Patterns of Vowel Epenthesis in Nigerian Spoken English

Author

Kofoworola A. Adedeji

Affiliation:

Department of English University of Lagos. Akoka, Lagos State, Nigeria

Abstract

Vowel epenthesis is a regular speech reduction phenomenon across both formal and casual speech styles among English native speakers. However, in non-native speaker environments such as Nigeria, this phonological process is often categorized as a mispronunciation, a deviant interference features symbolic of basilectal-level spoken English. This study examines the patterns and motivations of epenthesizing in both lento and allegro naturally-occurring English spoken by Nigerians, analyses utterance tokens within a theory of conversational phonology and concludes that in a similar vein with other speech reduction processes, Nigerian L2 vowel epenthesis is an attempt at achieving a more natural phonology. Phonotactics-dependent and metrical-dependent vowel epenthesis were identified as major patterns of this process, thereby providing support for the argument that this speech phenomenon is an authentic part of the Nigerian spoken English phonological system and not merely a deviant, interference-induced mispronunciation.

Keywords: vowel epenthesis, Nigerian L2 speakers, formal and casual speech

1.0 Introduction

Heterogeneity within language is a major focus of the field of variationist sociolinguistics while one of the more widely explored linguistic levels of plurality is phonetics/phonology. Studying pronunciation differences is very fruitful as marked differences in accent features are highly discernible to both the expert and the layman.

Research of accent differences across space and time as well as across stylistic boundaries have been done at both English native speaker and non-native speaker environments. Stylistic variation involves variation in the speech of individual speakers (intra-speaker variation), encompassing (a) dialects and; (b) registers. While dialect refers to differences in usage associated with a particular community of speakers, registers pertain to differences in usage with particular situations of use. An illustration of register-related variation is the assimilation form of '*going to*' pronounced as '*gonna*' as a speaker changes speech style from a formal situation such as speaking at a public seminar, to a casual one such as chatting with a close friend. In addition, the speaker may also change from one language variety to another, specifically between 'lects'. Varieties correlating with social variables i.e. social class, gender and education are lects which speakers shift between. Consequently, there are two typologies of style-shifts: shift into and out of different language varieties and shifts in usage levels associated with these varieties. Style-shifts also correlate with the axis of time and self-awareness. The length of time in which the speaker shifts between varieties may vary between momentary/short and extensive. A brief style-shifting of a few seconds involving perhaps a phrase said in the local parlance by a professor during a class and more extensive style-shifting stretching to a few hours such as features of an acrolectal standard English accent used by a corporate executive while at business meetings during the day and a 'local' Igbo- accented English used by the same executive at home with his family in the evening. Self-awareness is an indication of the amount of attention the speaker pays to their speech itself as opposed to its content i.e. how they are saying something rather than what they are saying. It refers to self-consciousness or otherwise of the use of features.

A speaker may be completely self-aware in the deliberate use of a language variety or he may be unconscious of his style-shifting between language varieties. Deliberate style-shifting hinges on the appropriation of perceived benefits by the speaker. Potential benefits generally include: appropriating a desired group identity, to express concepts that may be more easily accessible in another language variety and for interjections and sentence-fillers etc. However, these are factors related to the speech situation/audience and are therefore external to the speaker. A powerful yet intrinsic motivation

which indicates that speakers are not merely reactionary to the audience or speech situation is the need to project a personal identity while shaping their interaction with the audience. Creativity in the use of stylistic resources not only ensures that the desired identity is projected; it also ensures that the speaker is able to control interpersonal interactions during the speech event. Considering the various extrinsic and intrinsic factors involved in style-shifting, Schilling-Estes (2004:376) states that stylistic variation involves "shifts in usage levels for phonological and morphosyntactic features, typically arranged along a vernacular-standard continuum, across different speech situations, delimited either according to their relative formality or the composition of the speaker's audience.

