iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** X **Month of publication:** October 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Factors Influencing Voice Onset Time (VOT): Voice Recognition

Jasdeep Kaur
Assistant Professor, Department Of Physics, A.S College, Khanna, Punjab

Abstract-In this paper we study the methods to evaluate the features which might estimate the closeness of acoustic signals. For measuring Voice onset time, we were considered 3 stop consonants in word- initial position and two abutted vowels /a/ and /i/. These 3 stop consonants were divided into: Voiced (/b, d, g /) and Voiceless (/p, t, k /) categories. The utterances were acoustically analyzed via PRAAT. The VOT values of the target stops were obtained from the waveform and verified with the spectrogram. As for the spectrographic readings, VOT intervals from the beginning of the release burst to the onset of voicing were analyzed. It has been found that voice- onset time for voiceless consonants is larger than the voiced consonants. Further, VOT get changed with effect of place on articulation .It has been found that Velar has largest VOT following in descending order by Alveolar and Labial. In addition to it, effect of vowel height on voice onset time was studied, which reveal that High vowel /i/ have larger voice- onset time than Low vowel /a/.
Key words-Phonetics, Speech communication, Speech recognition, Artificial intelligence, Acoustics.

I. INTRODUCTION

One of the most difficult aspects of performing research in speech recognition by machine is its interdisciplinary nature. The most important discipline that has been applied to speech recognition problem is acoustics. Acoustics is science of understanding the relationship between the physical speech signal and physiological mechanisms (human vocal tract mechanism) that produced the speech and with which the speech is perceived (human hearing mechanism). The description, modeling and explanation of speech communication in the languages of the world is main goal of phonetics which is “The study of the spoken medium of language” in the broadest Sense. According to Diehl (1991), what has been considered to be purely phonetic is also phonological in character; that is to say, the domains of phonetics and phonology overlap significantly.

In most languages the written text does not correspond to its pronunciation so that in order to describe correct pronunciation some kind of symbolic presentation is needed. Every language has a different phonetic alphabet and a different set of possible phonemes and their combinations. The acoustic signals are converted by listener into a sequence of words and sentences. The most familiar language units are words. They can be thoughts of as a sequence of smaller linguistic units, phoneme, which are the fundamental units of phonology. Phonemes are abstract units and their pronunciation depends on contextual effects, speaker's characteristics, and emotions.

The phoneme can be characterized as:

Basic concept of phonetics
Smallest unit of language, existing as such speech sound which is capable differentiating one word from another, or one grammatical form from another.
Speech sound that makes a difference in meaning
A class or family of sounds regarded as a single sound and represented in transcription by the same symbol
Abstractional and generalized in character exists in our minds as an abstraction and at the same time is generalized in speech in the form of its allophones
Some efforts to construct language-independent phonemic alphabets were made during last decades. One of the best known is perhaps IPA (International Phonetic Alphabet) which consists of a huge set of symbols for phonemes, suprasegmentals and tones/word accent contours.

The study of speech communication is the study of human symbolic behaviour in many forms. Speech is the oldest academic discipline and one of the most modern in its concern with interpersonal relationships. Communication helps us form relationships, allows cultures to evolve, and encourages understanding among people; in short, communication is the thing that makes us human.

Without it, we would perish. The study of speech is based on the assumption that one's ability to communicate in an effective manner is vital to successful human interaction. Four major areas are included in the speech discipline, each of which focus on unique characteristics of particular situations. Interpersonal communication includes the study of symbolic behaviour in dyadic, two person, and relationships. Group communication concentrates on the small group of three to seven persons. Organizational communication examines the effects that organizational structure and membership have on human communication. Rhetoric and public address is the study of discourse and its role in shaping public perceptions and practices. All areas emphasize effective oral and written communication. Phonetics helps us to know how children learn the sounds of their first language, what can be done in case of speech and hearing defects, how is speech perceived and produced. Speech is one of the common modes of intelligence communication from one place to another. Speaking to each- other is one of the most important things that we human beings do. Each of us has a Mind, a private world filled with thoughts, feelings and memories; the faculty of expressing thoughts by words or articulate sounds are called Speech. Speech refers to the processes associated with the production and perception of sounds used in spoken language. Speech consists of complex sound signals produced by the human vocal apparatus-an apparatus, which has the capability of producing a wide variety of speech. To study the various voice recognition systems of Hindi language, and after seeing the pressing need of the day, it has become necessary to fabricate a voice recognition system of Hindi language. To develop a voice recognition system for Hindi language, we have to study the various acoustic cues or parameters of Hindi speech sounds.

