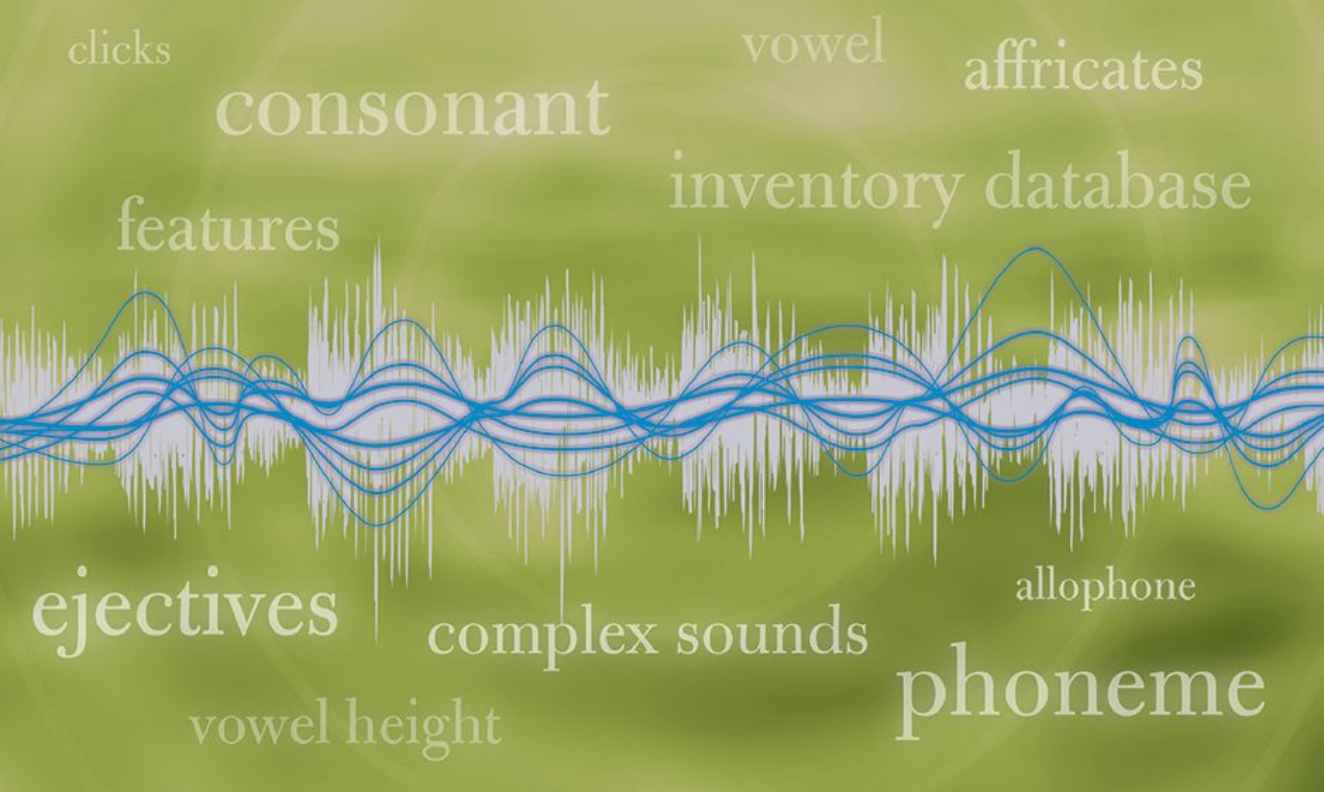
OXFORD
LINGUISTICS

A THEORY OF PHONOLOGICAL FEATURES

SAN DUANMU



A Theory of Phonological Features

A Theory of Phonological Features

SAN DUANMU

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2016

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2015948898

ISBN 978–0–19–966496–2

Printed in Great Britain by
Clays Ltd, St Ives plc

Contents

Preface

The goal of this study is to determine a feature system that is minimally sufficient to distinguish all consonants and vowels in the world's languages. Evidence is drawn from two databases of transcribed sound inventories, UPSID (451 inventories) and P-base (628 inventories).

