SPEECH SYNTHESIS FOR URDU VOWELS USING HLSYN

M. USMAN AFZAL

ABSTRACT

This paper tries to give the brief overview of different kinds of speech synthesis systems (formant, concatenative and articulatory). General steps, which are involved in the synthesis, are discussed. Moreover, the Klatt synthesizer is also discussed in some detail.

This paper also includes the synthesis of Urdu oral vowels (ɑ, ɛ, æ, ʊ, i, e, ɔ, o, ə, ɪ, ʊ) using High Level synthesizer (HLSyn). Results show that HLSyn is good at synthesizing the Urdu oral vowels, but it does not give the control over bandwidth and intensity, which is a hurdle to archive natural sound.

1. INTRODUCTION

A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read *any* text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system.

There are different kinds of TTS systems. Formant synthesizer is a potentially powerful approach to speech synthesis. This approach has the advantage of extreme flexibility. Concatenative synthesizers possess a very limited knowledge of the data they handle: most of it is embedded in the segments to be chained up. It is generally highly intelligible. Articulatory synthesis is accomplished by computing the desired speech sounds directly from the physical structure and movement of the vocal tract.

2. Literature Review

2.1. Urdu Consonantal & Vocalic Sounds

Urdu, the national language of Pakistan, is spoken and understood by a large section of population of South Asian countries. Hindi is phonetically similar to Urdu, but it differs in its Script and historical characteristics.

The word Urdu has a Turkish origin, meaning ‘camp or army with its follower’. It is popularly regarded as offspring of Persian. It borrows words from different languages to expand its vocabulary. Major languages participating in the camp of Urdu are: Persian, Arabic, Portuguese and English (Saksena).

2.1.1. Review on URDU sounds

The pronunciation of Urdu varies from region to region and perhaps that is why there is no consensus in literature on the number of oral vowels in Urdu. According to Kachru (1990), there are seven long oral vowels, and three short oral vowels, and according to Bokhari (1991), there are seven long oral vowels, but seven short oral vowels. Bokhari contains many allophones of the corresponding long vowels as discussed by Kachru. Kachru maintains that the front low cardinal vowel [æ] exists as front middle low vowel [ɛ] in Urdu. As a result the back low cardinal vowel [ɑ] is shifted to the low center, making it [a] (Kachru, 1990). Alam also agrees with the long and short vowel distribution of Kachru. Bokhari and Alam list ten nasalized vowels including five short and five long nasalized vowels (Bokhari, 1985). Kachru (1990), on the other hand, has not listed any nasalized vowel, but mentions in the text that oral and nasal vowels contrast, and that nasalization is distinctive.

Collectively Kachru, Bokhari, Alam, and Hussain have listed forty-three (43) consonantal sounds of Urdu, out of which twenty-eight (28) sounds are agreed upon by all the above authors (see Appendix B). Kachru lists 37 consonants and has missed [ʔ, h, rʰ, nʰ, mʰ, lʰ]. Hussain (1997) lists 36 consonants and has missed [ŋ, ʈʰ, rʰ, nʰ, mʰ, lʰ, q]. Bokhari lists 36 consonants and he has missed [f, ʃ, ʒ, z, q, x, r]. Bokhari (1985) misses interestingly many basic sounds, which are listed by Kachru and Hussain (see Appendix B). Alam (1997) lists, most of all, 42 consonants and has missed only one consonantal sound [ŋ]. Overall, the controversial consonantal sounds are [ʔ, ʈʰ, rʰ, nʰ, mʰ, lʰ, ŋ].

All URDU consonants and vowels are listed in Appendix B.

2.2. Methods of Speech Synthesis

The different methods of speech synthesis discussed are:

1. Articulatory synthesis
2. Formant Synthesis
3. Waveform synthesis

2.2.1. Articulatory Synthesis

Articulatory Synthesis is a method of synthesizing speech by controlling the speech articulators (e.g., jaw, tongue, lips, etc.). Articulatory synthesizers attempt to model faithfully the mechanical motions of the articulators and the resulting distributions of volume velocity and sound pressure in the lungs, larynx, and vocal and nasal tracts (Bickley; 1999).

An all-out, “pure” phoneme-to-articulation-to-utterance machine would need to assemble the correct combination and timing of many speech muscle groups, corresponding to the different articulators (e.g. tongue, lips and glottal muscles). A true articulatory synthesizer could be developed, if no limits were placed on time and expense (Bickley; 1999).

In order to have a practical synthesizer capable of running on a reasonably priced computer at tolerable speed, “engineering” approximation to the speech production models have to be made. Most of the models, which were used in 1950s and 1960s, are based upon two-dimensional data from x-ray pictures. This data does not show the third dimension.

Articulatory synthesis is losing ground to other competing techniques because:

1. There still might be missing knowledge of speech production to push the accuracy of the models to the necessary level.
2. The problem of obtaining input/control data for our synthesis is to be solved.
3. There is no guarantee that the synthesizer produces high-quality speech. (Bickley; 1999)

The basic steps for Articulatory Synthesis are mentioned below.

