

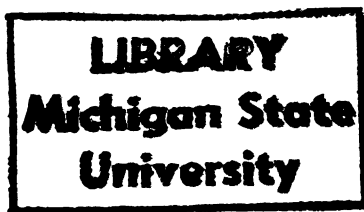


22147304

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 00569 4520



This is to certify that the

dissertation entitled

INTRASUBJECT VARIABILITY IN VOICE ONSET TIME  
AS A FUNCTION OF ALCOHOL INDUCED CHANGES  
IN PHYSIOLOGICAL STATE

presented by

Bradford L. Swartz

has been accepted towards fulfillment  
of the requirements for

Ph.D. degree in Audiology  
and Speech Sciences

Major professor

Date May 6, 1988



RETURNING MATERIALS:

Place in book drop to  
remove this checkout from  
your record. FINES will  
be charged if book is  
returned after the date  
stamped below.

--	--	--

INTRASUBJECT VARIABILITY IN VOICE ONSET TIME  
AS A FUNCTION OF ALCOHOL INDUCED CHANGES  
IN PHYSIOLOGICAL STATE

by

Bradford L. Swartz

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

1988



5/1/93-2

## ABSTRACT

### INTRASUBJECT VARIABILITY IN VOICE ONSET TIME AS A FUNCTION OF ALCOHOL INDUCED CHANGES IN PHYSIOLOGICAL STATE

By

Bradford L. Swartz

Voice onset time was first defined twenty-four years ago and has been the topic of a wealth of research since. It is known to be variable within and between individuals. But individual variability over time and as a function of change in physiological state has not been previously investigated. This study evaluated intrasubject voice onset time variability over time and under conditions of intoxication and sobriety for the /d/ and /t/ phonemes. Eight men and eight women of the same age range who had normal hearing and speech were the subjects. Voice onset time was determined using visual inspections of digital oscillograms. Obvious changes in time domain information established the points of stop closure release and onset of periodicity. This methodology yielded a test-retest reliability-of-measure of 1.46 [ms].

Results indicated there was no significant difference in voice onset time variability over time or as a function of intoxication versus sobriety. Level of



intoxication was not found to be correlated to differences in mean voice onset time. Three temporal features of voice onset time combined in various ways to form six voice onset time types. The predominant type used by all subjects was what is traditionally considered to be positive voice onset time. However male subjects used a greater diversity of these six types than females, and consistently demonstrated shorter voice onset times than females. Implications for further research are suggested.



Accepted by the faculty of the Department of Audiology and  
Speech Sciences, College of Communication Arts and  
Sciences, Michigan State University, in partial fulfillment  
of the requirements for the degree of Doctor of Philosophy.



May 6, 1988

---

Oscar I. Tosi, Ph.D.  
Dissertation Committee Chairman



This dissertation is dedicated to:

The memory of professor Richard L. Powell, philosopher,  
educator and friend, who helped me appreciate the magic of  
speech, and whose inspiration lives on;

and to

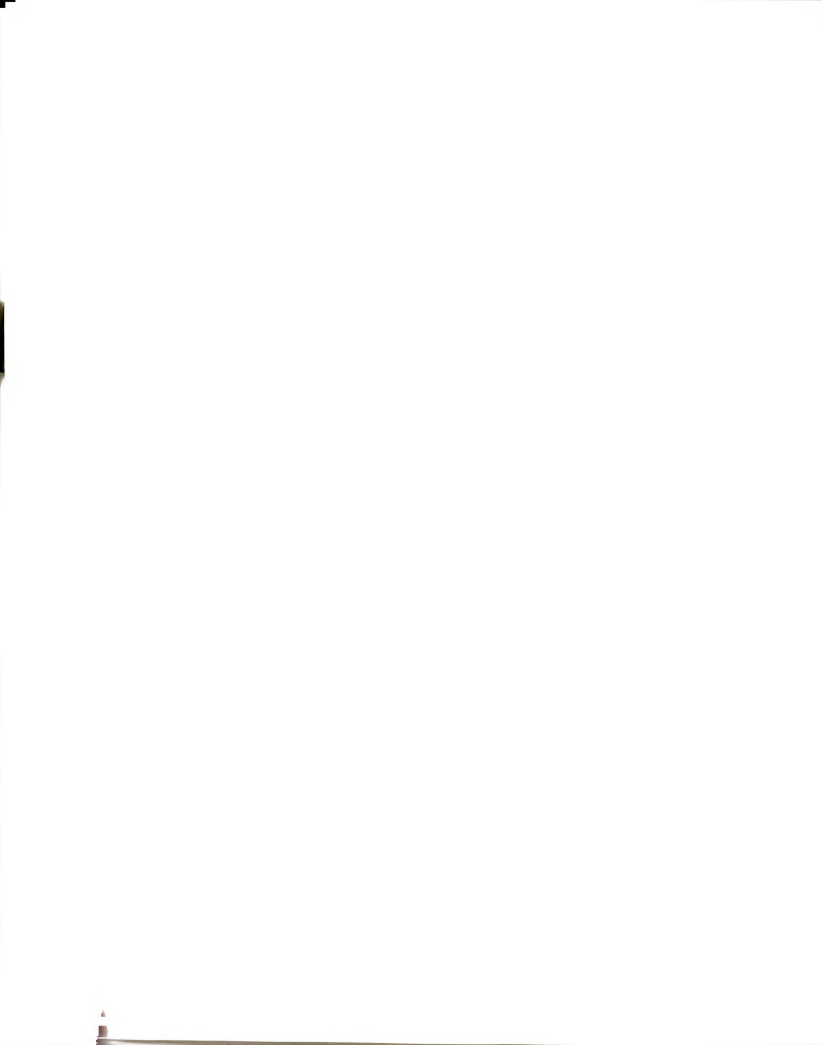
Judi my wife, and my daughters Meredith and Marissa. The  
love of my family is the greatest possession of all.