Linguistic elements such as the phonological, morpho-syntactic, lexical and pragmatic, para-linguistic ones such as intonation, extra-linguistic ones such as body language, gesticulations, and dressing are studied in relation to style-shifting. A continuum of language varieties ranging from standard (acrolect) through to intermediate varieties (mesolect) to non-standard exist in this dimension. There are a number of social variables which may be correlated with these lects and their manipulations tend to affect speakers' performances across the phonological, morpho-syntactic, and other elements of language. Within native speaker geographical spaces, social class has been established to be a determinant of language use but level of education is the prime determinant in non-native speaker environments.

1.2 Education and Spoken English in Nigeria

The interconnectedness of educational level and English proficiency levels in Nigeria has been underscored in multiple theoretical and empirical studies. A chronological review of the correlative comparison of level of education with spoken English shows that there is some level of agreement between scholars. Also, the basis for the identification of the varieties is the triad design proposed by Bickerton (1975) as a representation of three varieties of Guyanese creole which form a continuum rather than discrete varieties: the basilect, the mesolect and the acrolect. An early stage of creole development which is characterized by the use of 'broad' forms is referred to as the basilect while the acrolect is the variety close to standard English and

spoken by the highly educated while the mesolect is the name given to varieties between the basilect and acrolect, that is, intermediate forms.

A national classification which relates specifically to English pronunciation in Nigeria is Jowitt (2019) where a two level-proposal is proposed based on the need to "distinguish between the English of relatively more educated and that of relatively less educated persons in Nigeria" (2019:43). Using the terms acrolectal and non-acrolectal, the former is described as being equivalent to 'more educated' while the latter refers to 'less educated' and also represents a fusion of basilectal and mesolectal Nigerian spoken English. Jowitt clearly adopts these terms to clear up some of the confusion implied by the use of the term 'variety' and 'level' which have been used by scholars for many decades. These terminologies suggest discreteness of features which is misleading. As it has been agreed by many scholars, the speech of any speech community or group cannot be said to be limited to a particular set of linguistic features, hence the suggestion that classifications are lectal continuums and not discrete in any way.

2.0 Speech Rate and Variation

A significant difference between speech produced in isolation and that produced in natural connected speech is change in adjoining sounds. These changes have been extensively studied in native speaker English speech and include: assimilation, elision and liaison or linking. These connected speech processes have been extensively discussed by Wells (1982) in three volumes, Stampe (1979), Shockey (2003) etc.; and explorations of these processes in non-native speaker speech include Gut (2005), and Adedeji (2015). Studies have shown that these processes are language specific and according to Shockey (2003: 20):

Recalling the sound of Indian, Caribbean, Singaporean and African English, it is easy to convince oneself that while many people from these areas are native speakers of English, they do not sound like each other nor speakers of Standard Southern British and are hence likely to have very different conventions for casual speech. (Shockey 2003:20)

The conventions and processes of the Standard British English accent have been identified by Shockey and Wells as: Reduction in unstressed syllables including absorption and suppression of the schwa, reduction of closure for obstruents, tapping, voicing and devoicing of certain segments, consonant and vowel alternation, cluster simplification, alteration of /t / and /d/, relocation of the nasals, reduction of [ð], h-dropping and palatalization.

Processes which are specific to Nigerian English include those identified by Gut (2005), Ubong & Soneye (2010), and Adedeji (2015) as various assimilatory processes i.e. progressive, regressive, metaphony, anticipatory vowel nasalization, deletion of segments, glide formation, voicing and devoicing and pre-voicing of consonants, cluster simplification, metathesis, and the insertion of epenthetic vowels. Some of these processes are summarized below as presented in Adedeji (2015) and Gut (2004).

Contact assimilation: Two types of contact assimilation processes are attested to in Nigerian spoken English:

Progressive assimilation: [opunnidɔ] /ð/ *becomes* /n/ 'open the door'

Regressive assimilation: [reppepe] /p/ becomes /d/ 'red pepper'

(ii) Distant assimilation: Two types are also identifiable.