- Excitation sources
- Models of the vocal tract
- Acoustic losses
- Webster's equation for acoustic radiation

2.2.2. Waveform Synthesis

Concatenative Synthesis methods result in highly intelligible and potentially very natural-sounding speech. Such methods consist of storing, selecting, and smoothly concatenating snippets of speech in and from their specific acoustic inventory (Bickley; 1999). Sound segments, which comprise of transition from the center of one phoneme to center of the next, are called dyads or diphones. The possibility of using them as the basic units for concatenation was first noted in the mid 1950 (Witten; 1982).

Imagine that one would record, in all desired voices, all existing words spoken in all sensible intonations. Such a system would be a truly smart speech play back system that would sound absolutely natural (Witten; 1982). Such a system would require drastic storage and speed. Voice coding techniques are used to improve storage. The recorded signal is coded efficiently, in as few bits as possible for the desired quality, and stored with the nametags for retrieval.

Various coding technique are used to represent the acoustic inventory units used for concatenative synthesis. LPC methods were widely used ten to twenty years ago; more recently, a waveform coding technique called pitch-synchronous overlap add (PSOLA) has been popular (Bickley; 1999).

Following are the basic steps for waveform synthesizers:

- Speech units (diphones, demisyllables, words, etc.)
- Parameterization and storage
- Concatenation
- Linear prediction

2.2.3. Formant Synthesis

Formant synthesizers, derive an approximation to a speech waveform by a simpler set of rules formulated in the acoustic domain. These are also called rule-based synthesizers.

Formant (Rule-based) synthesizers are mostly in favor with phoneticians and phonologists, as they constitute a cognitive, generative approach of the phonation mechanism. The broad use of the Klatt synthesizer, for instance, is principally due to its invaluable assistance in the study of natural speech characteristics, by analytic perception of rule-synthesized speech. What is more, the existence of relationships between articulatory parameters and the inputs of the Klatt model make it a practical tool for investigating physiological constraints (Dutoit; 1999).

Synthesizers of this type are based on the source filter theory of speech production. They utilize formant tracks, “basic” formant (F1, F2, F3) patterns for each phoneme, and track of F0 that specify source variations and other aspects of articulation. Because spoken phonemes are articulated in a context of adjacent phonemes and various stress patterns, their needs to be rules for modifying the basic phoneme formant patterns to create natural connections between them in the synthesized speech (Bickley; 1999).

The model is parametrically oriented. Notably, there are forty (40) odd parameters available to control the KLSYN88 (SENSIMETRICS, 1997). These parameters are in the appendix A.

There are two types of Formant synthesizers.

1. Parallel formant synthesizers.
2. Cascaded formant synthesizers.

Parallel Formant Synthesizers

In one type of configuration, called a parallel formant synthesizer, the formant resonators that simulate the transfer function on the vocal tract are connected in parallel as shown in the figure Each formant resonator is preceded by an amplitude control that determines the relative amplitude of a spectral peak (formant) in the output spectrum for both voiced and voiceless speech sound (Klatt, 1979).

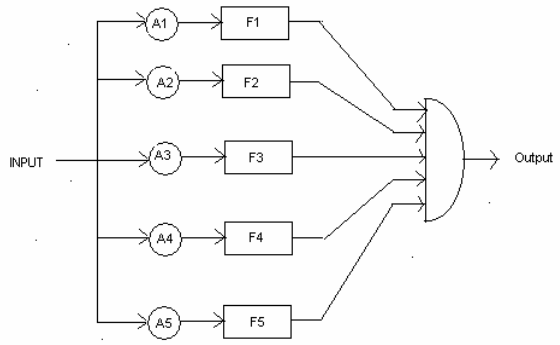


FIGURE 1 The transfer function of the vocal tract may be simulated by a set of digital formant resonators connected in parallel where each resonator must be produced by an amplitude controller A.

Cascaded Formant Synthesizer

In the second type of configuration, called cascaded formant synthesizer, sonorant is synthesized using a set of formant resonator connected in cascade, as shown in the picture (Klatt, 1979).

FIGURE 2 The transfer function of the vocal tract

Text-to-Speech Synthesis by Formant Synthesizer (Klatt Synthesizer)

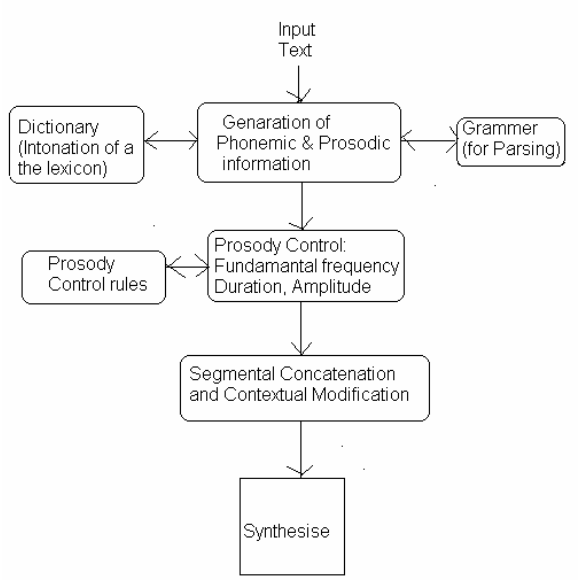


FIGURE 3 Step of Speech Synthesis

Speech Synthesizer function can be divided into following parts:

1. Generation of Phonemic and Prosodic Information
2. Prosody Control (Fundamental Frequency, Duration, Amplitude)
3. Segmental Concatenation and Contextual Modification
4. Signal synthesis (Klatt)

2.2.4. Generation of Phonemic and Prosodic Information

The major part of the speech synthesis system is the Text Analysis. The text analysis is itself composed of:

- A pre-processing module, which organizes the input sentences into manageable lists of words. It identifies numbers, abbreviations, acronyms and idiomatic and transforms them into full text when needed.
- A morphological analysis module, the task of which is to propose all possible part of speech categories for each word taken individually, on the basis of their spelling. Inflected, derived, and compound words are decomposed into their elementary grapheme units (their *morphs*) by simple regular grammars exploiting lexicons of stems and affixes.
- The contextual analysis module considers words in their context, which allows it to reduce the list of their possible part of speech categories to a very restricted number of highly probable hypotheses, given the corresponding possible parts of speech of neighboring words. This can be achieved with *n-grams, which* describe local syntactic dependences in the form of probabilistic finite state automata (i.e. as a Markov model).
- Finally, a syntactic-prosodic parser, which examines the remaining search space and finds the text structure (i.e. its organization into clause and phrase-like constituents), which more closely relates to its expected prosodic realization (Dutoit; 1999).

The Letter-To-Sound (LTS) module is responsible for the automatic determination of the phonetic transcription of the incoming text.

It requires a number of dictionaries:

- Pronunciation dictionaries refer to word roots only. They do not explicitly account for morphological variations (i.e. plural, feminine, conjugations, especially for highly inflected languages, such as French), which therefore have to be dealt with by a specific component of phonology, called *morphophonology*.
- Some words actually correspond to several entries in the dictionary, or more generally to several morphological analyses, generally with different pronunciations. This is typically the case of heterophonic homographs, i.e. words that are pronounced differently even though they have the same spelling.
- Pronunciation dictionaries merely provide something that is closer to a *phonemic* transcription than a *phonetic* one (i.e. they refer to phonemes rather than to phones). As denoted by Withgott and Chen (1993): "*while it is relatively straightforward to build computational models for morph phonological phenomena, such as producing the dictionary pronunciation of '*electricity*' given a baseform '*electric*', it is another matter to model how that pronunciation actually sounds*".

- Words embedded into sentences are not pronounced as if they were isolated. Surprisingly enough, the difference originates not only in variations at word boundaries (as with phonetic liaisons), but also on alternations based on the organization of the sentence into non-lexical units, that is whether into groups of words (as for phonetic lengthening) or into non-lexical parts thereof. Many phonological processes, for instance, are sensitive to syllable structure.
- Finally, not all words can be found in a phonetic dictionary: the pronunciation of new words and of many proper names has to be deduced from those of already known words (Dutoit; 1999).

Prosody Generation

The term *prosody* refers to certain properties of the speech signal, which are related to audible changes in pitch, loudness, and syllable length. Prosodic features have specific functions in speech communication. The most apparent effect of prosody is that of focus. For instance, there are certain pitch events which make a syllable stand out within the utterance, and indirectly the word or syntactic group it belongs to will be highlighted as an important or new component in the meaning of that utterance. The presence of a focus marking may have various effects, such as contrast, depending on the place where it occurs, or the semantic context of the utterance.

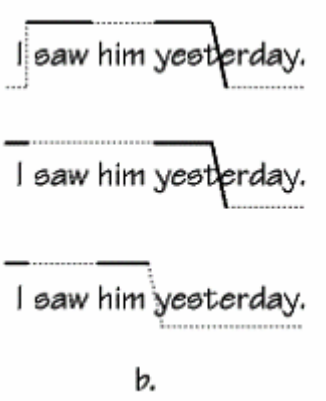


FIGURE 4 Different intonation patterns (a and b)

a. Focus or given/new information;
b. Relationships between words (saw-yesterday; I-yesterday; I-him)

Prosodic features create a segmentation of the speech chain into groups of syllables, or, put the other way round, they give rise to the grouping of syllables and words into larger chunks. Moreover, there are prosodic features, which indicate relationships between such groups, indicating that two or more groups of syllables are linked in some way. This grouping effect is hierarchical, although not necessarily identical to the syntactic structuring of the utterance (Dutoit; 1999).

2.3. Prosody Control

Speech has special voice features that are inherent to communication. When we talk with someone we usually want to convey two things: some objective information or the WHAT and some sort of attitude or HOW we feel about it. Both of these aims are accomplished by speaking with special variation of voice pitch and rhythm, called prosodic feature (Bickley; 1999).

Prosodic features play a strong role in the linguistics code of communication. For example, foreign speakers of English who articulate well all of the vowels and consonants are not easily understood unless they speak with correct English rhythms and intonation. The reason of our difficulty is that utterances are formed under defined rules of rhythm and intonation. These prosodic rules are unique for every language (Bickley; 1999).