A. Voice Onset Time (VOT)

Refers the time that elapses between the release of the articulators for a stop and the onset of vocal cord vibration of the following segment. This period is usually measured in milliseconds (ms). It is useful to distinguish at least three types of VOT which are shown in the schematic diagram (fig.1) below:

1) Negative VOT: where the onset of vocal fold vibration precedes the plosive release. If the voicing starts before the release (i.e., during the closure phase) of the stop, then the result is described as 'voice lead' (or 'prevoiced) and is given a negative VOT value. This is the case of voiced plosives.

2) Zero VOT: where the onset of vocal fold vibration coincides (approximately) with the plosive release. Voiceless unaspirated plosives have zero voice-onset time.

3) Positive VOT: where there is a delay in the onset of vocal fold vibration after the plosive release. If the voicing starts after the release of the stop, then the result is 'voice lag' and is described with a positive VOT value. All voiceless aspirated plosives have positive value of voice- onset time.

The amount of lag is important to separate voiceless unaspirated ('short lag') from voiceless aspirated ('long lag').

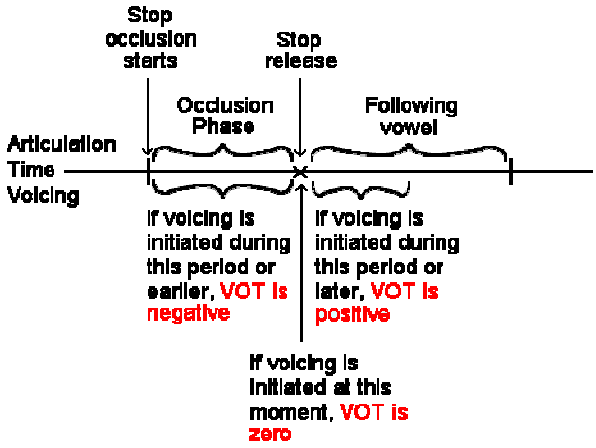


Figure 1. A graphical representation of the VOT of voiced, voiceless unaspirated and voiceless aspirated plosives

This diagram defines Voice Onset Time (usually referred to as VOT). VOT can be positive, zero or negative and the timing is always relative to the stop release burst. English /p, t, k/ are produced with a long lag; /b, d, g/, on the other hand, may be produced with a short lag.

B. Factors Having Significant Effect On Voice- Onset Time

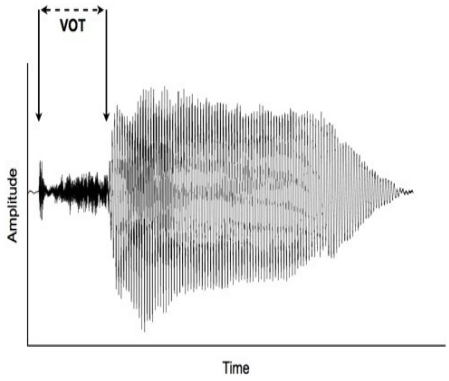
Voicing of stop,
The place of articulation of the target stop,
The height of the vowel following the target stop, and
Stress.

Voice- onset time is greater for voiceless stop as compare to voice stop. In relation to the former, it has been stated that the lag is longer as we move the place of articulation from front to back (i.e., from bilabial stop to alveolar stop and then to velar stop). This is due to degree of abruptness of the pressure drop upon the release of the stop. The more sudden (abrupt) the pressure drop is, the sooner the voicing of the next segment starts. Consequently, this results in less aspiration (i.e., shorter lag). When we look at stops at the three places of articulation, we see that the tongue dorsum separates more slowly (less abrupt) from the velum for /k/ than the tip from the alveolar ridge for /t/, or from the lips for /p/. Voice- onset time increases as we move the place of articulation from front to back. In addition to the place of articulation of the stop, some studies suggest the effect of the height of the following vowel. Specifically, greater lag was observed when stops were followed by high vowels (more open articulations) than when they are followed by low vowels (narrower opening). The reason for this effect again is related to the abruptness of the pressure drop. High vowels have a more obstructed cavity than low vowels. Since the high tongue position that is assumed during the stop closure in anticipation of a subsequent high vowel would result in a less abrupt pressure drop, a stop produced as such will have longer lag than the one produced before a low vowel. A factor that we have some reason to expect will be significant for the timing of voice onset is one of stress. All phonetic descriptions of the English stops indicate that stressed voiceless (/p, t, k/) stops have greater voicing lag than the unstressed voiceless (/p, t, k/) stops. Therefore stress is an important parameter for voice- onset time.