Feature systems have been offered before that use data from many languages. For example, Trubetzkoy (1939) cited some 200 languages, Jakobson et al. (1952) cited nearly 70 languages, and Maddieson (1984) used a database of 318 languages. However, a fundamental methodological question remains to be addressed: Let X be a sound from one language and Y a sound from another language. How do we decide whether X and Y should be treated as the same sound or different sounds? It is well known that, when X and Y are transcribed with the same phonetic symbol, there is no guarantee that they are phonetically the same. Similarly, when X and Y are transcribed with different phonetic symbols, there is no guarantee that they must be different sounds. Moreover, even if there is phonetic evidence that X and Y are notably different, there is no guarantee that they cannot be treated as the same sound in a language. Without a proper answer to the methodological question, generalizations from cross-language comparisons are open to challenges.

In this study, I offer an answer, using the notion of contrast: X and Y are treated as different sounds if and only if they contrast in some language (i.e. distinguish words in that language). For example, [l] and [r] contrast in English (as in *lice* vs. *rice*); therefore, we must represent them with different transcriptions and different features. In addition, when [l] and [r] occur in a language where they do not distinguish words, such as Japanese, we can distinguish them, too (as "allophones"). On the other hand, if X and Y never contrast in any language, there is no need to distinguish them, not even as allophones, even if they have an observable phonetic difference. For example, Ladefoged (1992) observes a difference between [θ] used by English speakers in California, whose tongue tip is visible, and [θ] used by those in southern England, whose tongue tip is not visible; but if the two kinds of [θ] never contrast in any language, there is no need to distinguish them, either in transcription or in features. Similarly, Disner (1983) observes that the [i] in German has a slightly higher tongue position than that in Norwegian. But again, if the two kinds of [i] never contrast in any language, there is no need to distinguish them either. Non-contrastive differences are not left aside but will be addressed as well, and explanations will be suggested without assuming feature differences.

I make no prior assumption as to what the system should look like, such as whether features should be binary or innate. Instead, the proposed method is explicit

and theory-neutral. The work can be laborious to carry out, though: It requires repeated searches through sound inventories of all available languages in order to find out whether a difference in phonetics or in transcription is ever contrastive in any language. In addition, to guard against errors in original sources or in the databases, we shall re-examine all languages that seem to exhibit an unusual contrast.

The resulting feature system is surprisingly simple: Fewer features are needed than previously proposed, and for each feature, a two-way contrast is sufficient. For example, it is found that, if we exclude other factors, such as vowel length and tongue root movement, a two-way contrast in the backness of the tongue is sufficient; the same is true for the height of the tongue. This result is quite unexpected, because even the most parsimonious feature theories, such as Jakobson et al. (1952) and Chomsky and Halle (1968), assume three degrees of tongue height, and many assume three degrees of tongue backness as well. Nevertheless, our conclusion is reliable, in that the notion of contrast is uncontroversial, the proposed procedure is explicit, and the result is repeatable.

Representing contrast is not the only purpose for which features are proposed. In particular, features have been proposed to describe how sounds are made (articulatory features) and how sounds form classes in the phonology of a language (class-based features) as well. This study focuses on contrast-based features for two reasons. First, contrast lies at the core of phonology. Second, contrast-based data are the least controversial and are large in quantity and high in quality. I shall attempt to interpret contrast-based features as articulatory gestures, too, but I shall say little of class-based features. It has been proposed that contrast-based features, articulatory features, and class-based features should correspond to each other (Halle 1995), but that remains an ideal. Before the ideal is confirmed, differences among the three feature systems, once highlighted, provide fertile grounds for further research.

Acknowledgments

The theory of phonological features is one of the first topics a student of linguistics comes across, yet one soon realizes that there are many unresolved questions. Naturally, many phonologists have searched for answers, and some have devoted most of their careers to it. If some progress is made in this book, it is very much the result of collective effort. I am glad to be part of it, and have many people to thank.

Starting from the beginning, I would like to thank those whose classes introduced me to the field of phonetics and phonology: David Crystal, Francois Dell, William Hardcastle, Jim Harris, Michael Kenstowicz, Dennis Klatt, Donca Steriade, Ken Stevens, Peter Trudgill, Moira Yip, Victor Zue, and above all Morris Halle. I was fortunate to have you as my teachers.

I would also like to thank my colleagues at the University of Michigan, in Linguistics and beyond. It is a privilege to share a great research and teaching environment with you.