For synthesis of natural-sounding speech, it is essential to control prosody, to ensure appropriate rhythm, tempo, accent, intonation and stress. Segmental duration control is needed to model temporal characteristics just as fundamental frequency control is needed for tonal characteristics. In contrast to the relative scarcity of work on speech unit generation, many quantitative analyses have been carried out for prosody control. Specifically, quantitative analyses and modeling of segmental duration control have been carried out for many languages (Sagisaka).

Traditional statistical techniques such as linear regressive analysis and tree regression analysis have been used for Japanese and American English, respectively. In one instance a feed-forward neural network has been employed to predict the interactions between syllable and segment level durations for British English. In this modeling, instead of attempting to predict the absolute duration of segments directly, their deviation from the average duration is employed to quantify the lengthening and shortening characteristics statistically (Sagisaka).

2.3.1. Intonation Analysis

Intonation analysis involves the following basic steps:

- Determination of the pitch contour. It is often said that pitch rises on a question and falls on a statement (Witten; 1982).
- Dividing the utterance into tone groups.
- Choosing the tone syllable, or major stress point, of each one.
- Assigning a pitch contour to each tone group.

In English, stressed vowels often get higher pitch and greater intensity than unstressed vowels in the same word. But it is not necessary in all languages. In Italian, stressed vowels in non-final open syllables are typically longer then unstressed one (Napoli, 1996).

2.3.2. Rhythm Analysis

Rhythm Analysis is concerned with the decision of locating foot boundaries within English text. Every language has its own rules to determine the stress.

One way to give a formal statement of the stress pattern is to talk about five parameters:

1. Boundedness
2. Quantity sensitivity
3. Headedness at the foot level
4. Headedness at the word level
5. Direction

We can make a set of rules using these five parameters.

There are major factors in prosodic features:

1. The fundamental voice frequency of vowels, responding to the amount of sub glottal pressure and vocal fold tension, is higher in stressed syllables than in unstressed syllables.
2. The intensity of the vowels in stressed syllables is higher than in unstressed syllables.
3. The relative intensity of F2, F3 and higher formants is grater in vowels of stressed syllables (Witten; 1982).

2.4. Segmental Concatenation and Contextual Modification

Prosodic features like stress affect segmental concatenation; but a notation, which indicates stressed syllables, is insufficient to capture this influence. Furthermore, contextual modification of segments, i.e. the co-articulation effects that govern allophones of phonemes, is explicitly rendered.

2.5. Synthesizer

A software formant synthesizer is described that can generate synthetic speech using a laboratory digital computer. A flexible synthesizer configuration permits the synthesis of sonorant by either cascade or parallel connection of digital resonators, but friction spectra must be synthesized by a set of resonators connected in parallel (Klatt, 1979).

The synthesizer design is based on an acoustic theory of speech production proposed in Fant (1960) and is summarized in Figure 5.

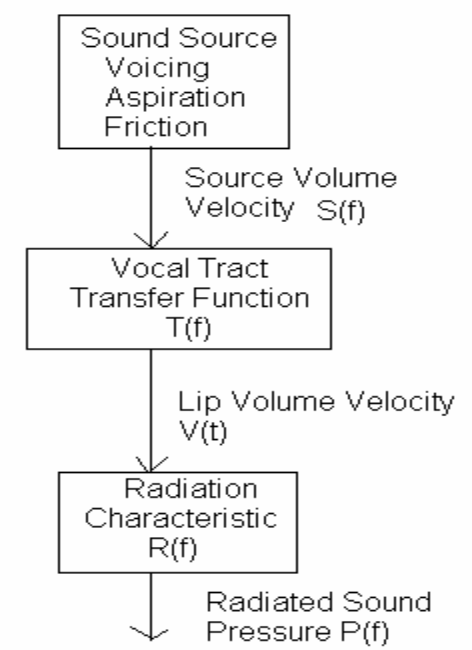


FIGURE 5 The output spectrum of a speech sound, P (f), can be represented in the frequency domain as a product of a source spectrum S (f), a vocal tract transfer function, T (f), and a radiation characteristic, R (f).

According to this view, one or more source of sound energy is activated by the buildup of lung pressure. Each sound source excites the vocal tract, which acts as a resonating system analogous to an organ pipe. Since a vocal tract is the linear system it can be characterized in the frequency domain by a linear transfer function, T (f), which is a ratio of lip-plus-nose volume velocity, V (f) to source input, S (f). Finally, the spectrum of the sound pressure that would be recorded some distance form the lips of the talker P (f), is related to the lip-plus-nose volume velocity, V (f) by a radiation characteristic, R (f), that described the effects of directional sound propagation from the head (Klatt, 1979).

2.5.1. Cascaded versus Parallel Synthesizer

The first advantage of the cascade connection is that the relative amplitudes of formant peaks for vowels come out just right (Fant, 1965) without the need for individual amplitude controls for each formant. The disadvantage is that one still needs a parallel formant configuration for the generation of fricatives and plosive bursts (the vocal track transfer function cannot be modeled adequately by five cascaded resonators when the sound source is above the larynx) so that cascade synthesizers are generally more complex in overall structures.

The second advantage of the cascade configuration is that it is a more accurate model of the vocal tract transfer function during the production of non-nasal sonorant. As will be shown, the transfer functions of certain vowels are difficult to match using a parallel formant synthesizer. A parallel synthesizer is particularly useful for generation of stimuli that violate the normal amplitude relations between formants (Klatt, 1979).

Klatt Parameters are discussed in Appendix A.

3. METHODOLOGY

For the purpose of synthesis of URDU vowels, the source filter theory of speech production has to be implemented. HLSyn is used to answer that description.

3.1. Speech Synthesizer

For the synthesis of Urdu vowels HLSyn (High-level Parameter Speech Synthesis System version 2.2) is used, which employs the Klatt synthesizer. HLSyn converts the high level parameters into the Klatt parameters by itself.

HLSyn provides an integrated environment for specifying, creating, analyzing and comparing synthetic speech files using high-level synthesis. HLSyn program also supports formant synthesis using the underlying Klatt-type cascade-parallel formant synthesizer.

Following is the list of HL parameters, with brief description.

f1, f2, f3, f4
First four natural frequencies of vocal tract. These are the natural frequencies when the velo-pharyngeal port is closed,

f0
Fundamental frequency of vocal-fold vibration. This HL parameter is usually identical to the KL parameter F0.

ag
Area of glottal opening.

al
Cross-sectional area of constriction formed by the lips during the production of labial consonants.

ab
Cross-sectional area of constriction formed by the tongue blade during the production of coronal consonants.

an
Cross-sectional area of velo-pharyngeal port.

ue
Rate of increase of vocal-tract volume that is actively controlled during the constricted interval for an obstruent consonant.

3.2. Subjects

A test was arranged to check the correctness of the synthesized sound of vowels. Eleven Vowels are arranged randomly five times each and ask from a group of 5 Urdu native people to tell what they perceive. That result is composed in a confusion matrix.

Furthermore to confirm the data, the spectrogram of Actual sounds and synthesize sound were compared.

3.3. Data Recording and Processing

All the acoustic analysis of the speakers was carried out on the X-Waves 5.3, a collection of digital speech-processing tools designed for Linux users. PAART, another speech analysis software was also used.

4. RESULTS

4.1. Confusion Matrix

	ɑ	ɛ	æ	ʊ	i	e	ɔ	o	ə	ɪ	ʊ
ɑ	25										
ɛ	6	13	5			3			4		
æ	2		23								
ʊ				20				5			
i					25						
e			1			24					
ɔ	1						20	4			
o				4				21			
ə		3							22		
ɪ										25	
ʊ											25

5. DISCUSSION

By looking at the results, we conclude that, Back high vowels [o] and [u] are difficult to perceive. F1 and F2 are overlap. F1 and F2 of [u] are between 250 to 350 and 800 to 900 respectively. And the F1 and F2 of [o] are between 325 to 425 and 850 to 950 respectively. There is also involving the difference of pitch and intensity. The Intensity of [o] is higher then [u]. It is very difficult to control the intensity using HLSyn. Another thing, which was concluded, was that, the error in perception of [o] in place of [u], when [u] comes after the [ɑ] vowels in the test. [ɑ] is the back low cardinal vowels and [u] is the back high cardinal vowel.

By looking the result, [ɛ] is between [æ] and [e]. It means that [ɛ] sound is difficult to perceive and mostly the listeners mapped it on the one three vowels ([æ], [e], [ə]). [ɛ]. It is easily recognized when it is assumed to be a short vowel. If it is assumed it as a long vowel then the listeners usually map it on [æ] or [e].

5.1. Limitation

HLSyn is a good tool to synthesize the oral vowels. The only problem is it is a high level tool. It does not give the control on the bandwidth and intensity, which is a hurdle in the synthesis of natural sound.

6. SUMMARY

The different techniques of speech synthesis were presented and the formant synthesizer was discussed in some detail. Steps for synthesis of speech were described and the results of synthesis efforts of Urdu vowels with HLSyn were given in the form of spectrograms and confusion matrix.

7. REFERENCES

Bickley, C. Ann, S. and Schroeter, J. 1999, p.325-335. “The Acoustics of Speech Communication”, Surry

Witten, H. 1982 p 361 “Principle of Computer Speech” University of Calgary, Canada

Dutoit, T. 1999 “A Short Introduction to Text-to-Speech Synthesis” TSTC Lab

SENSIMETRICS 1997 “High-Level Parameter Speech Synthesis System” Massachusetts

Klatt, D. 1979 “Software for a cascade/parallel formant synthesizer” Cambridge

Saksena, Ram Babu A History of Urdu.

Hussain, S. 1997. Phonetic Correlates of Lexical Stress in Urdu

Bokhari, S. 1985. Phonology of Urdu Language

Kachru, Yamuna 1990. *Hindi-Urdu* in *The Major Languages of South Asia, The Middle East and Africa*, edited by Bernard Comrie.

Sagisaka, Y. “Spoken Output Technologies” ATR, Japan
www.csku.cse.ogi.edu

Napoli, D. 1996 “Linguistics An Introduction”, Oxford University

Bokhari, Sohail 1991. Urdu Zubaan ka Soti Nizaam.

