How Abstract is Phonetics?

Osamu Fujimura

Department of Speech and Hearing, The Ohio State University
osamu.fujimura@gmail.com

Abstract

Assuming a generative principle of description for speech utterances, in particular a syllable-based phonological representation and the C/D model of phonetic implementation of utterances, the basic question is discussed: how abstract phonetic representations as an optimal symbolic description of speech utterances should be. Several examples from a variety of languages in the world, including Kaingang (a Je language spoken by some indigenous Brazilians), which requires a phonological specification of oral syllables, as opposed to the default nasal syllables, are discussed. The advantage of using syllabic features and their implementations for concise phonetic description for particular languages is advocated.

Index Terms: syllable features, C/D model

1. Introduction

The mapping from phonological properties such as (distinctive) features to signal properties such as position of an articulator is very often not transparent but quite opaque. Not only the mapping is not linear or even monotonic, but also the mapping relation is not always one-to-one. This paper examines such opaqueness. We may need alternative phonological variables as the input or alternative phonetic variables or signal parameters as the output, or both, in order to improve the relationship. While such an improvement is strongly desirable for applications as well as for basic understanding of human behaviors, substantially more complex phonetic implementation algorithms than usually assumed are necessary for precisely describing properties of language and speech.

2. Phonology and Phonetics

Paul Kiparsky [1] asked this question: how abstract is phonology? The answer must depend on how powerful phonetics can be, assuming that there is a mapping process, i.e., phonetic implementation of phonological implementation as a generative description [2]. Current approaches in phonology use what is called underspecification [3] in phonological representation of linguistic forms to different extents of abstraction. The larger is the domain of oppositional patterns considered, the more underspecified phonological representations can be. Development of very large databases makes this exploration feasible, given current computational power.

To transform any such phonological representation into concrete speech signals, we need phonetic implementation algorithms for the given language. In addition, for speech utterances to be represented, we need to add substantially more specification about speaker conditions and utterance situations. In the C/D model [4], this added utterance-specific information is represented by setting (symbolic or numeric) system parameter values.

Phonetics is a mixture of symbolic and numeric information. Generative phonetics (phonetic implementation) has to work with the phonological representation of linguistic forms in a hierarchical phrasal representation as its input, and multidimensional variables as the output. Basically phonetic signals are continuous variables in multiple dimensions, but it contains symbolic control variables as parameters for the system operation.

To claim that a phonetics is totally concrete and objective is a falacy. A human adult, including trained phoneticians, cannot be free from biases from his/her personal linguistic experience in the native language, or at best familiar languages.

2.1. Prosodic Effects on Vowel Articulation

It may be said that depending on how abstract the phonological representation, and how powerful the phonetic implementation system, the number of distinct vowels in the paradigm varies. In English, the concept of schwa as the reduced vowel is widely used, but how many different schwas exist at what level of abstraction is still an interesting question.

In English, among syllables that are not reduced, i.e., with full vowels as their nuclei specified with vocalic features such as front-back and low-high, the articulation varies reflecting the degree of syllable stress. This degree is represented in the C/D model by a continuously variable scalar quantity called syllable magnitude, and it is quantitatively represented by the height of the syllable pulse in the CD diagram (see below for examples). Syllables in rhythmically weaker position in a phrase is phonetically, i.e. numerically, reduced [5].

Fig. 1: Effect of contrastive emphasis on vowel articulation

In Fig. 1, two utterances of the sentence ‘It’s six five seven America Street’ are compared. The ordinate represents the height of a pellet attached on the tongue blade. The upper time function represents an utterance with a contrastive emphasis on ‘six’ and the lower on ‘America’. The depths of valleys representing the vowels in corresponding syllables reveal substantial differences in tongue surface height for the same vowel in the same segmental context, solely depending on the stress that is attached to the pertinent word. The phrase-initial schwa of ‘America’ has substantially larger mouth opening than for the mid front vowel of the second syllable, only when the word is emphasized [4, 6]. Presumably, the...
phrase-initial effect is stronger than the lexical stress and syllable reduction effects in this example.

