining a large number of phonological distinctions that have been merged by a majority of American speakers. In fact, it maintains more phonological distinctions than RP … In conclusion, it can be observed that Skinner's Good Speech is careful, conservative and non-rhotic. It is full of rare and minority usage, and as such, is a prestige norm rather than a standard norm. The target area of the prestige form apparently is intended to be New England. (Elliot 2000: 19)

Some of these idiosyncrasies are visible in Table 1, which compares ATS low vowels to those of RP, Eastern New England (ENE), and New York City English (NYC).

---

the theatre standard) by being rhotic. Furthermore, authors such as Modiano (1996) use it to describe an Anglo-American mixture used especially by European EFL learners.

This last accent is included here because along with ENE it provides a backdrop for the variety spoken by FDR, a New Yorker by birth who received his education from Groton School and Harvard College.

	DRESS	TRAP	BATH	PALM	START	LOT
ATS (Skinner 1942)	e	æ	a	ɑː	ɑə	ɒ
RP (Jones 1964)	e	æ	ɑː	ɑː	ɑː	ɒ
ENE ([Boston]; Johnson 2010)	ɛ	æ	eə (pre-nasal) / aː (other)	aː	aː	ɒː
NYC (Newman 2014)	ɛ	æ	ɛə	ɑə	ɑə	ɑ

Table 1. The low vowels in four varieties of English

Table 1 suggests that as regards low vowels ATS and RP agree on almost all points except for two features shared with NYC and ENE. First, ATS speakers are expected to use [ɑə] in START, as in NYC. Second, the recommended pronunciation of the BATH vowel is [a], a quality resembling that used in ENE PALM, START, and non-prenasal BATH. However, the data are admittedly simplified, especially as regards BATH and TRAP, i.e. the reflexes of Middle English short *a*, in ENE and NYC. A few additional comments are therefore in order.[6]

The BATH set comprises those words in which Middle English short *a* underwent tensing/lengthening, whereas TRAP represents the lax/short reflex. Varieties of English differ as to which contexts or words (if any) were affected by tensing, or in Labovian terms, which short-*a* system they use (Labov et al. 2006: 173–184), but there is also variation in terms of the resulting vowel quality. The BATH set in ATS (referred to as the Ask-words by Skinner 1942/1990) agrees with that of RP in that it includes mostly words in which the vowel is followed either by a voiceless fricative or by /m/ or /n/ followed by an obstruent.[7] The difference lies in the quality: whereas RP uses back [ɑ] (= PALM, START), the vowel postulated for ATS is front [a], which McLean (1952: 176) describes as "the lowest front vowel" and "intermediate between æ and ɑː but … closer to æ". In ENE the BATH vowel is traditionally pronounced as [aː] (or even [ɑː]) before a fricative, and raised to [eə] before a nasal (/m/ and /n/, and less commonly /ŋ/) with or without a following obstruent (the nasal system; Labov et al. 2006: 174). Similar raising affects BATH in NYC, although here the set traditionally includes words where the vowel precedes a voiceless fricative, a voiced plosive, or a front nasal (the split system; Labov et al. 2006: 173).

[6] For state-of-the-art reviews of this topic based on early sources, see Johnson, Durian (2016) on ENE and Coggshall (2017) on NYC.

[7] That this demarcation of BATH is not in perfect agreement with what is observed in ENE (Johnson 2010; see also Miller 1953) indicates that RP *must* have had a significant impact on ATS, contrary to what the proponents of the latter would like to claim (see Knight 2000 for an overview).

The debate on [a(:) ~ ɑ(:)] in BATH has a long history in America (Penzl 1940) and Grandgent (1899: 214) traces its prestigious status to the early 19th century. Miller's (1953) data collected in the 1940s and early 1950s indicate that by that time it was in steep decline in favour of [æ] even in ENE. There is some evidence to indicate that [a(:) ~ ɑ(:)] in BATH was also found in early 20th century NYC (see Coggshall 2017: 104 for an overview), but at least some of that may be attributed to hypercorrection in the direction of Tilly's "World English" taught in NYC schools at that time (Thomas 1932: 325). In the latter half of the 20th century Nunberg (1980: 171) observes that in Boston BATH with [ɑ:] "is a stigmatized feature when used by working-class speakers, but is regarded as an old-fashioned affectation when used by members of the upper class", whereas in New York "the pronunciation is not heard even among upper-class speakers".

Finally, while this is not the focus of this article, a few comments are needed on the relationship of ATS as prescribed by manuals to its realization in actors' speech. It is likely that the pronunciation styles used on stage and screen differed, especially in view of the obvious differences between the two media. Unfortunately, a detailed analysis of pronunciation in early 20th century theatre practice is largely impossible because of the scarcity of data. However, American film speech has recently become of interest to sociophoneticians, beginning with Elliott's (2000) study of (non-)rhoticity in the 1930s–1970s. More relevant to our topic, Boberg's (2018, 2020) two studies of on-screen speech reveal a shift away from Eastern patterns in favour of Western ones, including a gradual decline of the NYC split-*a* system in favour of the nasal system. In a follow-up book, Boberg (forthcoming: 126–127)[8] makes a few comments on "Mid-Atlantic" English as discussed above, referring, however, to actual onscreen practice rather than ATS. Among the features of the former he notes the pronunciation of BATH words with [a:] and an unrounded LOT vowel. Note that the latter is in stark contrast with [ɒ] advocated by Skinner (1942/1990).

2.3 Previous studies of FDR's accent

The only detailed phonetic analysis of FDR's pronunciation that I am aware of is that by Brandenburg and Braden (1952), who studied his overall voice quality and selected segmental features. They begin with a general remark that his pronunciation was that of the upper-class of the NYC area, rather than a "Groton-Harvard" accent, as popularly believed. At the same time, in their discussion of his pronunciation of BATH words they note what is possibly a trace of the time he spent in New England:

> He did not use the so-called "broad" *a*. He might say *ask* [æsk] or *vast* [væst], or he might say [ask] or [last] or [pasɪŋ]. Both of these pronunciations, as well as [ɑ], are accepted in eastern dialect. The tendency to use the [a] may have been developed during the period of Groton and Harvard, since [a] in "broad" *a* words is exceedingly common in New England, but seldom found in the New York area. In no sense was his pronunciation extremely broad eastern dialect. (Brandenburg, Braden 1952: 28)

[8] Since Boberg (forthcoming) has not appeared as of writing these words, I quote from the samples accessible via Google Preview.

There is no discussion of the other low vowels in Brandenburg and Braden's article, but among the transcriptions they provide is the word *congress* rendered as [kaŋgrəs], hinting at an unrounded LOT, although both ATS and RP have [ɒ] here (see in Table 1 above).

It is interesting to note in this context a comment on FDR's accent made by none other than William Tilly in his polemic with Jaime de Angulo in *Le Maître Phonétique*. De Angulo (1933: 12) suggested that to promote the use of the International Phonetic Alphabet in America the journal should not only publish Southern British transcriptions, which sound affected to American listeners, but also include transcriptions of the American variety. Interestingly, his postulate was seconded by Daniel Jones in a note under de Angulo's letter (1933: 12–13). In his response Tilly (1933: 57) cites a number of ATS manuals to support his claim that the transcriptions published in the journal are adequate for the American audience and includes the following suggestion to de Angulo: "ɪf hi: wɪl lɪsn tə ðə ɹeɪdɪoʊ spi:tʃɪz əv pɹezɪdənt *ɹoʊzəvelt ənd hɪz waɪf, hi: wɪl əgen hɪə ðɪs 'sʌðən ɪŋglɪʃ daɪəlekt.'"

3. Data

3.1 The nature of the material

The Fireside Chats (FCs) were chosen for the current analysis for several reasons, the most important being their overall superior sound quality compared to the other recordings of FDR's voice. Furthermore, they had been recorded in comparable conditions, in a quiet room and by means of a professional microphone. The fact that they form a series of speeches with the same audience in mind made it likely that the style and manner of speaking and the overall delivery would be comparable, even if the content was not. This approach to material collection was previously adopted by Harrington et al. (2000), who chose the Queen's Christmas broadcasts for similar reasons.

When it comes to the genre, the FCs are examples of scripted speech. It is known, also from FDR's own reports (see Brandenburg 1949), that they were prepared by a team of advisers and then revised carefully by the president, who was very much aware of the importance of the prosodic organization of speech. Consequently, the speeches are written in a formal register and both their content and delivery are highly controlled. One might therefore infer that the recordings provide a glimpse into FDR's less spontaneous pronunciation style most likely different from his everyday conversational style.

3.2 Recording selection and quality

Eight FC recordings were selected from the "Recorded speeches and utterances of Franklin D. Roosevelt, 1920–1945" collection made available online by the FDR Presidential Library and Museum (n.d.). The most important criterion was recording quality, but the intention was also to obtain a sample spanning the entire period when

File	Informal title	Recording and broadcast date	Duration (mm:ss)	
			original	after pre-processing
afdr013.mp3	FC1: On the Banking Crisis	12 March 1933	13:04	12:40
afdr022.mp3	FC7: On the Works Relief Program	28 April 1935	28:38	26:54
afdr096.mp3	FC10: On New Legislation (incomplete)	12 October 1937	16:12	14:41
afdr107.mp3	FC12: On the Recession	14 April 1938	45:15	37:54
afdr149.mp3	FC14: On the European War	3 September 1939	14:29	10:56
afdr223.mp3	FC16: On the "Arsenal of Democracy"	29 December 1940	38:49	36:58
afdr238.mp3	FC18: On the Greer Incident	11 September 1941	28:04	27:43
afdr278.mp3	FC25: On the Fall of Mussolini	28 July 1943	29:01	28:25
		Total duration:	213:32	196:11

Table 2. The recordings chosen for analysis

the FCs were broadcast. The selected recordings were downloaded as 128kbps/44kHz mono MP3 files listed in Table 2. The titles and the numbering are taken from the transcripts published by the Miller Center (2016).

The lineage of the files, that is the quality of the analogue source material and the way it was digitized, except for the parameters of MP3 compression, remain unknown. Furthermore, the very use of lossy compression makes the files deficient compared to their source, be it digital or analogue, although studies have shown that MP3 compression at 128kbps does not significantly affect F1 and F2 measurements (Bulgin et al. 2010). In view of this and given the overall good quality of the recordings, the lack of access to lossless or uncompressed copies, and the existence of earlier phonetic studies utilizing MP3 audio (e.g. Przedlacka, Ashby 2019), the files were deemed adequate for the purpose of this study.

4. Method

Semi-automatic forced-alignment and vowel extraction was performed in DARLA (Reddy, Stanford 2015), using the Montreal Forced Aligner (McAuliffe et al. 2017) and FAVE-Extract (Rosenfelder et al. 2014). Vowel extraction was performed with DARLA's default settings, excluding stop-words (listed on the DARLA website), unstressed vowels, and vowels whose F1 and F2 bandwidths exceed 300Hz. Praat software (Boersma, Weenink 2021) was used for annotation checking and correction.

The recordings were first converted to WAV format. While this does not improve the signal parameters, the reliance on uncompressed audio ensures that further modifications to the file can be introduced without generation loss caused by repetitive application of the compression algorithm upon saving the file. Furthermore, WAV enables greater annotation precision than MP3.

The WAV files were then edited to remove any speech not by FDR himself, such as introductions and closing remarks by the radio announcer. Also removed were false starts, self-corrections, mispronunciations, and portions of the recordings where the president's voice was inaudible or distorted. As a result, the overall duration was reduced by 17 minutes and 21 seconds. Transcripts of the eight speeches were obtained from the Miller Center (2016), corrected to remove any inconsistencies with the recordings and saved as plaintext files without punctuation or upper-case letters, as per DARLA recommendations.

The WAV files along with corresponding plaintext transcriptions were then fed into DARLA for forced-alignment in two passes. The resulting TextGrid files were revised manually at the phrase level after the first pass, and at the word- and phone-level after the second pass. The latter stage included the correction of transcription errors introduced by DARLA's reliance on the CMU (Carnegie Mellon Univesity) pronouncing dictionary, which is largely based on current Standard American English. One of the consequences is that all recordings are treated as if they represented a rhotic accent, and the existence of a rhotic segment is assumed even in cases when there is none. The correction involved removing the "R" annotation and readjusting

the segment boundaries. The only exception was the NURSE vowel, in which the vowel + /r/ sequence is treated as a unitary segment "ER" by the CMU dictionary. In certain cases the transcription of the vowel itself had to be revised. Examples include obvious Briticisms in FDR's pronunciation, such as the PRICE vowel in *either* and *neither*, instead of FLEECE, or the words *ma'am* and *Nazi(s)*, for which the CMU dictionary assumes [æ] and [ɑ:] respectively, while FDR's pronunciation was quite the opposite. The location of word stress was also corrected if necessary, because automatic stress assignment creates some issues, particularly with noun-verb pairs differentiated by stress placement.

The revised TextGrids along with WAVs were then uploaded to DARLA for vowel extraction. This procedure yielded spreadsheet files with vowel data (n=11,279), including unnormalized formant measurements taken at 20%, 35%, 50%, 65%, 80% of token duration and information about context.

The data was then manipulated manually by removing all suspicious tokens, including those shorter than 50 ms, possibly unstressed/reduced, or tokens for which formant readouts were unreliable. The decision was also made to ignore all words with /aʊə/ or /aɪə/, as they were subject to varying degrees of smoothing in the recordings, as well as all tokens of the word *British*, in which a tendency for metathesis was observed. The resulting dataset (n=10,045) was then manually organized into vowel classes. Since ATS serves as the reference accent, the vowel class division strictly follows Wells's standard lexical sets (Wells 1982). This is particularly relevant for the TRAP and BATH sets, which in ENE and NYC are demarcated differently than in RP and ATS (see section 2.2 above). However, BATH itself is plotted as two subsets, DANCE (tokens before nasal-obstruent clusters) and PASS (prefricative). This is motivated by the tendency of prenasal short-*a* to undergo raising and fronting in American English (see Labov et al. 2006: 174–175). Furthermore, CLOTH is subdivided into BORROW (prerhotic) and BOSS (other), due to an audible difference in pronunciation. The words were assigned to the sets by consulting Wells (1982 1: 127–168), Kenyon (1940) and Kenyon, Knott (1949).

As shown in Table 3 the distribution of tokens in the vowel classes is uneven, which is expected given the source of data. The PALM set is strikingly underrepresented with only two items: *ma'am* and *Colorado*. Beside these, there were also nineteen tokens of *Nazi*(*s*) and one of *Nazism*, but FDR pronounced these with the TRAP vowel, so they are included in that set. Because of the small token count, the PALM vowel was ignored in our analysis.

This dataset was then analyzed statistically in R (R Core Team 2021) with the help of RStudio 2021.09.0, build 351 (RStudio Team 2020). For plotting, the nucleus and the glide of each token were represented by F1 and F2 at 35% and 65% of token duration respectively (approximating the 1/3 and 2/3 measurement points used e.g. by Kendall, Fridland 2017). For modelling, only the nucleus was taken into consideration. Vowel plots were created with the help of the ggplot2 package, ver. 3.3.5 (Wickham 2016), to obtain a general picture of FDR's overall vowel system and the word-class differences in the low vowels. Then models of mean F1 and F2 for selected pairs of vowels using linear mixed effects regression were fit using lme4

KIT	1024	PASS	85	THOUGHT	275	NEAR	219
		DANCE	20				
DRESS	1605	BOSS	80	GOAT	650	SQUARE	95
		BORROW	36				
TRAP	771	NURSE	430	GOOSE	342	START	153
LOT	413	FLEECE	592	PRICE	468	NORTH	158
STRUT	818	FACE	1212	CHOICE	70	FORCE	170
FOOT	64	PALM	2	MOUTH	213	CURE	80

Total = 10,045

Table 3. Representation of vowel classes according to lexical sets

ver. 1.1.27.1 (Bates et al. 2015) and the significance of differences based on lexical set was tested using the lmerTest package, ver. 3.1.3, which derives the p-values using the Satterthwaite approximation (Kuznetsova et al. 2017). Marginal and conditional R^2 values were computed using the r.squared() function from the MuMIn package, ver. 1.43.17 (Bartoń 2020).

5. Results

5.1 Overview

Some preliminary conclusions may be drawn from Figure 1, which presents vowel trajectories by plotting each vowel's mean nucleus and glide. The NEAR, SQUARE and CURE vowels are diphthongs (expectedly so, given that FDR is a non-rhotic speaker), and so are PRICE, MOUTH and CHOICE, as well as FACE and GOAT. These features are overall consistent with both ATS and RP. There is also noticeable movement in FOOT, as well as NORTH, FORCE, THOUGHT and CLOTH, which coincide in the vowel space. The same is the case with LOT and START, which appear to be merged, unlike in ATS. Another difference from the theatre standard (or from NYC English) is the lack of a centring trajectory in START. BORROW seems to occupy an intermediate position between STRUT and START/LOT, but its trajectory is in the direction of the latter.

Focusing on low front monophthongs, little movement may be observed in DRESS, TRAP, and the two subsets of BATH, that is, PASS and DANCE, but the means of the latter two are clearly distinct, with DANCE pronounced on average higher than PASS. This is consistent with the tendency for raising in prenasal contexts mentioned above.

For more detail, let us examine the distribution of individual DANCE and PASS tokens against density plots of FDR's other low vowels in Figure 2. Prenasal TRAP is shown as individual tokens, and the distribution of the other TRAP words is presented as a density plot.

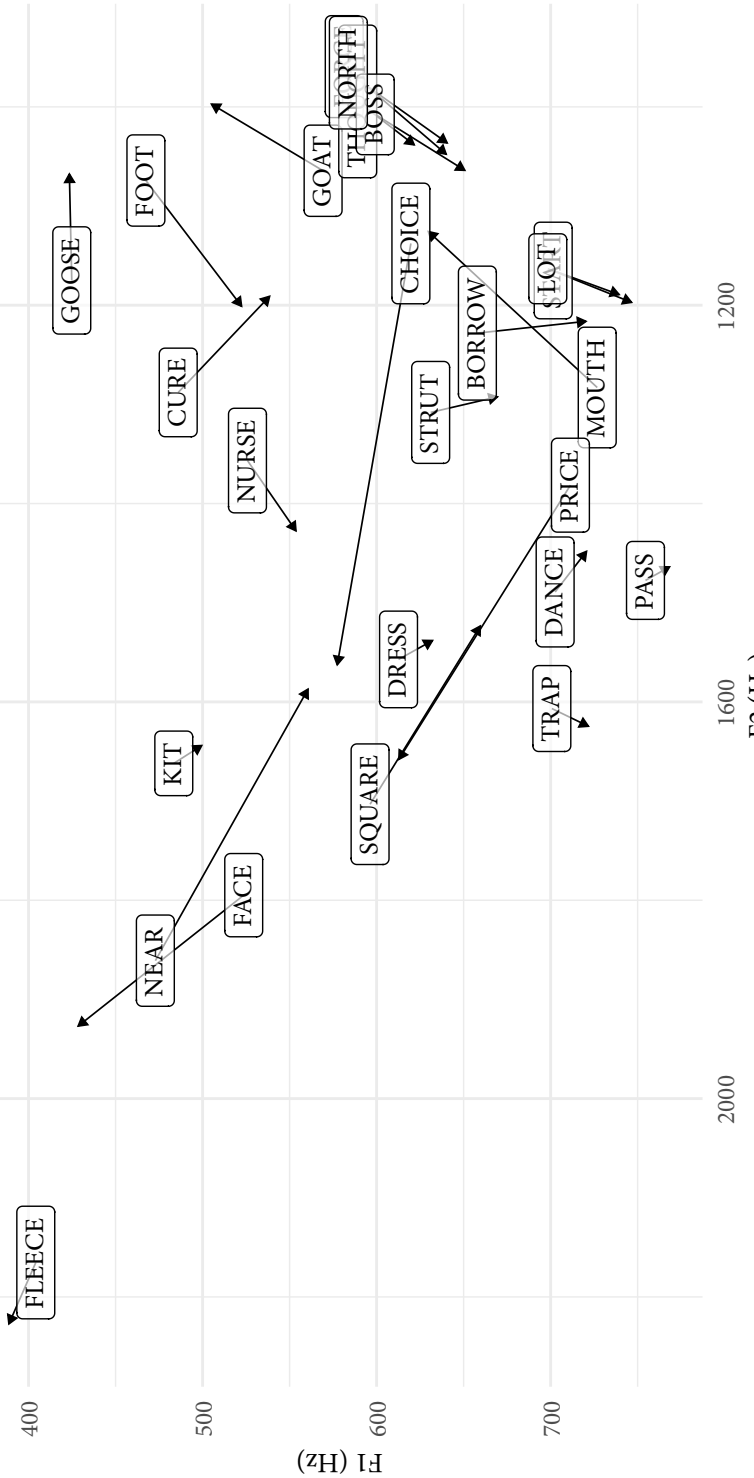

Figure 1. Mean F1 and F2 (in Hz) measured at 35% (nucleus) and 65% (glide) of token duration

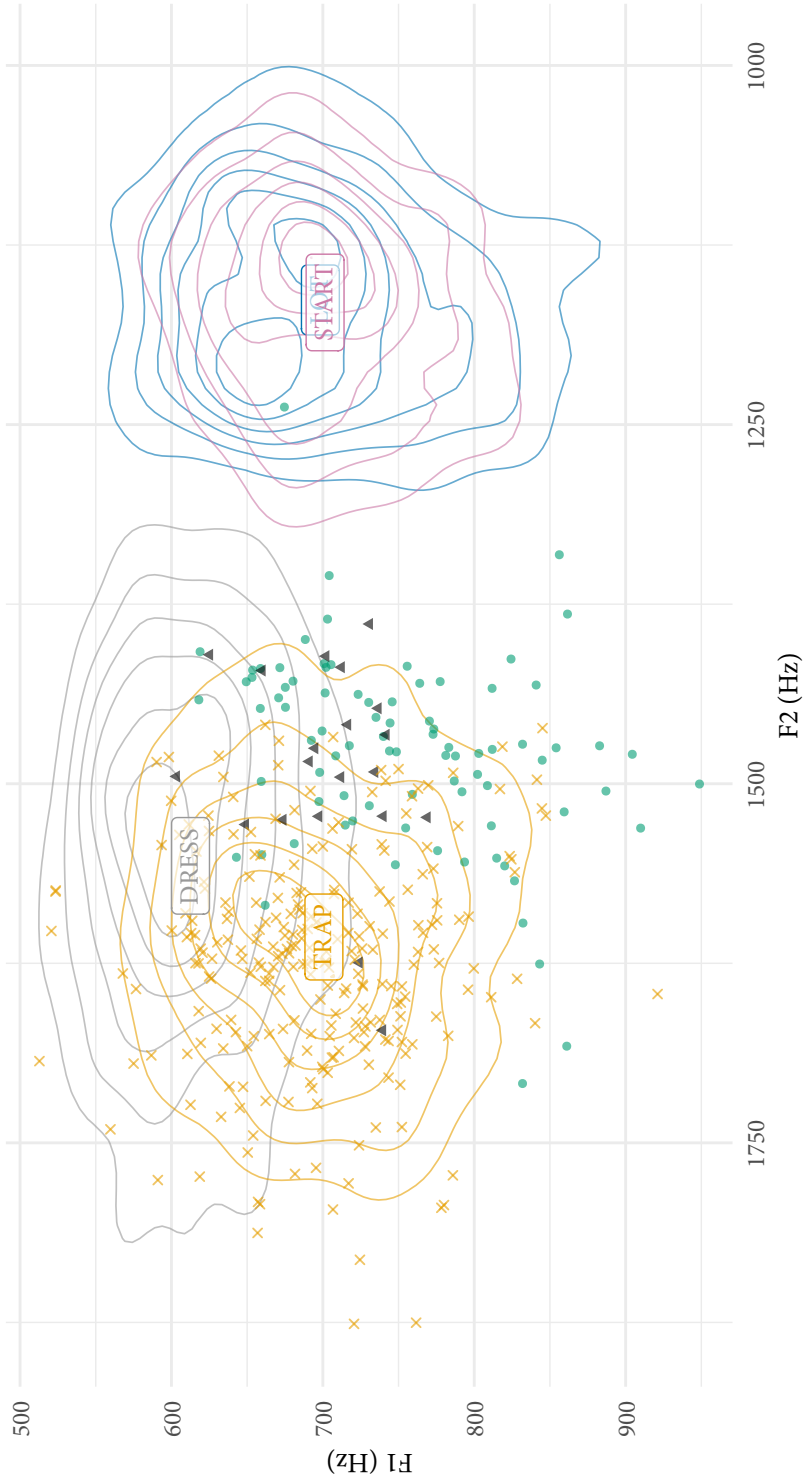

Figure 2. DRESS, non-prenasal TRAP, START and LOT density plots with individual tokens of DANCE, PASS, and prenasal TRAP (black triangles = DANCE; green circles = PASS; yellow asterisks = prenasal TRAP; labels indicate means of the entire sets)

The clustering of START and LOT on the right strengthens the conclusion that they are qualitatively merged. In the front area, DRESS and TRAP show considerable allophonic variation and partial overlap, with some raising of the latter. However, the data do not follow the exact pattern of a nasal system, since prenasal TRAP tokens (252 in total) cluster around the centre of non-prenasal TRAP. However, some effect of nasality is visible in the two subsets of BATH: PASS is variable so that some tokens are relatively low and others raised, but DANCE is clearly raised. At the same time, most of its tokens are retracted compared to prenasal TRAP (the two clearly front instances near the centre of TRAP are both tokens of *example*). This may suggest that DANCE and TRAP behave as separate classes, although the small number of DANCE tokens weakens any conclusions.

The solitary token from the PASS set that falls within the distribution of START/LOT is the only instance of the word *rather* in our material. Since this qualitatively belongs to that latter vowel, it will be excluded from further analysis.

5.2 Statistical models

As explained above, mixed effects models were used to assess the significance of lexical-set-based differences between selected pairs of vowels in terms of F1 and F2 in a manner similar to Johnson, Durian (2016).[9] Only interconsonantal tokens were included.[10] The following pairs of vowels were analysed: START and LOT (as a control case given their apparent clustering), DRESS and TRAP, PASS and TRAP, and PASS and START/LOT. DANCE was ignored, because the small token count does not ensure reliable results. The models estimated mean F1 and F2 and included the fixed effects of lexical set as well as place and manner of the preceding and following consonant with by-recording and by-word varying intercepts due to sound quality differences and uneven lexical representation respectively. In those cases where the effect of lexical set turned out to be significant the conclusion to be drawn is that membership in a given lexical set explains a significant difference between the two vowels in terms of height (F1) or backness (F2) when the other variables are taken into account.[11]

---

[9] Overall the method resembles the one used by Nycz (2011), although in her approach significance of lexical set is tested by means of likelihood ratio tests rather than the Satterthwaite approximation. We prefer the latter as it is believed to be more accurate (Sonderegger 2020: 327). Although recent studies recommend other methods of estimating the degree of vowel merger, esp. the Pillai score (e.g. Kelley, Tucker 2020), mixed effects regression is preferred for our data, which clearly violate the independence criterion (multiple tokens per word and multiple tokens per recording). Note, however, that while the Pillai score measures vowel overlap, the method used here indicates vowel difference, so the two are not directly comparable.

[10] A reviewer indicates that following sonorants (liquids and nasals) are known to influence vowel formants in American English. The models take that into account by including following manner as a fixed effect. Furthermore, removing preliquid tokens did not change the conclusions and therefore we report models fitted to the fuller dataset. The influence of nasals on BATH is factored out by considering PASS and excluding DANCE.

[11] Full model summaries are available at the project's OSF archive: https://osf.io/jy2fe/.

Vowel pair	n	F1		F2	
		Difference (Hz)	p	Difference (Hz)	p
START : LOT	139 : 397	9.32	NS	-0.27	NS
DRESS : TRAP	1479 : 690	-93.82	< .001	-8.92	NS
PASS : TRAP	72 : 690	14.42	NS	-72.25	< .001
PASS : START/LOT	94 : 536	47.62	< .001	324.56	< .001

Table 4. Estimated F1 and F2 differences for pairs of low vowels. The differences along both dimensions are given relative to the second vowel class in the pair

The relevant aspects of the models are summarized in Table 4. The analysis leads to the following conclusions:

1. there is no evidence in our dataset of a significant qualitative distance between START and LOT;[12]
2. DRESS is significantly higher than TRAP, but there is no significant difference in backness;
3. PASS is significantly backer than TRAP, but no significant difference was found in height;
4. BATH is significantly fronter and lower than START/LOT.

6. Discussion and prospects

The conclusion in (1) above shows that the qualitative difference found in RP and prescribed in ATS cannot be confirmed in FDR's speech. Auditorily his START/LOT vowel is clearly unrounded, which ties in with Brandenburg and Braden's (1952) transcription, and contradicts Tilly's (1933) comment. In connection to our main research question, the qualitatively merged START/LOT indicates that to the extent that FDR's accent is a good illustration of the north-eastern standard, the latter was certainly not identical to ATS. The conclusion in (4) is in line with expectations and needs no further discussion.

The conclusions in (2) and (3), whereby the vowels in each pair are only distinguished along one of the dimensions, are consistent with their acoustic variability. In other words, the consequence of their partial overlap is that lexical set membership as postulated for ATS (largely on the basis of RP) does not always predict either the height (PASS vs. TRAP) or the backness (DRESS vs. TRAP) difference between vowel categories.

[12] Because F2 weighs more heavily on vowel quality than F1, when both are measured on a linear scale, Euclidean distances in raw Hz are difficult to interpret, and therefore they are not included in the chart. However, in the case of START and LOT the distance is so small (9.33 Hz) that it reinforces the conclusion of qualitative identity. I am grateful to one of the reviewers for raising this point.

As for Brandenburg and Braden's (1952) claim that FDR's BATH vowel was variable between [æ] and [a], this is indeed visible in the plot in Figure 2 in relation to the prefricative tokens (PASS). Prenasal BATH (= our DANCE) appears to be raised in its entirety, but at the same time largely retracted compared to prenasal TRAP. However, due to the low token count for DANCE, only PASS could be shown statistically to be different on average from TRAP. However, the patterns observed in our data only partly agree with the ones described in ATS manuals. To put it differently, Skinner's (1942/1990) Ask-words are never pronounced with British [ɑ] by FDR (with the sole exception of *rather*) and only occasionally with [a].

The main reason for the discrepancies between FDR's pronunciation and ATS is that the former is a spoken variety, whereas the latter is a prescriptive construct. In other words, the phonetic features of ATS were determined by Tilly and his followers based on their perceived prestige, whereas FDR acquired his accent naturally through exposure to the British-influenced upper-class patterns of NYC and ENE. However, the variety he may have heard from his teachers was certainly not ATS. Except for a short span at a public school in Germany, he was home-schooled until 1896 when he entered Groton. He completed his education with a BA degree in history from Harvard in 1903, that is fourteen years before Tilly came to New York.

The question remains open as to exactly whom McLean means when she praises "the speech of the intelligent, cultivated classes", or in other words to what extent the speech patterns of people such as FDR helped determine the features of ATS. British influences are obvious in both, but ATS is far closer to its elder overseas sibling both in systemic and phonetic terms. It cannot be ruled out that being an Australian, Tilly left his own British-like stamp on the variety he helped create.

It should be noted that due to space limitations the current article overlooks several issues. Further research is needed to understand fully the behaviour of TRAP and BATH, especially to determine the patterns of phonetic or lexical conditioning of the raised variants. Moreover, the analysis above ignored the aspect of vowel duration. While START and LOT overlap qualitatively, they may still be distinguished in terms of length as is the case e.g. for some ENE speakers contemporaneous with FDR studied by Johnson and Durian (2016). Finally, the focus of this article has been very narrow, leaving other vowel categories in FDR's accent open to investigation.

References

Abad-Santos A. 2017. Jackie Kennedy's strange, elegant accent, explained by linguists. – *Vox* 7 February 2017. [https://www.vox.com/culture/2017/2/7/14442410/jackie-kennedy-accent-natalie-portman; accessed: 16 August 2021].

Aschmann R. 2018. American English dialects. [https://aschmann.net/AmEng/; accessed: 16 August 2021].

Bartoń K. 2020. MuMIn: Multi-model inference. R package version 1.43.17. [https://CRAN.R-project.org/package=MuMIn; accessed: 14 November 2021].

Bates D., Mächler M., Bolker B., Walker S. 2015. Fitting linear mixed-effects models using lme4. – *Journal of Statistical Software* 67.1: 1–48.

Beaman K.V., Buchstaller I. (eds.). 2021. *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies*. New York: Routlege.

Boberg C. 2018. New York City English in film: Phonological change in reel time. – *American Speech* 93.2: 153–185.

Boberg C. 2020. Diva diction: Hollywood's leading ladies and the rise of General American English. – *American Speech* 95.4: 441–484.

Boberg C. (forthcoming). *Accent in North American film and television: A sociophonetic analysis*. Cambridge: Cambridge University Press. [Google Preview; https://books.google.pl/books?id=h2lHEAAAQBAJ; accessed: 24 November 2021].

Boersma P., Weenink D. 2021. *Praat: Doing phonetics by computer.* [Computer program. Version 6.1.41; https://www.fon.hum.uva.nl/praat/; accessed: 25 March 2021].

Bonfiglio T.P. 2002. *Race and the rise of Standard American*. Berlin, New York: De Gruyter Mouton.

Brandenburg E. 1949. The preparation of Franklin D. Roosevelt's speeches. – *Quarterly Journal of Speech* 35.2: 214–221.

Brandenburg E., Braden W.W. 1952. Franklin D. Roosevelt's voice and pronunciation. – *Quarterly Journal of Speech* 38.1: 23–30.

Bulgin J., De Decker P., Nycz J. 2010. Reliability of formant measurements from lossy compressed audio. [Poster presented at The British Association of Academic Phoneticians Colloquium, University of Westminster; https://research.library.mun.ca/684/; accessed: 12 October 2021].

Coggshall E.L. 2017. *Short-*a *in the sixth borough: A sociophonetic analysis of a complex phonological system in Jersey City.* [PhD Diss., New York University].

De Angulo J. 1933. lak əv əmɛɹɪkn tɛksts. – *Le Maître Phonétique* 48.41: 12–13.

Deacon D. 2007. World English? How an Australian invented 'Good American Speech'. – Deacon D., Damousi J. (eds.). *Talking and listening in the age of modernity: Essays on the history of sound*. Canberra: ANU Press: 73–82.

Elliott N.C. 2000. *A sociolinguistic study of rhoticity in American film speech from the 1930s to the 1970s*. [PhD Diss., Indiana University].

Fallows J. 2015a. Language mystery redux: Who was the last American to speak this way?. – *The Atlantic*, 6 June 2015. [http://www.theatlantic.com/national/archive/2015/06/language-mystery-redux-who-exactly-was-the-last-american-to-speak-like-this/395090/; accessed: 18 November 2021].

Fallows J. 2015b. That weirdo announcer-voice accent: Where it came from and why it went away. – *The Atlantic*, 8 June 2015. [https://www.theatlantic.com/national/archive/2015/06/that-weirdo-announcer-voice-accent-where-it-came-from-and-why-it-went-away/395141/; accessed: 18 November 2021].

FDR = Presidential Library and Museum. n.d. *FDR Audio Recordings*. [https://www.fdrlibrary.org/utterancesfdr; accessed: 31 March 2021].

Grandgent C.H. 1899. From Franklin to Lowell. A century of New England pronunciation. – *Publications of the Modern Language Association of America* 14.2: 207–239.

Harrington J. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. – *Journal of Phonetics* 34.4: 439–457.

Harrington J. 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. – *Laboratory Phonology* 9: 125–144.

Harrington J., Palethorpe S., Watson C. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. – *Journal of the International Phonetic Association* 30.1–2: 63–78.

Harrington J., Palethorpe S., Watson C. 2005. Deepening or lessening the divide between diphthongs: An analysis of the Queen's annual Christmas broadcasts. – Hardcastle W.J., Beck J.M. (eds.). *A figure of speech*. Mahwah (NJ): LEA: 227–261.

Hickey R. (ed.). 2016. *Listening to the past: Audio records of accents of English*. Cambridge: Cambridge University Press.

Johnson D.E. 2010. *Stability and change along a dialect boundary: The low vowels of Southeastern New England*. Durham: Duke University Press.

Johnson D.E., Durian D. 2016. New England. – Hickey R. (ed.). *Listening to the past: Audio records of accents of English*. Cambridge: Cambridge University Press: 257–297.

Jones D. 1964. *An outline of English phonetics*. [9th edition]. Cambridge: W. Heffer & Sons.

Kelley M.C., Tucker B.V. 2020. A comparison of four vowel overlap measures – *The Journal of the Acoustical Society of America* 147.1: 137–145.

Kendall T., Fridland V. 2017. Regional relationships among the low vowels of U.S. English: Evidence from production and perception. – *Language Variation and Change* 29.2: 245–271.

Kenyon J.S. 1940. *American pronunciation: A textbook of phonetics for students of English*. [8th edition]. Ann Arbor: George Wahr.

Kenyon J.S., Knott T.A. 1949. *A pronouncing dictionary of American English*. Springfield (MA): G. & C. Merriam.

Knight D. 2000. Standard Speech: The ongoing debate. – *Voice and Speech Review* 1.1: 31–54.

Kuznetsova A., Brockhoff P.B., Christensen R.H.B. 2017. lmerTest Package: Tests in linear mixed effects models. – *Journal of Statistical Software* 82.13: 1–26.

LaBouff K. 2008. *Singing and communicating in English: A singer's guide to English diction*. New York: Oxford University Press.

Labov W., Ash S., Boberg C. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin, New York: De Gruyter Mouton.

MacNeil R., Cran W. 2005. *Do you speak American? A companion to the PBS television series*. New York: Nan A. Talese/Doubleday.

McAuliffe M., Socolof M., Mihuc S., Wagner M., Sonderegger M. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. – *Proceedings of Interspeech 2017*: 498–502. [https://www.isca-speech.org/archive/interspeech_2017/mcauliffe17_interspeech.html; accessed 27 November 2021].

McLean M.P. 1952. *Good American speech*. New York: E.P. Dutton and Company.

McWhorter J. 2015. On American r-lessness. – *Language Log*, 11 June 2015. [https://languagelog.ldc.upenn.edu/nll/?p=19486; accessed: 12 November 2021].

Miller V.R. 1953. Present-day use of the broad a in Eastern Massachusetts. – *Speech Monographs* 20.4: 235–246.

Miller Center 2016 = Presidential Speeches | Miller Center. 21 November 2016. [https://millercenter.org/the-presidency/presidential-speeches; accessed: 31 March 2021].

Modiano M. 1996. *A Mid-Atlantic handbook: American and British English*. Lund: Studentlitteratur.

Newman M. 2014. *New York City English*. Berlin, New York: De Gruyter Mouton.

Nunberg G. 1980. The speech of New York City upper class. – Shopen T., Williams J.M. (eds.). *Standards and dialects in English*. Cambridge (MA): Winthrop Publishers: 150–174.

Nycz J. 2011. *Second dialect acquisition: Implications for theories of phonological representation*. [PhD Diss., New York University].

Penzl H. (1940). The vowel-phonemes in "father, man, dance" in dictionaries and New England speech. – *The Journal of English and Germanic Philology* 39.1: 13–32.
Przedlacka J., Ashby M. 2019. Comparing the Received Pronunciation of J. R. Firth and Daniel Jones: A sociophonetic perspective. – *Journal of the International Phonetic Association* 49.3: 381–400.
Queen R. 2015. *Vox popular: The surprising life of language in the media*. Chichester: John Wiley & Sons.
R Core Team 2021 = *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. [http://www.r-project.org/index.html].
Rathe A. 2020. How the Golden Age of Hollywood tried to make everyone sound filthy rich. – *Town & Country*, 3 May 2020. [https://www.townandcountrymag.com/leisure/arts-and-culture/a32292809/mid-atlantic-accent-golden-age-of-hollywood/; accessed: 18 November 2021].
Reddy S., Stanford J. 2015. A web application for automated dialect analysis. – *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*: 71–75. [https://aclanthology.org/N15–3015.pdf].
Rosenfelder I., Fruehwald J., Evanini K., Seyfarth S., Gorman K., Prichard H., Yuan J. 2014. *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2*. Zenodo. [available at: https://doi.org/10.5281/ZENODO.22281; accessed: 27 November 2021].
RStudio Team 2020 = *RStudio: Integrated development environment for R*. Boston: RStudio, PBC.
Safire W. 1987/1991. Locust Valley Lockjaw. – Safire W. (ed.). *Coming to terms*. New York: Doubleday: 160–162. [Originally published in *The New York Times Magazine*, 18 January 1987].
Skinner E.W. 1942/1990. *Speak with distinction*. [Revised with new material added by Timothy Monich and Lilene Mansell]. New York: Applause Theatre Book Publishers.
Smith B.T. 2012. Aristocratic American (Mrs. Roosevelt's accent). – *Dialect Blog*, 25 August 2012. [http://dialectblog.com/2012/08/25/aristocratic-american-accent/; accessed: 18 November 2021].
Sonderegger M. 2020. *Regression modeling for linguistic data*. Draft 0.91. [https://doi.org/10.17605/OSF.IO/PNUMG; accessed: 9 October 2021].
Taylor T. 2013. The rise and fall of Katharine Hepburn's fake accent. – *The Atlantic*, 8 August 2013. [https://www.theatlantic.com/entertainment/archive/2013/08/the-rise-and-fall-of-katharine-hepburns-fake-accent/278505/; accessed: 18 November 2021].
Thomas C.K. 1932. Jewish dialect and New York dialect. – *American Speech* 7.5: 321–326.
Tilly W. 1933. stændəd əmeɹɪkən pɹənʌnsɪeɪʃn. – *Le Maître Phonétique* 48.43: 57.
Wells J.C. 1982. *Accents of English*. [3 vols]. Cambridge: Cambridge University Press.
White E.J. 2020. *You talkin' to me? The unruly history of New York English*. New York: Oxford University Press.
Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer.
Wikipedia 2021. Mid-Atlantic accent. – *Wikipedia, the free encyclopedia*, 5 October 2021. [available at: http://en.wikipedia.org/wiki/Mid-Atlantic_accent; accessed: 12 October 2021].
Withers-Wilson N. 1993. *Vocal direction for the theatre: From script analysis to opening night*. New York: Drama Book Publishers.
Wolfram W., Schilling N. 2015. *American English: Dialects and variation*. [3rd edition]. Chichester: John Wiley & Sons.