Haqqi, S. “Farhang-e-Talffuz”, Muqtadra Qaumi Zubaan. ISBN 969-474-153-X.

Feroz-ul-Lughat Urdu, ISBN 969-000-514-6

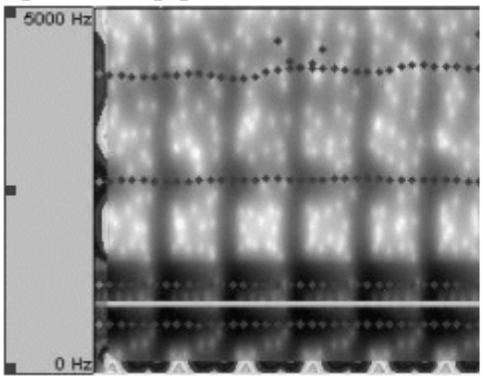
Saksena, Ram Babu A History of URD

8. APPENDICES

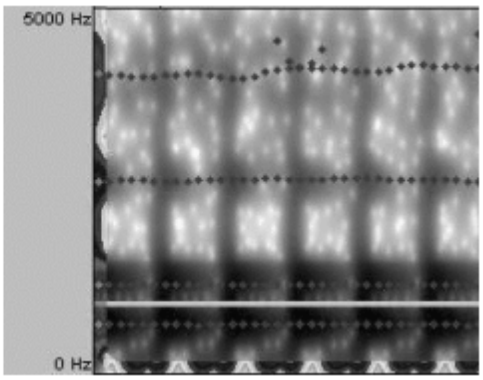
8.1. Appendix A

8.1.1. Spectrograms

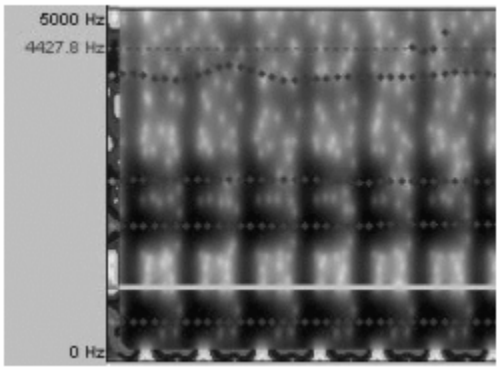
Synthesize [ɑ]



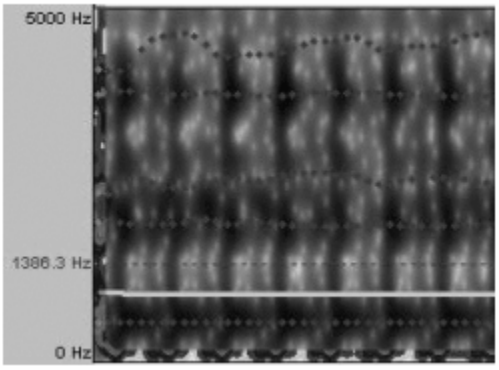
Actual ɑ



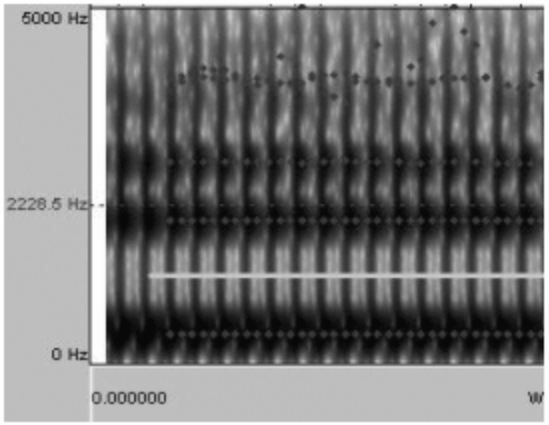
Synthesized [æ]



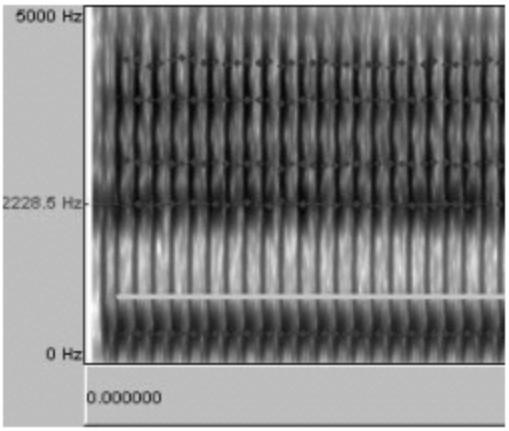
Recoded æ



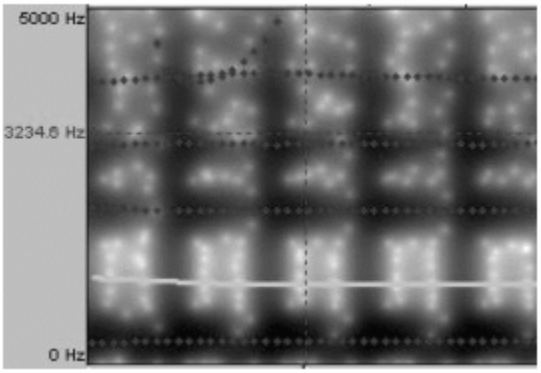
Synthesized [e]