3. The C/D Model

The C/D model is an effective description of phonetic implementation as a system superimposing different phonetic variables, as an approximately linear system up to the very last stage of variable conversion called signal generator. This last process transforms articulatory control variables to physico-acoustic variables through a strongly nonlinear process, including collision between the crucial active articulator, such as the tongue tip for obstruents, and the passive articulator, such as the hard palate. The CD diagram displays semi-quantitatively, like a cartoon, the articulatory control actions for each control variable independently from others. This is possible since the strong interactions among physical (from muscular activity to acoustic) variables in movement pertain to the signal generation component of the model [5, 6], after the displayed level of the CD diagram for presentation and interpretation of the complex phenomena.

In particular, the concept of threshold values of control variables as time functions plays a critical role. The position change for each active (i.e., moving) articulator is displayed in CD diagrams as if the curve for an impulse response function (IRF) for a consonantal gesture represented the center of gravity of the pertinent articulatory organ rather than its periphery, which becomes in contact with the passive articulator and deforms nonlinearly, strongly interacting with other articulatory variables according to anatomical situations.

Fig. 2: A CD diagram for ‘strict’ in isolation

An articulatory closure of the vocal tract by the apex of the tongue, for example, is formed, when a threshold value for height of the tongue tip is exceeded in terms of position of the center of gravity. The center of gravity of the pertinent part of the tongue continues to move after the peripheral point has come to stoppage due to collision. The tongue tip control variable is, in effect, a representation of such smoothly moving anatomical subsystem, while the threshold value, indicated in the CD diagram by a horizontal cursor line shows when the periphery closes up the vocal tract by collision of an articulatory stop, for example. Such a motion of the physical organ beyond the collision threshold does not have any significant acoustic consequences for usual discussion. But it does have significance in determining the force of articulatory contact during closure and in the shape of the vocal tract in detail that is hidden behind the closure. Fig. 3 displays a CD diagram. Fig. 2 shows a CD diagram for the word ‘strict’, spoken in isolation. Fig. 3 is for ‘splatter’ in American English, which involves an ambisyllabic /t/ between the syllables [8, 9].

Fig. 3: A CD diagram for ‘splatter’ //SplaTR// (English)

In the case of the laryngeal adduction/abduction, which controls voice onset and offset, there is a threshold value of the glottal aperture that determines the moment of the onset and offset, given a subglottal pressure value set, which is usually assumed to be held approximately constant during a voiced period. The aperture threshold value is different for voicing onset and offset, due to hysteresis of the vocal fold tissue (see Fig. 3., for the word ‘splatter’ spoken in isolation).

In Figs. 2 and 3, the vertical cursor line is drawn at the moment when the control variable crosses such a threshold line, indicating the moment of discontinuous event such as articulatory release, voice onset, etc. The nonlinear nature of the physical speech production system is crucial for determining the mapping between articulation/phonation and acoustic (spectrographic) events, but considerations of control functions are approximately free from such complications due to severe system nonlinearity. Even though the computation of the signal generator component of the C/D model is exceedingly complex and computationally heavy, the process is purely physical and in principle it is solvable [10].

4. Minimizing Representation

4.1. English ‘strict’ //Strik.t// as an example

Consider the word ‘strict’ in English as an example. The traditional representation by phonemes is /strikt/, where each alphabetic letter corresponds to either vowel or consonant. The letter /t/ here represents a lax, high, front vowel. Phonemes comprise a linear string, each of them being treated as an independent comparable concatenative unit. For the string of phonemes to represent a syllable, the minimal form for utterance, we cannot exchange the position of any of the letters with another within the same string in this case. There is a large amount of other redundancy in this representation for an English syllable. Given the condition it be an English syllable, for example, the first unit has to be /s/, not any other phoneme. The last phoneme must be /t/, nothing else. The second phoneme must be /p/ or /t/ or /k/, regardless if it is a real word, and the next phoneme could be either /l/ or /r/, nothing else.
A more rational phonological representation with much less redundancy is a representation of the syllable as an unordered set of phonological features: \{spirantized, apical\} for onset, \{dorsal\} for coda \{stop\} being implied and \{tense\} being an unmarked default choice, \{stop\} for s-fix, and \{high, front\} \{lax\} being implied by not having a glide feature specified for coda) for the nucleus. In English, \{stop, fricative, spirantized\} are comparable obstrener features for choice in onset, coda, or s-fix, \{nasal, lateral, rhotic, glide\} are another set for a choice of sonorant feature. An s-fix is separable from the rest \{syllable core\}, and the only choice of its feature is \{stop\}, \{fricative\}, \{spirantized\} or \{marginally, always as a morpheme with a specific meaning\} \{interdental\}.

The feature \{spirantized\} is a complex manner feature, which is implemented as a preceding unvoiced frication gesture combined with a place specification (for \{p/, t/, k/\}) as an immediately succeeding stop \{unvoiced oral in coda and unvoiced oral or nasal in onset\}. Interestingly, this temporal sequencing of frication \{preceding\} and oral closure \{succeeding\} gestures is maintained for the \{spirantized\} feature when it is specified for the syllable onset. Note in contrast, in \{pri/\} vs. \{pIr/\}, the sequence of the frication and lateral phonemes \{//u/\} must be inverted between onset and coda \{sonority cycle constraint [11]\}. The gesture complex for the feature \{spirantized\} is more solidly integrated in its implemented gesture organization with fixed temporal order regardless its position in the syllable, whereas the combination of lateral and fricative consonantal gestures are segmentally separate, obeying the sonority cycle principle.

In automatic speech recognition, for example, such constraints corresponding to the high amount of redundancy should be exploited, if the object to be identified has to be a sequence of phonemic segments. As a linguistic representation, the phonemic representation is simply not effective.

5. Syllable Feature \{oral\} in Kaingang

Kaingang is a Je language currently spoken by indigenous inhabitants in Brazil. There are different dialects; what I try to analyze here, tentatively with a very limited knowledge of the language and no experience of speaking with native speakers but solely based on reports, in publication and personal communication, of an expert of this language at Campinas University, near Sao Paolo: Wilmar da Rocha d’Angelis [12, personal communication]. His analysis seems to show a very deep insight into the nature of this language, in particular proposing a marked feature of orality of vowels, while nasality is unmarked for Kaingang syllables. Our analysis suggested here is a reinterpretation of his description, based on the C/D model as the available phonetic implementation process, using a method of phonetic representation of phonetic phrases.

5.1. Kaingang syllable \#/dUg#/ 

Let us start with an apparently typical, though strange for most speakers of major European languages, phonological form of a word, which d’Angelis phonologically represents as \#/nu\#, with the vowel specified phonologically as \{oral\}. His phonetic (IPA) transcription is \{ndug\}. We transcribe it here by the phonemoidal transcription (see Fujimura and Erickson [1997]) as \#/dUg#/., where the capital letter for the vowel identifies the entire syllable to be specified as \{oral\}. In d’Angelis’ segmental analysis, the vowel is phonologically marked as \{oral\} when it is not transcribed with explicit tildes for nasal, which is a default situation in Kaingang. In our \{syllable\} based phonemoidal transcription, a syllable marked for \{oral\} is represented by a capital letter for this language.

Our phonological representation by syllable features (superscripts O/C for onset/coda) and IPA transcription are: \{apical\}, \{lax\}, \{dorsal\}, \{nasal\}, \{oral\}, \{ndug\}.

Unlike most other languages, Kaingang uses the phonetic status \{nasal\} as the default syllable status. According to D’Angelis, the feature \{oral\} marks the vowel /\u/ phonologically; it oralizes also inside parts of the nasal stops of onset and coda. The syllable is underlyingly specified as /\nu/ to produce /ndug/ phonetically.

In Fig. 4, this situation is reinterpreted as a CD diagram, where the base function of the syllable is implemented abstractly \{within phonetics\} by a rectangular time function, along with the onset and coda gestures represented by the IRFs superimposed near the left and right margins, respectively. The IRFs are triggered by onset and coda pulses, respectively, erected at the left and right edges of the syllable domain \{see the triangle at top\}. One integral \{closing and releasing\} gesture of the tongue tip is observed acoustically as a sequene of /\n/ in the onset, because it is divided into successive phonetic segments at the time point when the oral nucleus gesture \{the tinted area\} predominates over the velum lowering gesture for /\n/ in onset. The \{ghi\} sequence of coda is similarly explained. In Kaingang, if I understand it correctly, lax \{voiced\} stops, /\b/, /\d/, /\g/ occur only in a syllable marked as oral, though they become partially nasal, while simple nasal stops, /\m/, /\n/, /\j/, appear only in the context of \{default\} nasal syllables. Tense obstruents can occur in both oral and nasal syllables.

![Image](image.png)

Fig. 4: A CD diagram for \#/dUg#/ for a word ‘stomach’.

Using a syllable-based feature analysis with the C/D model, the unique phonetic patterns in Kaingang, as observed above, can be interpreted simply and naturally, assuming:

1. Syllables as a whole are phonetically nasal by default, except for inherent orality of margin \{onset or coda\} gestures specified for an obstruct feature \{stop or fricative\}. Obstruent gestures, unless marked as \{lax\} are, by default, tense and voiceless and free from any nasalization, even when the syllable is not specified \{oral\}.

2. When the feature \{lax\} is assigned to onset or coda of a syllable, a voiced stop gesture by the crucial articulator is implemented, according to the concomitant place feature specified. In this case, the stop closure is nasal at the outside edge of the syllable next to the syllable boundary, producing what may be interpreted by the traditional phonemic view as a homorganic segmental sequence like //\n/\ or /\ndi/\ (the sharp here stands for syllable boundary).
3. Syllable boundaries are inherently |nasal| (in conformity with the natural breathing requirement). The partial nasality of stop closure, when the feature {lax} is specified, can thus be interpreted as a coarticulatory nasalization of the stop closure due to the nasality of the adjacent boundary. The lax feature makes the velum control particularly susceptible to such contextual influence.

4. The phonological feature {oral}, when applicable, is assigned only for the entire syllable, according to our view.

5. If {oral} is not specified, the entire syllable is produced with the phonetic status |nasal|. An obstruent is by default unvoiced, unless specified by the feature {lax}.

6. The feature {stop} implies a nasal stop if and only if the syllable is not specified for {oral} (or {lax} by implication).

5.2. //big// /miɡʃi/ (/cat/)

Fig. 5 interprets a disyllabic form for the word meaning ‘cat’, [m-inkʃ]-i, which is a compound noun: [m-tn] (tiger) + [i-j] (small). (the tilde is attached above [i] in IPA). Our syllable feature representation is given under the diagram. Both syllables are |nasal| having no {oral}. Therefore, each has a nasal vowel. The second syllable is denasalized in onset by {fricative}, which evokes the IRF for lifting up the velum. The first syllable is specified by {labial}, {stop}, {dorsal}, {front, high, front). Since the entire syllable is |nasal| by default, both onset and coda are nasal stops. The phonological representation is /big//, and the margins are |voiced| and |nasal| phonetically. In this language, if a syllable is not marked {oral}, then {stop} is implemented (by the crucial articulator according to the place specified) as a nasal stop. The second syllable is also |nasal| by default; the phonemic representation shows it by the lower-case for the vowel: //i//. The onset is a coronal (palatal, distinct from apical) fricative //f//, which is tense despite the nasal environment because it is {fricative}. Generally in this language, an obstruent is {tense} (unvoiced) by implication, unless a stop is specified for the feature {lax} in my understanding.

Kaiingang does not have any {fricative} with {lax} and no {voice} can be specified in this language, and therefore {lax} implies {stop}. The obstruent manner feature that can be specified for a margin is {stop}, {fricative}, or {lax}. The specification of {lax}, like other obstruent manner features, must be paired with a place feature: {labial}, {apical}, {coronal}, or {dorsal}. The feature {rhotic}, along with other sonorants, is oral if and only if the syllable is oral, otherwise being nasalized.

In the dialect we are concerned here (spoken around Sao Paulo), when the preceding nasal syllable has a {stop} in coda, and the succeeding syllable has an obstruent onset other than {lax}, implying {unvoiced}, voicing control for the preceding syllable (there cannot be a tense coda in Kaiingang) is terminated usually prior to the onset of the friction. This means, if the preceding nasal syllable has a (nasal) stop gesture in coda, part of the articulatory closure becomes unvoiced and an articulatory release may be heard for what would be called an epenthetic tense stop around the boundary.

In the CD diagram, note, around the boundary between syllables 1 and 2, the |nasal-jinal| switching in velum position step function coincides with voicing cessation, which in turn coincides with the joining point of the two syllable triangles.

6. Conclusion

Following the descriptive methodology of generative phonology [2], the C/D model [4] promises an effective descriptive framework for phonetic implementation. Phonetics has to be abstract as well as concrete.

7. References