Next, I would like to thank other phonologists, broadly defined. Some I have interacted with directly and some I got to know through their works. You have been wonderful companions in a journey of exploration for a better understanding of the sound patterns of language.

Just as much, I would like to thank my students over the years, whose open-minded questions often helped me clarify my explanations, revise my views, and see new questions that had been hidden by familiar jargons.

Finally, I would like to thank Michael Kenstowicz, Keren Rice, and three anonymous reviewers for their valuable comments on the book proposal, Bert Vaux and Jeff Mielke for their very careful comments on the draft copies of the book, other colleagues for their comments on various parts of the book (Steve Abney, George Allen, Hans Basbøll, Pam Beddor, Iris Berent, Arthur Brakel, Charles Cairns, Nick Clements, Andries Coetzee, Jennifer Cole, Stuart Davis, Ik-sang Eom, Shengli Feng, Mark Hale, Bruce Hayes, Jeff Heath, Yuchau Hsiao, Fang Hu, José Hualde, Di Jiang, Ping Jiang, Pat Keating, Hyoyoung Kim, Andrew Larson, Qiuwu Ma, Bruce Mannheim, James Myer, Wuyun Pan, Eric Raimy, Rebecca Sestili, Jason Shaw, Yaoyun Shi, Ken Stevens, Will Styler, Sally Thomason, Jindra Toman, Hongjun Wang, Hongming Zhang, Jie Zhang, Jun Zhang, Jianing Zhu, and Xiaonong Zhu—and this surely is an incomplete list, I am afraid), and the staff at (or working for) Oxford University Press, in particular Sarah Barrett, John Davey, Lisa Eaton, Julia Steer, Vicki Sunter, and Andrew Woodard, for their professionalism at every stage of the publication process. You have helped make this book far better than what was originally conceived.

For copyright permissions I am grateful to the following:

Cambridge University Press for Fig. 1.2 (Karl Zimmer and Orhan Orgun, "Turkish," *Journal of the International Phonetic Association* 22(1–2) (1992), p. 44), also reproduced as Fig. 4.1.

Sandra Disner for Fig. 1.3 (Sandra Ferrari Disner, "Vowel quality: the relation between universal and language-specific factors," doctoral dissertation, UCLA, 1983; distributed as UCLA Working Papers in Phonetics 58, p. 67, Fig. 3.24).

AIP Publishing LLC for Fig. 2.2 (Gordon E. Peterson and Harold L. Barney. 1952. "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America* 24(2) (1952), p. 182, Fig. 8).

Abbreviations and terms used

Commonly used abbreviations

*	ill-formed
[ ]	a feature, often one with binary values, such as [round]; a phonetic transcription
ATR	advanced tongue root
C	consonant
CC	unit of two consonants; geminate consonant
CG	unit made of a consonant and a glide, such as [kw] or [k^w]
CL	unit made of a consonant and a liquid ([l] or [r]), such as [pl]
CV tier	tier representing consonant and vowel positions
G	glide
H	high tone, which is the same as the feature value [+H]
HL	falling tone, which is a combination of H and L
IPA	International Phonetic Association; International Phonetic Alphabet
L	low tone, which is the same as the feature value [–H]
LH	rising tone, which is a combination of L and H
N	nasal consonant
NC	unit made of N and C; pre-nasalized consonant
P-base	database of phoneme inventories and sound classes (Mielke 2004–7)
UPSID	UCLA Phonology Inventory Database (Maddieson and Precoda 1990)
V	vowel
VX	rime made of a vowel and another sound, such as VV [ai] or VC [an]
X	timing slot; unit of time
X tier	tier representing time units

Active Articulators

Articulator	Common name	Other terms
Lips	Lips (or lower lip)	Labial; Lower lip
Tip	Tip of the tongue	Coronal; Tongue blade
Body	Body of the tongue	Dorsal
Velum	Velum	Soft palate
Root	Root of the tongue	Radical
Glottis	Glottis	Vocal cords; Vocal folds; Laryngeal
Larynx	Larynx	

1

Introduction

A fundamental assumption in phonology is that, at some level of abstraction, speech is made of a sequence of sounds, or consonants and vowels. Under this assumption, numerous studies of a language begin by listing its consonants and vowels. The goal of this study is to determine a minimally sufficient feature system that can distinguish all consonants and vowels in the world's languages. Before I introduce the methodology in Chapter 2, it is necessary to address some preliminary questions: What are speech sounds? What are features? Can we compare sounds and features across languages? Do we have adequate data for the task?