C. Measurement Of VOT

The utterances are acoustically analyzed via PRAAT. The VOT values of the target stops were obtained from the waveform and verified with the spectrogram. The beginning of the lag (positive VOT) was identified by a sharp spike where the waveform changes from quiescent to transient; the end point (onset of vocal cord vibration) is determined from where the waveform becomes periodic (figure 2). As for the spectrographic readings, VOT intervals from the beginning of the release burst to the onset of voicing were analyzed. In other words, VOT is a property of stop consonants (e.g., /b/ and /p/) defined as the time between the release of the stop and the onset of subsequent vocal fold vibration. To measure VOT, one looks for the release of the stop consonant on the waveform, this is usually marked by a sudden spike of a periodic energy - and then looks for the onset of vocal fold vibration, which is marked by high amplitude, periodic energy. VOT is calculated as the time between these two points.

The picture on next page shows a waveform for the utterance “tea”; on this waveform, VOT for /**t**/ is indicated.



II. OBJECTIVES

To study the various writing operated enquiry systems of Hindi language. We are interested in finding out acoustic cues in the final position of stop consonants. Voice recognition system of a language which will be useful for various voice operated enquiry systems. So a thorough study of Acoustic parameters is a must. Utilization of speech for communication between man and machine has been significant requisite and the main motivation factor behind developing speech interactive systems. The objectives of the proposed study are:

To select a sizeable number of Hindi speech sounds involving words with 3 consonants at word- initial position with two abutted vowels /a/ and /i/.

To record these Hindi speech sounds in sound- proof lab, in order to get the required Database.

To make use of adopted softwares, to digitize these Hindi speech sound samples.

To measure voice onset time from the spectrograms of different Hindi meaningful syllables.

III. RESULTS

The Voice onset time of 3 stop consonants in word- initial position of meaningful Hindi syllables with two abutted vowels /a/ and /i/ for tokens spoken in isolation by speakers is presented in Table 1, Table 2and Table 3.

	/a/	/i/	Average
Voiced	41	44	42
Voiceless	48	52	50
Ratio	0.85	0.84	0.84
Difference	-7	-8	-8

Table 1 Voice –onset time (in m sec) for voiced and voiceless words:

It has been found from Table 1 that voice- onset time for voiceless consonants (Avg 50 msec) is larger than the voiced consonants (Avg 42 msec).

The bar graph shown in figure 2 confirms the result of Table 1.

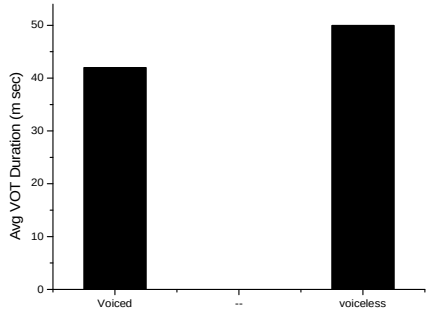


Figure 2.Bar graph showing avg. values of VOT (in m sec) of both voiced and voiceless stops.

Volume 3 Issue X, October 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

	Place of articula-tion	Voiceless	Voiced	Avg.	VOT (msec)
/i/	Velar	67	52	60	48
	Alveolar	52	44	48	
	Bilabial	38	35	37	
/a/	Velar	63	48	56	44
	Alveolar	48	37	43	
	Bilabial	35	32	34	

Table 2 shows Effect of Vowel height on voice – onset time. It has been found from table that High vowel /i/ have larger voice-onset time (Avg 48 msec) than Low vowel /a/ (Avg 44 msec). Another interesting aspect comes into play from table 2 that bilabial stops have minimum value of voice onset time for both the low as well as for high vowel. Voice onset time for alveolar stops was more than bilabial but less than velar stops.
The graph as shown in the figure.3 confirms the results of Table 1 and Table 2.