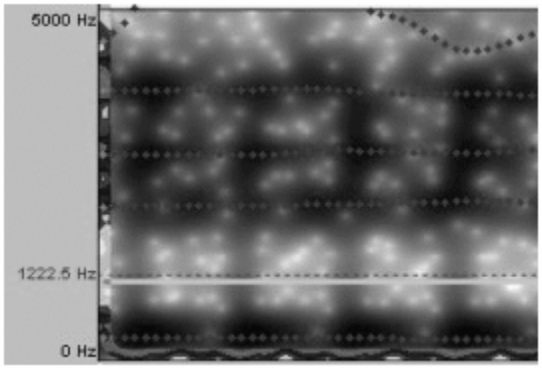
Recorded [e]



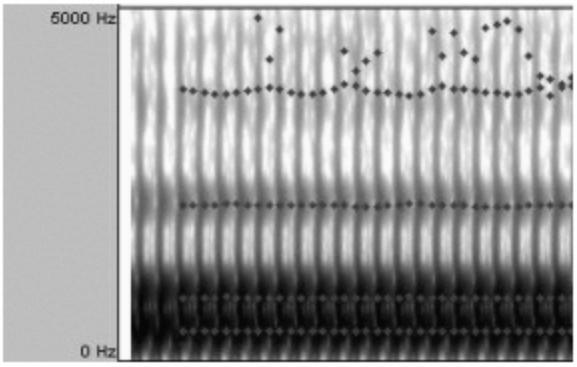
Synthesized [i]



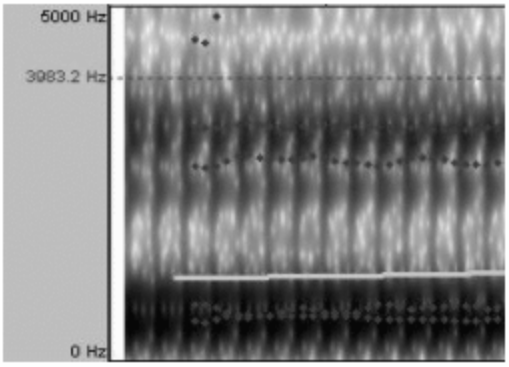
Recorded [i]



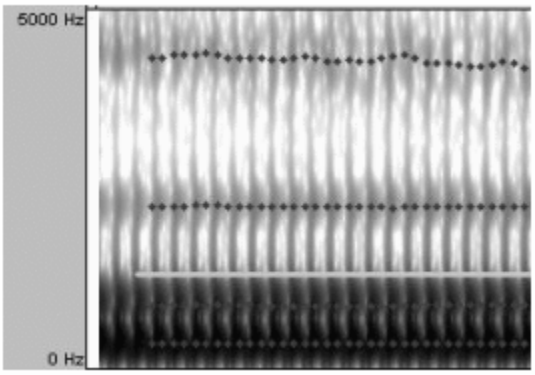
Synthesized [o]



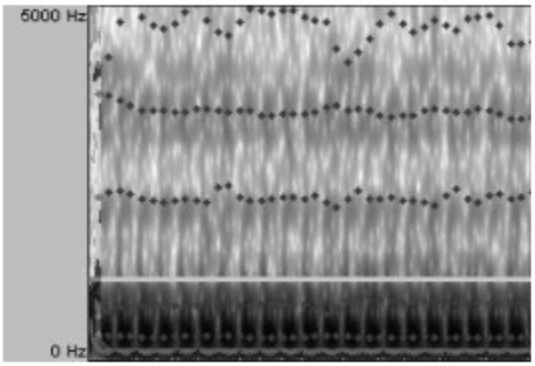
Recorded [o]



Synthesized [ʊ]



Recorded [ʊ]



8.2. Appendix B

TABLE A.1 Klatt Parameters

No	Name	Meaning	default	Min	Max
1	AV	Amplitude of voicing	0	0	80
2	AF	Amplitude of Frication	0	0	80
3	AH	Amplitude of Aspiration	0	0	80
4	AVS	Amplitude of Sinusoidal voicing	0	0	80
5	F0	Fundamental Frequency	0	0	500
6	F1	First Formant	450	150	900
7	F2	Second Formant	1450	500	2500
8	F3	Third Formant	2450	1300	3500
9	F4	Fourth Formant	3300	2500	4500
10	FNZ	Frequency of Nasal Zero	250	200	700
11	AN	Amplitude of Nasal formant	0	0	80
12	A1	Amplitude of F1 (Parallel only)	0	0	80
13	A2	Amplitude of F2 "	0	0	80
14	A3	Amplitude of F3 "	0	0	80
15	A4	Amplitude of F4 "	0	0	80
16	A5	Amplitude of F5 "	0	0	80
17	A6	Amplitude of F6 "	0	0	80
18	AB	Amplitude of Cascade/Parallel Bypass	0	0	80
19	B1	Bandwidth of F1	50	40	500
20	B2	Bandwidth of F2	70	40	500
21	B3	Bandwidth of F3	110	40	500
22	SW	Parallel/Cascade switch*	0	0	2
23	FGP	Frequency of Glottal Pole	0	0	600
24	BGP	Bandwidth of Glottal Pole	100	100	2000
25	FGZ	Frequency of Glottal Zero	1500	0	5000
26	BGZ	Bandwidth of Glottal Zero	6000	100	9000
27	B4	Bandwidth of F4	250	100	500
28	F5	Fifth Formant Frequency	3850	3500	4900
29	B5	Bandwidth of F5	200	150	700
30	F6	Sixth Formant Frequency	4900	4000	4999
31	B6	Bandwidth of F6	1000	200	2000
32	FNP	Frequency of Nasal Pole	250	200	500
33	BNP	Bandwidth of Nasal Pole	100	50	500
34	BNZ	Bandwidth of Nasal Zero	100	50	500
35	FRA	Second Glottal resonator bandwidth	200	100	1000
36	SR	Sampling rate	10000	5000	20000
37	NWS	Number of samples per frame	50	1	200
38	GAI	Overall Gain control	48	0	80
39	NFC	Number of cascaded formants	5	4	6
40	CO	Overall gain control (db)	0	80	47

8.3. Appendix C

TABLE B.1 Consonants

Type	Place	Manner	Sound Symbol	Minimal Pairs
Stops	Bilabial	Voiceless	p	pap, bap, bʰap
		Voiced	b	pəlna, pʰəlna
		Aspirated Voiceless	pʰ	pan, man
		Aspirated Voiced	bʰ	bənda, pʰənda
		Nasalized Voiced	m	əbər, əmər
				pʰaɽa, bʰaɽa
				pʰəlna, məlna
				bʰərɑ, mərɑ
		Aspirated Nasalized Voiced	m^h	*ɾʊmʰẽ*
	Dental	Voiceless	t̪	t̪al, t̪ʰal, d̪al
		Voiced	d̪	t̪ali, t̪ʰali
		Aspirated Voiceless	t̪ʰ	t̪ar , d̪ʰar
		Aspirated Voiced	d̪ʰ	t̪an, t̪ʰan, d̪ʰan, nan
		Nasalized Voiced	n	d̪aɽ, d̪ʰaɽ
				d̪am, nam
				t̪ʰən, d̪ʰən
		Aspirated Nasalized Voiced	n^h	
	Alveolar	Voiceless	ʈ	ʈalna, ɖalna
		Voiced	ɖ	ʈaɽ, ʈʰaɽ
		Aspirated Voiceless	ʈʰ	ʈal, ɖʰal
		Aspirated Voiced	ɖʰ	ɖənɖa, ʈʰənɖa
				ɖal, dʰal, ʈʰal
	Velar	Voiceless	k	kali, gali
		Voiced	g	kaja, kʰana
		Aspirated Voiceless	kʰ	kaʈi, gʰaʈi
		Aspirated Voiced	gʰ	gana, kʰana
				gaʈi, gʰaʈi
				kʰəɽa, gʰəɽa
	Uvular	Voiceless	q	qələm, ʔələm
	Glottal	Voiceless	ʔ	ʔəlim, qəlim
Fricatives	Labio-Dental	Voiceless	f	faʃ, laʃ
		Voiced	v, w	vaɖı, ɖaɖı
	Alveolar	Voiceless	s	ser, zer
		Voiced	z	zer, ser
	Palatal	Voiceless	ʃ	ʃaɖı, saɖı
		Voiced	ʒ	ʒala, nala
	Velar	Voiced	ɣ	ɣərib, qərib
	Uvular	Voiceless	x	xana, gana
	Glottal	Breathy Voiceless	h	hamı, namı
Affricates	Alveolar	Voiceless	tʃ	tʃəla, dʒəla
		Voiced	dʒ	dʒəla, tʃəla

Trills	Palatal	Voiced	r	muqərər, muqərəb
			r^h	
Flap	Palatal	Voiced	ɽ	baɽ, bal
			$ɽ^h$	
Approximants	Back	Central	j	gaja, gana
	Middle	Lateral	l	lana, dʒana
			l^h	

TABLE B.2 Vowels

Type	Position		Sound Symbol	Minimal Pairs
	Front/Back	High/Low		
Long	Front	High	i	bin, bən, bɪn
		Middle	e	beʃərəm, baʃərəm
		Low	æ	bæl, bəl
	Back	High	u	pura, para
		High-Middle	o	bona, bəna
		Low-Middle	ɔ	pɔd̪a, pæd̪a
		Low	a	ban, bin
Short	Front	High	ɪ	d̪ɪn, d̪an
		Middle	ɛ	sɛhər, səhər
	Back	High	ʊ	sʊn, sən
	Middle	Middle	ə	kəlɪ, kʊlɪ
Nasalized Long	Front	High	ĩ	pɛhnĩ, pɛhna
		Middle	ẽ	kəhẽ, kəha
		Low	æ̃	hæ̃, hi
	Back	High	ũ	$k^h\widetilde{au}$, $k^h\widetilde{ao}$
		High-Middle	õ	k^hanõ, k^hana
		Low	ã	ləɽkɪã, ləɽkɪõ