(a) The first occurs in Igbo English in the form of vowel harmony. It is the same Igbo phonological process of harmonisation of vowels involving the open and close pharyngeal vowels which is transferred into English by Igbo English speakers as in the example of: [folo] /o/ for /ɒ/ 'follow'

(b) The second one is where the vowel in the second syllable becomes a high front vowel: rizin /i/ for /ə/ 'reason'

Bi-directional/Fusional Assimilation: This involves the creation of an output from a fusion of two segments which can be illustrated as: $S_x S_y S_z$

(where S= Segment; z = a combination of the features of x and y)
An example is: /i + l/ = j; leading to [ɔje] 'oil'.

(iv) Anticipatory vowel nasalization: [dazĩ] 'dancing'

(v) Consonant lenition or weakening: [dizemba] 'December'

(vi) Consonant fortition or strengthening: There are 5 possibilities:

pre-voicing of the consonant preceding the –*ed* morpheme in past tense forms such as: inkrizd   'increased'

voiceless consonants becoming voiced under the influence of a following voiced sound: keziz 'cases'

voicing of [ks] sequences such that [ks] [gz]: $\longrightarrow$
magzimɔm 'maximum'

devoicing of final consonants: wiθ 'with'

Devoicing of /-z/ in plurals and third person singular forms:
rods   'roads'

(vii) Homorganic assimilation: [ɔrait] 'all right'

(viii) Deletion: Three types of deletion processes occur:

(a) Aphaeresis: [jumɔ] 'humour'

(b) Syncopation: [wekɔm] 'welcome'

(c) Apocopation: [nɔma] 'normal'

3.0 A Conversational Speech Theory

Phonologists have made attempts at providing explanations of reductions in casual speech, usually from the perspective of major phonological theories. Harris (1969) is an early attempt to characterize phonological reduction as a function of rate through a hierarchical ordering of speech rates correlated with speech styles as follows.

Figure 1: Table showing speech rate across styles

Speech Rate	Speech Style	Function	Phonological Features
Largo	Very slow, deliberate & over-precise	Correcting a misunderstanding or communicating with a foreigner	No word final/initial consonant assimilation i.e. [mismo]
Andante	Moderate, natural	Delivering a lecture or teaching a class	Partial assimilation across word boundaries i.e. [mis^{z}mo]
Allegretto	Fast, casual, colloquial	Natural speech for many situations	Assimilation and nasal assimilations across word boundaries i.e. [mishmo]
Presto	Very fast	Completely unguarded	Nasal, voicing and lateral assimilation, deletions [mi mo]

Adapted from Harris (1969)
*Mismo /mismo/ is the Spanish word for *same* or *right*

Harris' position is informed by a Generative Phonology/distinctive features perspective whereby a set of basic underlying forms which contain a lot of phonetic features (i.e. nasal, labial, glottal) are subjected to phonological rules and the outcome may be insertion, deletion or neutralization; thus implying that a word or morpheme has a number of surface realisations which are predictable by speech rate and style, leading to casual forms of that word or morpheme being derived from the careful form. Accordingly, a faster speech rate means a higher probability of reductions which correspond to the four discrete styles: Largo, Andante, Allegretto and Presto.

In the same vein of Generative Phonology in which surface forms are derived from an underlying form through application of rules, Zwicky (1972) adds that the rules which generate casual speech forms are not variable but ordered in an hierachichal manner to yield particular outputs in a sequence of: (i) a process absent at one level is

7

manifested in the next (b) a process optional at one level becomes obligatory in the next (iii) a process applies to a larger class of segments, or in a wider context; and (iv) processes are reordered so as to maximalize applicability (Zwicky 1972:610). Moreover, the interdependence of rate and style is suggested as there is the 'freezing of such processes in slow speech (1972:607), leading to the existence of recognizable discrete levels such as natural careful speech and natural casual speech.

The explanation given by Shockey (2003) on the theory of casual speech is only partial in its agreement with earlier postulations in that, although there is a high correlation between fast speech rate and the presence of reduction phenomena, the differences between careful speech and casual speech are inconsistent and does not easily lend itself to prediction by rules as asserted by Generative phonologists that "given two productions of the same sequence of words, one fast, one slow, the faster one will probably show more reduction, but not always in such a way that you could regard the slower version as an input to some rules which will produce the faster version" (Shockey 2003:51).