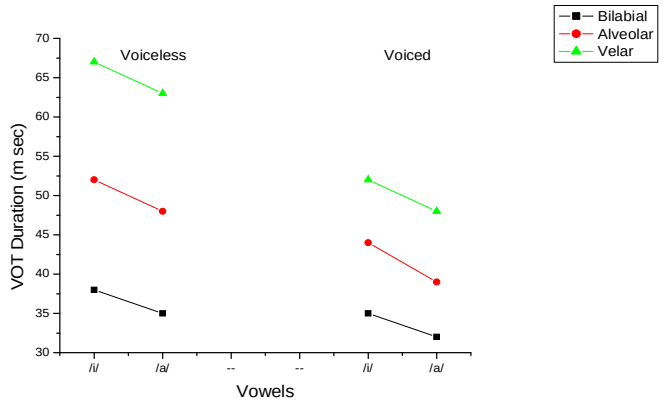


Figure 3. Graph shows VOT for voiceless and voiced stops which show variation for high and low vowels for three different stop categories.

Ohala et.al (1992). Reported initial results of an acoustic phonetic study of Hindi designed to further text- to- speech synthesis of the language. Some phonetic parameters are described for e.g., the effect of stop voicing on preceding vowel duration, the effect of vowel height on intrinsic vowel duration. They measured vowel duration: effect of voicing, vowel duration: effect of aspiration and VOT (voice onset time). For measuring VOT of both Voiced and voiceless stop in four places of articulation with three vowels /a/,/i/ and /u/, they concluded that Velars show greater VOT than the other places of articulation with all vowels.
Yavas (2008) in his paper defined voice onset time as the time that elapses between the release of the articulators for a stop and the onset of vocal cord vibration of the following segment. It has been stated that the lag is longer as we move the place of articulation from front to back (i.e., from bilabial to alveolar and then to velar).
In addition to the place of articulation of the stop, it also suggests the effect of the height of the following vowel. Specifically, greater lag was observed when stops were followed by high vowels (more open articulations) than when they are followed by low

vowels (narrower opening).

Lisker et. al. Concluded that voice onset time increases as the place of articulation moves back in the mouth. They found differences that show certain regularity: the intervals by which voice onset either leads or lags behind release are almost always significantly shorter in sentences, whether the voicing begins before or after release. They concluded that voiced and voiceless categories are very clearly differentiable on the basis of variations in the timing of voice onset when they occur initially in isolated words.

Cho and Ladefoged observed that velar stops always have a longer VOT. Furthermore, in both aspirated and unaspirated stops, VOT is shortest before bilabial stops and intermediate before alveolar stops, with the exception of the unaspirated stops in Tamil and the aspirated stops in Cantonese and Eastern Armenian. They have shown variations of VOT for 18 languages associated with a difference in the place of articulation.

Khattab in this paper presents findings from an instrumental study of Voice Onset Time (VOT) production in 3 English-Arabic bilingual children and 3 monolingual controls from each language. The subjects were all aged between 5 and 10, and the recordings were made. The aim was to find out whether bilinguals acquire separate VOT patterns for each language and whether these patterns are parallel to the monolingual ones. In English, the monolingual and bilingual children produce their VOICELESS stops with long lag and their VOICED stops with short lag, and VOT increases as the place of articulation moves further back in the mouth.

IV. CONCLUSION

The following conclusions were made from the above study:

Voice onset time for voiceless consonants is larger than the voiced consonants.

Effect of place of articulation on VOT: The VOT varies as:

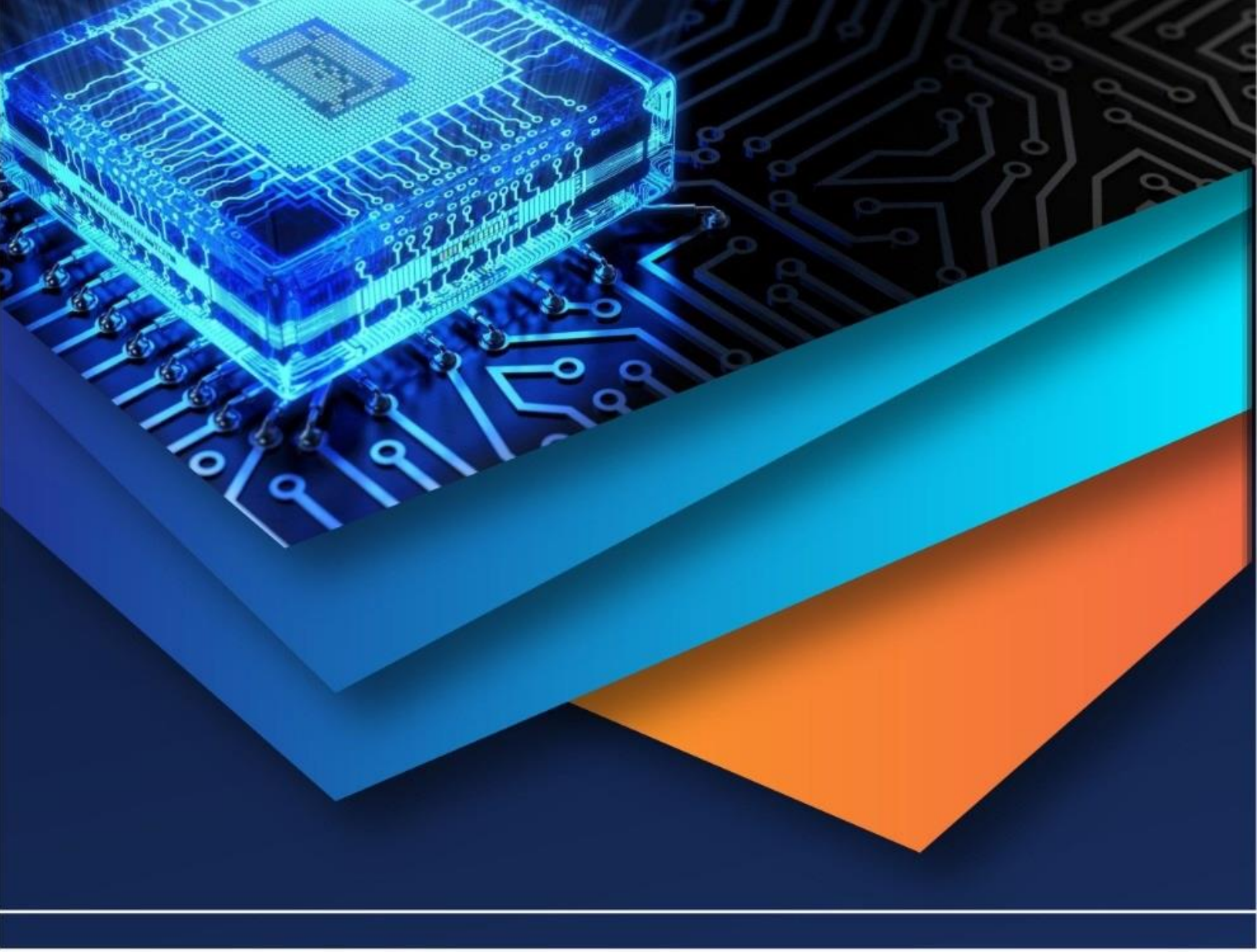
Velar > Alveolar > Bilabial

Effect of Vowel height on voice onset time: VOT for

High vowel (/i/) > Low vowel (/a/)

REFERENCES

[1] P.Ladefoged, Vowel and Consonants-An Introduction to Sounds of Languages, Wiley-Blackwell, ISBN 0-631-21412-7, (2001) 1-2.
[2] L.Rabiner and B.H. Juang, Fundamentals of Speech Recognition, PTR Prentice-Hall International, Inc., New Jersey 07632, ISBN 0-13-285826-6, (2007) 11-20.
[3] M.Ohala and J.J.Ohala,, “Phonetic universals and Hindi segment duration”, published in Second International Conference on Spoken Language. (1992)
[4] M.Yavas, “Factors Influencing the VOT of English Long Lag Stops and Interlanguage Phonology”, Florida International University, United States. (2008).
[5] L. Lisker, “Closure duration and the intervocalic voiced-voiceless distinction in English”, Language, (1957) , 33(1), 42-49.
[6] T.Cho, and P.Ladefoged, “Variation and universals in VOT: evidence from 18 languages”, Journal of phonetics, (1999) 27, 207-229.
[7] G.Khattab,” VOT production in English and Arabic bilingual and monolingual children”, Leeds working papers in Linguistics, ,(2000) 8, 95-122.
[8] R.L.Diehl, “The role of Phonetics within the study of language”, Phonetica,(1991) 48, 120-134.
[9] P.A.Keating, “Phonetic representations in a generative grammar” , Journal of phonetics, (1990),18, 321- 334.
[10] H.Kim, J.Cole, H.Choi, and M.H. Johnson, “The Effect of Accent on Acoustic Cues to Stop Voicing and Place of Articulation in Radio News Speech”, Proceedings of Speech Prosody, (2004) 29-32.
[11] Neti, C.; Rajput, N.; Verma, A. , (2002) , “A Large vocabulary continuous speech recognition system for Hindi”, Proceedings of the National Conference on Communication. Mumbai, India



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY