

Ksenia Shalnova
OutsideEcho, UK
ksenia@outsideecho.com

Phone Inventory and G2P rules

Keywords: *Phoneme set, Phone set, G2P rules, diphone inventory creation, Speech Synthesis*

Contents

1. Introduction	3
2. Phoneme and allophone (phone) inventory	3
3. Suprasegmentals	4
3.1. Lexical Stress (Primary Stress)	4
3.2. Secondary Stress (Non-primary Stress)	5
3.3. Phrasal Stress	5
3.4. Tones	5
3.5. Intonation Patterns for Phrases	6
References	6
APPENDIX (Diphone Inventory creation)	7

1. Introduction

This document describes the procedure for creating phone inventory and G2P rules for the particular language. Although the procedure is language-dependent, it seems possible to outline several universal features to be taken into account.

2. Phoneme and allophone (phone) inventory

Phonemes (minimal linguistic units with a distinctive and constructive function) could be determined on the basis of "minimal pairs" when sound substitutions cause differences of word meaning, e.g., /di/ *dis* - /de/ *de* - /dE/ *dais* in French. It is important to notice that a little bit palatalised [d] in *dis* the French is not distinguished by native speakers from the non-palatalised [d] in *de/dais*. The reason is that these 2 sounds do not differentiate the meaning and, therefore, represent one phoneme.

In order to determine phonemic inventory in case when the minimal pair can not be found, it is possible to use the similar phonetic context or word "destruction" criteria, when changing of one phoneme into another one generates nonsense words.

Grapheme (grapheme combinations)	Phoneme (IPA symbol)	Sound description	Context	Example	
				Orthography	IPA transcription
<i>ou</i>	/u/	close, back, rounded	any	<i>goutte</i>	/gut/
<i>o</i>					
	ɔ	open-mid, back, rounded	before all graphemes, but <i>n</i> and <i>u</i>	<i>dot</i>	/dɔt/
	/ɔ̃/	open-mid, back, rounded, nasalized	before <i>n</i>	<i>long</i>	/lɔ̃/

Table 1. Grapheme-to-phoneme rules for vowels (Example from French).

Phoneme realizations should be distinguished from the so-called **free variation**. Free variation occurs when the variants of one phoneme can occur in the same phonetic environment. The principle of complementary distribution is broken, but the changing of one variant into another one does not influence the word meaning. These variants are usually differentiated stylistically/regionally/socially, e.g., different amount of aspiration in the English final consonant /p/ - *stop* [stop^h].

Allophones (phoneme realizations in speech) are mutually exclusive, i.e. they do not occur in the same phonetic environment (it is a rule of complementary distribution).

Two sounds that occur in different phonetic contexts are in complementary distribution → allophones of the same phoneme. If two sounds that occur in the same phonetic context are in contrastive distribution, they are the allophones of different phonemes. For example, the labialised allophone [s^w] of the phoneme /s/ occurs only before rounded vowels [s^wɔft] *soft*, whereas non-labialised one – only before non-rounded vowels [si:] *sea*.

Allophones can be described by means of IPA (International Phonetic Association) symbols (see [2] and [3]). The aim of such a description is non-ambiguous description of each particular sound based on its articulatory and acoustic characteristics.

The main sound classifications are the following.

3.2. Secondary Stress (Non-primary Stress)

In most European languages non primary (or secondary stress) occurs in clitics and compound words. In case of quality difference, separate allophones are worth introducing into phone inventory.

Vowel phoneme	Example	
	Orthography	IPA transcription
/i:/	<i>readability</i>	[,ri:də'bilɪti]
...	...	

Table 2. The list of vowel that can carry secondary stress (Example from English).

It seems necessary to point out another type of not-primary stress -a rhythm, which is sensitive to the place of the main stress (e.g., in Garawa, Seri etc).

Rule	Example
Primary stress falls on initial syllables, with secondary stresses falling on every other syllable thereafter.	/ˈmɛrɪˌpɛt/ 'beard'

Table 3. The rules for rhythm assignment (Example from Maranunggu).

3.3. Phrasal Stress

Phrasal stress is used to highlight a particular word in comparison to other words in a phrase (sentence)³. At present there is no typological classification for determining phrasal stress. The vowels that carry the main phrasal stress should be marked with a diacritic symbol.

For TTS systems, phrasal stress is usually determined on the basis of analysing syntactic structure that is closely related to Intonation Patterns in a particular language. Some isolating languages, like Chinese, have only phrasal stress (and no lexical stress) due to the fact that all words are monosyllabic.

3.4. Tones

Tonal languages can be subdivided into 2 types:

1. Languages that have got only lexical tones (Chinese). For better TTS quality, the vowels with different tones are worth including into the phone inventory as separate units (as any modification of tones will give unnatural sound).
2. Languages that have got both lexical and grammatical tones (Ibibio, Zulu). In this case the tone system is usually terraced. Some compromise should be found for including a sufficient number of vowels with different tone realisations as separate units into the phone inventory. Probably, it is worth starting with including vowels with different lexical tones.

³ Sentence comprises one or more phrase.

3.5. Intonation Patterns for Phrases

Sentence comprises one or more phrase. In European languages there are about 10 patterns for phrases realized in accordance with different sentence communicative types and the position within a sentence - sentence final/sentence internal positions. In G2P, intonation patterns can be marked at the phrase boundary.

Intonation pattern (rising/falling)	Comments	Example	
		orthography	F0 trajectory description (based on 4 levels)
Yes-no question (rising)		C'est le numero six ? (Is it number six?)	2: starts 1: penultimate syllable (possible, but not always) from 1 or 2 till 4: rise on the stressed syllable
Wh-question (falling)		Quand viendra-t-il? (When will he come?)	3 or 4: starts 1 or 2 : terminates
Finality (falling)	in declarative sentences	C'était en hiver. (It was in winter.)	2 or 3: starts 2: fall on the stressed syllable
Enumeration - (continuation rise - minor continuation)		L'automne a trois mois: <i>septembre, octobre, novembre.</i> (There are three months in autumn: September, October, december).	2 or 3: starts till 3 or 4: respectively rise on the stressed syllable
Major continuation (continuation rise)	at the phrase boundary	<i>La messe est terminée</i> et l'antique église se vide. (The mass is finished and the old church becomes empty.)	2: starts 4 or 5: terminates (1 phrase of the sentence)
Minor finality (continuation fall)	distinctive function	<i>Ton père, il est médecin?</i> (Your father, is he a doctor)?	2: starts 1.5: terminates (1 phrase of the sentence)
Exclamation (falling)
...

Table 4. Intonation Patterns for French.

References

- [1] Rob Goedemans, Harry van der Hulst & Ellis Visch "Stress Pattern of the World".
- [2] Handbook of the International Phonetic Association, 1999.
- [3] <http://www.arts.gla.ac.uk/IPA/ipachart.html> - description of the IPA symbols.

APPENDIX. DIPHONE INVENTORY CREATION

Diphone – sound segment composed of the transition from the middle of one phone to the middle of the next one (the plosive consonants are divided at the boundary of a stop and a burst). In order to obtain the diphone inventory for developing TTS system for a particular language it seems necessary to carry out several stages (Fig. 1). The output of the grapheme-to-allophone converter should correspond to the speech units present in the diphone inventory. Therefore, the process of the diphone inventory creation is closely related to the formalized description of the phoneme/allophone sets for each particular language. Such a description, which could be called the phonetic study, serves as a basis both for grapheme-to-allophone converter and for diphone inventory creation.

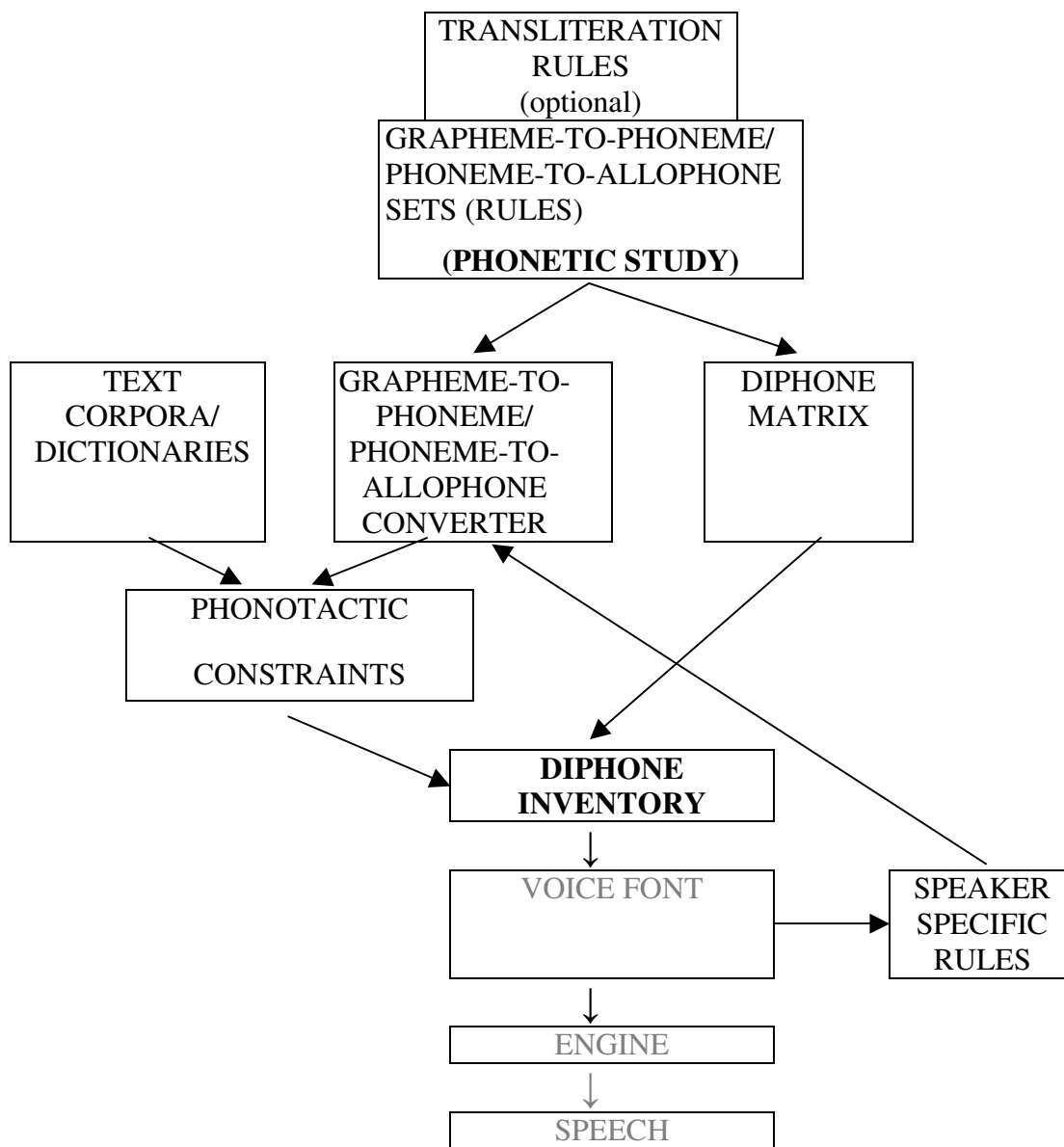


Fig. 1. The process of the diphone inventory creation.

Local Language Speech Technology Initiative

1. TRANSLITERATION (applicable for the languages with non-latin script)

2. GRAPHEME→PHONEME→ALLOPHONE (Phone) CONVERTER (based on a phone set)

3. SPEAKER SPECIFIC RULES

A speaker that has to be recorded for TTS voice may have certain peculiarities of pronunciation, i.e. specific allophone realizations. These peculiarities are usually due either to pronunciation variation within a normative language or to dialectal influence. When the speaker is selected, it seems necessary to analyse the speech and to adapt Phoneme→Allophone Converter to his voice.

4. PHONOTACTIC CONSTRAINTS

Theoretically the number of diphones for each language is equal to the number of all pair combinations of sounds (DIPHONE MATRIX) - minus phonotactic constraints - impossible sound combinations. The list of such impossible combinations could be obtained automatically on the basis of the processing of the transcribed TEXT CORPORA/DICTIONARIES. In order to obtain automatically PHONOTACTIC CONSTRAINTS it seems useful to develop first a grapheme-to-sound (allophone) converter.

5. DIPHONE INVENTORY

Creation of the diphone matrix is carried out on the basis of possible sound (allophone) combinations – minus PHONOTACTIC CONSTRAINTS.