Bolozky and Schwarzwald (1990) agree with previous theoretical claims there is a strong connection between speech rate and degree of reduction, but extend the argument by introducing the concept of attention*:*

 '...the role of increased rate of speech in reduction and assimilation may also be related to decreased attention to syllables of low prominence: since the attention focused on stressed syllables maintains their relatively long duration, speeding up the general rate of speech in the casual register will primarily be achieved by reduction of the low troughs of prominence (Bolozky and Schwarzwald (1990:23)

Also, in divergence to the generative phonology perspective, Stampe (1979) proposes the principle of Natural Phonology to account for casual speech reduction: suppression or repression of natural processes of articulation results in formal speech while lower degrees of suppression/repression produces more reduction hence casual speech. Increase and decrease in suppression will then determine different speech styles ranging from extremely formal to extremely

casual based on the quantity of reduction phenomena. It is claimed that 'processes apply in a random, nonlinear, sequential way' (1979:60). Cedargreen and Sankoff (1974) broaden this theory to include the theory of probabilities. They account for the existence or nonexistence of a condition as set of conditions as being the determinant of whether or not a rule (of reduction) will apply. Other attempts at explaining this theory include Hooper's (1976) Trace or Event principle of Hooper: disagrees with the view that rules which determine styles are ordered hierarchically, but that there is a system and direction to the application of phonological rules; the autosegmental approach (Goldsmith 1990), the metrical approach (Liberman 1975) and its extension Optimality theory (Kager 1999), Prince and Smolensky (1993) which both deal with changes to syllable boundaries in both formal and conversational speech. In its explanation of reduction, the *free-ranking* principle of OT suggests that different speech styles correspond to *constraints* which generate *outputs* ranging from formal to casual speech. The notion of statistical frequency also operates in OT. This principle suggests that phonological reduction processes can be ranked from most powerful to least powerful with frequency being the determinant of rank i.e. the more a common a phonological process is the more powerful and less violable it is.

4.0 Methodology: Data Selection

Data was collected from Tedtalk Nigeria videos and House 21 TV video comedy skits on Youtube.

Figure 2: Breakdown of Selected Data

S/N	Type	Name of Speaker	Topic	Date/Length of Video
1.	TedTalk/Formal	Sanusi Lamido	Overcoming the Fear of Vested Interest	18th August 2023/18mins.
2.	TedTalk/Formal	Aisha Babangida	Inspirational Values	16th August 2013/18mins.
3.	TedTalk/Formal	Charles Oputa (Charly	The Unstoppable You	3rd December 2015/20 mins.

		Boy)		
4.	TedTalk/Formal	Bishop Matthew Kukah	What Now?	30th August 2018/21 mins.
5.	TedTalk/Formal	Opeyemi Kareem	What's in your Bag?	1st August 2023/17 mins.
6.	Comedy/Casual	House21 TV comedy show featuring Toke Makinwa & Folarin Falana (Falz) as Mr. & Mrs. Sarumi	Marriage Therapy Sessions: One Season of 10 episodes	August 2020 - October 2020/8 mins – 12 mins per episode

Source: Author
*Total approx. time for TedTalk: 94 mins.
*Total approx. time for House21 TV comedy videos: 100 mins.

Methodology: Criteria for Selection of Speakers
Six sociolinguistic considerations were set in order to ensure that:
the data is representative of educated Nigerian speech and; (b) the data is a fairly homogenous sample of speech

The criteria are:
born in Nigeria
education is up to first degree level
primary and secondary education in Nigeria
not spent extended/significant time outside Nigeria
marked ethnic/geo-linguistic accent features were not considered
all TedTalks were delivered to a Nigerian audience

5.0 Data Presentation: Identification of Recurring Vowel-Epenthetic Features

Features which were observed to occur with some significance within both styles of speech were identified as mispronunciations. The utterances are presented below with transcriptions. FS stands for

formal speech, CS stands for casual speech, phrasal context is provided and the epenthesizing vowel is underlined.