1.1 Sounds and time

The discovery that words can be decomposed into sounds has made it possible to create writing systems in many languages, which in turn has had a profound impact on human civilization. Indeed, according to Goldsmith (2011: 28), phonemic analysis (the technique for figuring out the consonants and vowels of a language) remains the greatest achievement in phonology. Nevertheless, a number of questions remain.

1.1.1 Segmentation of speech

When we segment speech into consonants and vowels, we encounter two problems. First, there are prosodic properties, such as tones, that seem to be independent of, or attached to entities larger than, consonants and vowels. Second, phonetically, the boundaries between sounds are not always clear, because properties of one sound often spill into another (and vice versa)—a process called "coarticulation" in phonetics and "feature spreading" in phonology. For example, in *pan*, nasalization (a property of [n]) starts during [æ], not after it. Similarly, the tone of a syllable can extend to another, or shift from one to another. In response, Goldsmith (1976) proposes that speech is made of multiple tiers of features, each tier being autonomous. Features on the same tier have their own temporal sequence, but there is no common temporal sequence across all layers. The conclusion is that speech cannot be segmented into consonants and vowels. A similar view is expressed by Firth (1957)

A Theory of Phonological Features. First edition. San Duanmu.

and more recently by Fowler (2012), who suggests that we should focus on individual articulatory gestures instead of sounds or segments.

However, some well-known facts will be hard to account for if speech is not made of a sequence of sounds (after we set aside prosodic properties such as tone and syllable structure). For example, given two sounds A and B, languages can make a contrast between AB and BA, such as *tax* vs. *task*, *cats* vs. *cast*, *tea* vs. *eat*, and *map* vs. *amp*. In addition, it is possible to spell or transcribe speech with a sequence of letters or phonetic symbols, regardless of the language—a fact that would be quite unusual if speech is not made of a sequence of sounds. Moreover, language games can manipulate (i.e. move, insert, delete, or change) sound-sized units, and such games are found not only in languages that are written alphabetically, such as Pig Latin in English, but also in those that are not, such as Chinese (e.g. Chao 1931; Yip 1982; Bao 1990b). Finally, there is no evidence that it is easier to account for feature spreading (or coarticulation) if we reject consonants and vowels. Indeed, even within the multi-tiered approach to phonology, a special tier has been proposed that corresponds to traditional notions of consonants and vowels, such as the CV tier of McCarthy (1979) and Clements and Keyser (1983), or the X tier of Pulleyblank (1983) and Levin (1985). Therefore, I shall continue to assume that speech is, at some level of abstraction, made of a sequence of sounds.

1.1.2 Granularity of segmentation

A second problem in decomposing words into sounds is the granularity of segmentation: there is no agreement on how large (or small) a sound should be. For example, is the affricate [ts] one sound or two? Is the diphthong [ai] one sound or two? Is the triphthong [uai] one sound or three? Is [ʘx] (in the African language !Xóõ) one sound or two? Should the decisions be made on a language-specific basis? For example, can [ai] be one sound in some languages and two in others, even if it is phonetically the same?

Chao (1934) argues that phonemic analysis is inherently ambiguous and a unique solution is rarely possibly. The ambiguity has made some linguists doubt the reality of consonants and vowels. For example, after years of working on consonants and vowels, the prominent linguist Ladefoged (2001: 170) remarks that they are probably "scientific imaginations" after all.

Nevertheless, many ambiguities are resolvable if additional evidence is taken into consideration. For example, consider the syllable onset in Standard Chinese. If we exclude glides, only the following items are found [p p^h t t^h k k^h ts ts^h $tʂ$ $tʂ^h$ $tɕ$ $tɕ^h$ m n f s ʂ ɕ x ɻ]. If the affricates [ts ts^h $tʂ$ $tʂ^h$ $tɕ$ $tɕ^h$] are single consonants, we see a generalization: The Chinese onset allows just one consonant. If affricates are clusters of two sounds each, the generalization is lost. In addition, we face a new question whose answer is not so obvious: Why are some consonant clusters allowed while others not?

However, additional evidence is often ignored. As an example, consider two analyses of Chinese. You et al. (1980) propose that Chinese should not be segmented into consonants and vowels. Instead, we can treat each rime as a single sound, such as [au], [ai], [an], and [aŋ]. The advantage, they argue, is that we do not need to account for contextual variations of vowels (called "allophonic variation" in phonemic analysis), such as the variation of [a] in [au], [ai], [an], and [aŋ]. However, they fail to account for the fact that diphthongs and VC rimes are found in full syllables only, whereas rimes of unstressed syllables are half as long and have a simple vowel only (Lin and Yan 1988). If the difference between full and unstressed syllables is taken into consideration, it is clear that the former have two rime positions and the latter just one, so that [au], [ai], [an], and [aŋ] should all be split into two sounds each.

Next we consider vowels in Standard Chinese. According to Lee and Zee (2003), there are six monophthongs (such as [a] in [ma] 'mother' and [an] 'peace'), eleven diphthongs (such as [ai] in [mai] 'sell' and [ia] in [ia] 'duck' and [ian] 'smoke'), and four triphthongs (such as [uai] in [xuai] 'bad'). However, the analysis becomes problematic when we consider evidence from syllable structure. First, [mai] 'sell' and [xuai] 'bad' form a riming pair, as do [an] 'peace' and [ian] 'smoke'. This means that [uai] should be divided into [u] and [ai], because [ai] is a unit for riming. Similarly, [ian] should be divided into [i] and [an], where [an] is a unit for riming. Second, diphthongs like [ai] and [au] cannot be followed by a consonant, such as *[ain] or *[aun], whereas simple vowels can, such as [in] and [an]. This means that a diphthong is equal to two sounds. Thus, evidence from syllable structure suggests that both diphthongs and triphthongs should be decomposed into simple vowels. The decomposition yields better phonemic economy, too. For example, without decomposing diphthongs and triphthongs, there are twenty-one vowels in Standard Chinese (Lee and Zee 2003), whereas with the decomposition there are at most six, including a retroflex vowel (Duanmu 2007: 41).

Even if additional evidence is used, its interpretation is not always obvious. Let us consider two examples. We have just seen that diphthongs can be treated as two sounds when they are long. However, Cairns (p.c. 2013) suggests that some diphthongs are short. The example is New Yorkers' pronunciation of the vowel in *bath* and *cab* as [æᵊ]. Cairns considers the vowel to be a short diphthong, because [æ] is often treated as a "lax" vowel, and lax vowels are usually short in English. The question is whether this vowel is indeed short. Phonetically, [æ] is clearly a long vowel (Peterson and Lehiste 1960: 701). Phonologically, the New York [æ] undergoes "tensing" in such an environment (Benua 1995), and tense vowels in English are phonologically long, in part because they attract stress (Halle and Vergnaud 1987; Hayes 1995). Thus, there is good evidence that the New York [æᵊ] is not a short diphthong but a long one, which can be decomposed into two sounds.

Another reported example of short diphthongs is found in Gussmann's (2002) analysis of Icelandic. The argument is that the maximal rime size in Icelandic is VX (i.e. VV or VC); therefore, in a VVC rime, VV must be a short diphthong. It is worth noting that in VVC rimes, the C is typically a nasal. Such a case is not new. For example, Borowsky (1986) observes that in non-final English syllables the rime size is mostly limited to VX, although some exceptions are found, such as *pumpkin*, whose first rime is VCC, and *fountain*, whose first rime is VVC. In such rimes, the vowel is followed by a nasal, which can form a nasalized vowel (Duanmu 2008). Thus, the first rime in *pumpkin* is [ʌ̃p] and that in *fountain* is [aũ], an analysis independently proposed before (Malécot 1960; Bailey 1978; Fujimura 1979; Cohn 1993).

The examples show that different decisions on the granularity of segmentation can yield different consonants and vowels. When we examine databases of sound inventories, therefore, we should not take reported inventories at face value. Instead, we should be aware of the range of ambiguities and alternative interpretations. In addition, we should be cautious in drawing certain kinds of generalization, such as the number of consonants and vowels in a language, the average sizes of phoneme inventories across languages, or the total number of distinct sounds in all languages. Moreover, the size of a sound affects its feature analysis, too. For example, if [ʘx] is a single sound in !Xóõ, its feature analysis would be quite complicated. On the other hand, in a "cluster analysis" (Traill 1985: 208–11), [ʘx] is made of two sounds, a bilabial click [ʘ] and a velar fricative [x], and their feature analyses would be much simpler.

The granularity problem is related to what Chao (1934: 371) calls "under-analysis" and "over-analysis." In under-analysis, one treats what are "recognizably compound sounds" as a single sound, such as treating the affricate [ts] or [ts^{h}] as a single consonant. In over-analysis, one treats what is "one homogeneous sound" as two sounds, such as treating the American English vowel [ɚ] as [ə] plus [r]. But the terms "under-analysis" and "over-analysis" imply that (i) there is a "proper" analysis and (ii) we know what it is. In other words, Chao seems to assume that we already know what a sound is. The assumption is not obvious, as we shall see next.

1.1.3 Defining sounds by time

Given the assumption that speech can be segmented into a sequence of sounds, the simplest definition of sounds is as in (1), with examples in (2), where X is a time unit (McCarthy 1979; Pulleyblank 1983; Clements and Keyser 1983; Levin 1985). I use "features" as a cover term for phonetic properties (such as articulatory gestures), to be elaborated on later.

(1) Sounds defined by time
A sound is a set of compatible feature values in one time unit.

(2) Sample representations of sounds

Single sounds		*Two-sound units*		
Simple	*Complex*	*Diphthong*	*Long vowel*	
X	X	XX	XX	Timing tier
|	/\	| |	\/	
k	kp	a i	i	Gestures
[k]	[k͡p]	[ai]	[i:]	Transcription

The notion of time is both phonetic and phonological. Phonetically, one sound is shorter than two sounds of comparable nature and in comparable context. For example, a short vowel (which takes one time unit) ought to be shorter than a long vowel or a diphthong (which take two time units each). Phonologically, a long vowel or a diphthong occupies two rime positions, similar to a short vowel plus a consonant. This can be demonstrated in some well-studied languages, such as English and Chinese. In addition, a long sound often shows different phonological behavior than a short one. For example, a long vowel tends to attract stress but a short one does not (Prince 1990; Hayes 1995). Similarly, a vowel may be lengthened when it carries a contour tone but not when it carries a level tone (Ward 1944).

Phonetic duration can be influenced by various factors though, in particular the phonotactics of a language. For example, an English vowel is shorter before a voiceless consonant than before a voiced one (House and Fairbanks 1953; Peterson and Lehiste 1960); thus, [ɪ] is shorter in *fit* than in *fig*. In addition, high vowels are shorter than non-high vowels. More striking cases have also been reported, in which the duration of a two-consonant cluster is similar to that of a single consonant. For example, Browman and Goldstein (1986: 233) report that in English, the temporal gesture of labial closure is the same for [p b m mp mb] (as in *capper*, *cabber*, *cammer*, *camper*, and *camber*), "regardless of whether the consonantal portion is described as a single consonant (/b/, /p/ or /m/) or as a consonant cluster (/mp/ or /mb/)." Similarly, Maddieson (1989) reports that, while prenasalized stops are found to be longer than a single consonant in Luganda (Herbert 1975; 1986), they have similar timing patterns to those of single consonants in Fijian. Such cases can sometimes be explained by the phonology of the given language though. For example, it can be argued that a prenasalized stop in Luganda is a true consonant cluster (Herbert 1975; 1986). In contrast, the Fijian consonant inventory has [p t k], [m n ŋ], and [^{m}b ^{n}d ^{ŋ}g], but no [b d g] (Dixon 1988: 13); therefore, it is possible that the Fijian [^{m}b ^{n}d ^{ŋ}g] are in fact single consonants [b d g] (Herbert 1975; 1986). In English, vowels are nasalized before a nasal; therefore, the nasal closure itself becomes redundant if the following consonant can indicate its place of articulation. In other words, the closure pattern observed by Browman and Goldstein (1986) can be described by a rule [VNC] → [ṼC], where [NC] is a homorganic cluster.