Mispronunciations
[dɛpɒzɪt] FS

4.4 trillion naira in deposits

[sentɛd] FS
is centered to God

[stətʊs] FS
a much bigger status

[fəvɔraɪt] FS
favourite song of mine

[vaɪnjɑd] CS
the vineyard of God

[sɛkʃɔn] CS
prior to this our session

[fruidian] CS
Freudian slip

[ɛkspanʃieɪt] CS
you can expatiate

[griviɑns] CS
my grievance is

[lindʒərə] CS
even that lingerie

[kamɛra] FS
how to use a camera
[kʌmpɛni] FS
set up his company

[fɒltarin] FS
the faltering history

[rɛfɛrɛnsiz] FS
made the references

[peʃiɛns] CS
be patient with her

[mɑɡazin] CS
you are reading a magazine

6.0 Discussions Many instances of mispronunciations in the data are identifiable as the phonological process of vowel epenthesis. In a simple description, this process refers to any phonological process in which a vowel is added to an utterance. Examples from the data are: [sentɛd] scented; [stətʊs] status; [fəvɔraɪt] favourite; [kɑmɛrɑ] camera; and [kʌmpɛni] company. There are two types of vowel metathesis in the data which can be classified based on the function that the epenthetic vowel performs: Phonotactics-dependent vowel epenthesis and; Metrical-dependent vowel epenthesis A movement towards a more natural phonology of articulation ease is the usual motivation for all phonological processes including vowel epenthesis. Epenthesizing of vowels into CC codas to break up what seems to the natural intuition of the speaker to be an unwanted cluster is evident in the illustrative data below:

i.   [sentɛd] [ɪ →ɛ]
ii.   [stətʊs] [ə→ʊ]
iii.   [fəvɔraɪt] [Ø →ɔ]
iv.   [kɑmɛrɑ] [Ø→ɛ]
v.   [kʌmpɛni] [ə→ɛ]

In all the data, there is a substitution of a stressed vowel with an unstressed one or the epenthesizing of a vowel in a CC cluster confirming Hall's (2011:1611) explanation that "the function of vowel epenthesis is to repair an input that does not meet a language's structural requirements."   Determination of what constitutes a language's 'structural requirements' seems to rest with the Nigerian L2 speakers in this instance. The environment in which the epenthesis occurs conforms with the general principle of the process being very

common: (a) coda rather than onset clusters (numbers i – v); and (b) a cluster of an obstruent followed by a sonorant (numbers iii and v).

Two motivations for epenthesizing are connected to phonotactics and syllable size. Phonotactics-dependent vowel epenthesis assumes that the syllabification of hanging consonants is compulsory in order to conform to the CV open syllable structure requirements of Nigerian indigenous languages. This is significant in light of the mother tongue transfer features which are inherent in the ESL environment of the speakers. A similar explanation is the Licensing by Cue approach of Steriade (1994) which proposes that rather than the traditional view of phonological licensing which focuses on total conformity to syllable structure rules in pronunciation, the tendency is for speakers to epenthesize vowels to make consonants more easily identifiable because "a consonant that is not adjacent to a vowel is less easy to identify." (Hall 2011:1612). This provides support for the view that vowel epenthesis makes consonants more perceptible to the listener, thereby serving the two purposes of ease of articulation by the speaker and ease of utterance perception by the listener. Speech intelligibility is enhanced through this accommodation between the speaker and the listener.

The tendency of heavy syllables in NE seems to be the motivation for metrical-dependent vowel epenthesis. It is attested in the literature (Gut 2005; Adedeji 2015) of the trend towards maximum syllable weight leading to augmenting of smaller-sized syllable with vowel epenthesizing. This partially conforms with the conclusions reached by Martinez-Gil (1997:189) in the investigation of word-final epenthesis in Galician that the function of metrical-dependent vowel epenthesis is to create "a well formed bimoraic trochee at the edge of each intonational phrase". This type of poetic metrical foot consists of two moras or units of sound with the first syllable typically an unstressed syllable while the second syllable is stressed, creating a peculiar rhythmic pattern.

magazine [mɑgɑzin] ɑ+ɑ

faltering [fɒltɑrin] ɒ+ɑ

references [rɛfɛrɛnsiz] ɛ+ɛ

13

patience [peʃiɛns] e+iɛ

The vowel sequences in the data (i-iv) show the alternation of strong vowels in conformity to the NE syllable pattern of heavy/strong rather than light/weak.

Summary and Conclusions

The analysis of non-native speaker naturally occurring speech shows patterns of epenthesizing of vowels which are similar across speech styles of lento and allegro in NE speech. The main reason why the selected Nigerian speakers unconsciously interspersed coda clusters in their utterances is ease of articulation leading from an attempt at achieving a more natural phonology. Speech accommodation is also one of the inherent advantages of this CV alternation while speech intelligibility is also one of the desirable outcomes. Given these outcomes, the vowel epenthetic process should be considered an authentic feature of NE phonology.

References

Adedeji, Kofoworola. "Phonological Processes in the Nigerian Accent of English". *English Phonology in a Non-Native Speaker Environment: Theory, Analysis and Practice*. Eds. Adenike Akinjobi and Bukola Atolagbe, Lagos State University Press, 2015.

Bickerton, Derek. *Dynamics of a Creole System*. Cambridge University Press, 1975.

Bolozky, Shmuel, and Schwarzwald Rodrigue. "On Vowel Assimilation and Deletion in Casual Modern Hebrew". Biblical and Other Studies, vol. 12, 1990, pp. 23-48.

Cedergren, Henrietta, and Sankoff David. "Variable Rules: Performance in Statistical Reflection of Competence". Language, vol. 50, No. 2, 1974, pp. 333-355.
Goldsmith, John. *Autosegmental and Metrical Phonology*. Blackwell Publishing, 1990.

Gut, Stefan. *Language and Identity: The case of Nigerian English*. Peter Lang, 2004.

Gut, Stefan. *The Dynamics of Language and Culture*. Peter Lang, 2005.

Hall, Nancy. "Vowel Epenthesis". *The Blackwell Companion to Phonology*, edited by Marc van Oostendorp, Collin J. Ewen, Elizabeth Hume, and Keren Rice, John & Wiley Sons, 2011, pp. 1576-1596.

Harris, Zellig. *Methods in Structural Linguistics*. University of Chicago Press, 1969.

Hooper, Joan. *An Introduction to Natural Generative Phonology*. Academic Press, 1976.

Jowitt, David. *Nigerian English*. de Gruyter Mouton, 2019.

Kager, Rene. *Optimality Theory*. Cambridge University Press, 1999.

Liberman, Mark. *The Intonational System of English*. Cambridge University Press, 1975.

Martinez-Gil, Fernando. "Word-final Epenthesis in Galician". *Issues in the Phonology and Morphology of the Major Iberian Languages*, edited by Fernando Martinez-Gil and Alfonso Morales-Front, Georgetown University Press,1997, pp. 270-340.

Prince, Alan, and Smolensky Paul. *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science, 1993.

Schilling-Estes, Natalie. "Constructing Ethnicity and Social Identity Through Lnaguage Variation: A Sociolinguistics Approach." *The Handbook of Language Variation and Change*, edited by J. K. Chambers, Petr Trudgill, and Natalie Schilling-Estes, Blackwell Publishing, 2004, pp. 375-401.
Shockey, Linda. "Phonological Processes in English". Language, vol. 50, no. 2, 1974, pp. 323-340.

Shockey, Linda. *Sound Patterns of Spoken English.* Blackwell Publishing, 2003.

Stampe, David. *A Dissertation on Natural Phonology*. Garland Publishing, 1979.

Steriade, Donca. "Licensing by cue". Unpublished ms., University of California, Los Angeles, 1994.

Ubong, Ufot, and Femi Soneye. "Language Variation and Change in Nigerian English." Journal of Nigerian English Studies, vol. 5, no. 2, 2010, pp. 45-60.

Wells, John. *Accents of English*. Cambridge University Press, 1982.
Zwicky, Arnold. "Some Controversies in Syntax". Linguistic Inquiry, vol. 3, no. 3, 1972, pp. 397-415.