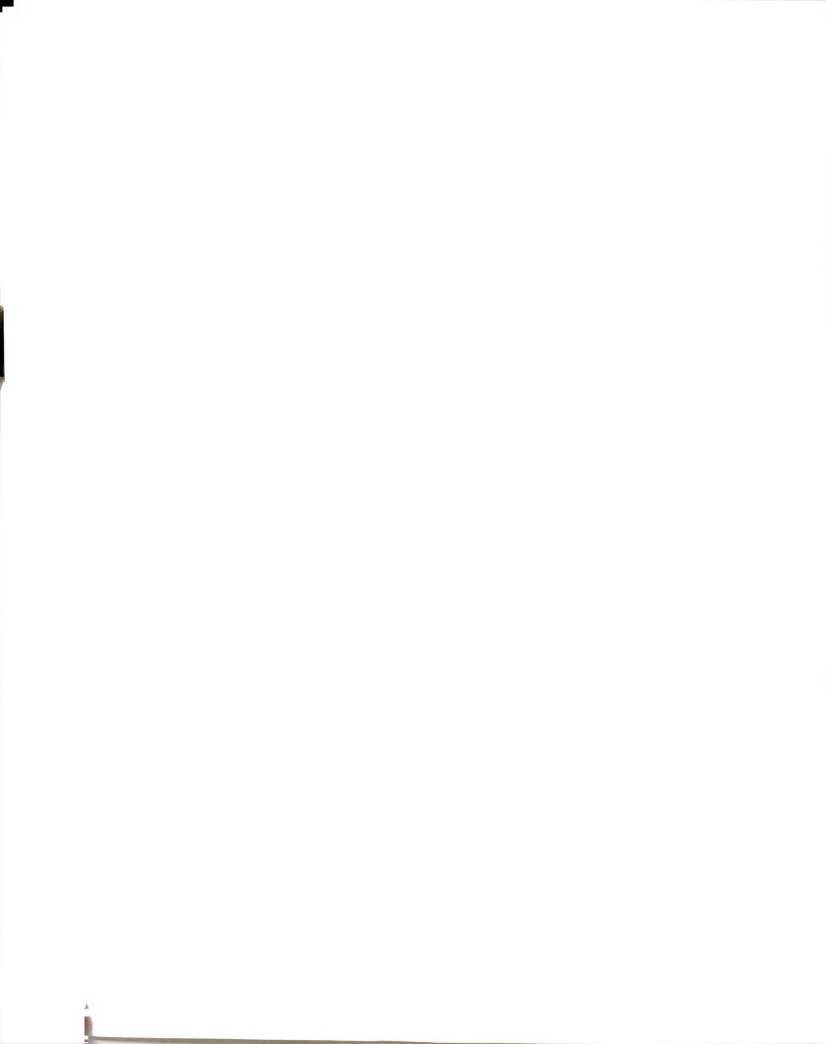


## ACKNOWLEDGMENTS

The completion of this dissertation is but the final step along a path which began seven years ago. That path could not have been traveled without the assistance of some important people. First and foremost the author thanks his wife Judi, whose love and encouragement have helped keep all the work in the proper perspective. Dr. Oscar Tosi has been this author's major professor, and has imparted wisdom as well as knowledge. Dr. Leo Deal served as academic advisor. His unwavering faith in the consistent yet part-time pursuit of this degree has been a valuable source of strength. The other members of the dissertation committee were Dr. Paul Cook and Dr. John Pumplin. This author thanks them for their direction and comments. Thank you also to WJR radio for over 39,000 miles of listening pleasure, to the people in the Department of Communication Disorders at Central Michigan University for their moral support over the years and the use of facilities during the preparation of this dissertation, to Marilyn Chamberlain and Emily Henderson for their many smiles, encouraging words, and parking tokens, and to Carl Lee for assistance in interpreting



the statistics in this study. A special thank you to Lt. Stan Dinius and the officers of the CMU Department of Public Safety who gave their time and talents to conduct the breathalyzer tests. Finally, thank you to my dissertation subjects, who gamely took part in this study. They now realize that drinking alone in a sound proof booth is really no fun at all.



## TABLE OF CONTENTS

LIST OF TABLES . . . . .	ix
LIST OF FIGURES . . . . .	xi
INTRODUCTION . . . . .	1
Null Hypotheses	
Operational Definitions	
Voice Onset Time (VOT)	
Physiology or Physiological State Change	
Intoxication	
Spectrographic and Oscillographic Analyses	
REVIEW OF THE LITERATURE . . . . .	14
METHODOLOGY . . . . .	23
Subjects	
Voice Onset Time Stimuli	
Data Collection	
Treatments and Physiological States	
Traditional VOT Measures	
Digital Signal VOT Analysis	
Types of VOT Measures	
Recording Methods	
Statistical Analysis	
RESULTS AND DISCUSSION . . . . .	37
Differences in Voice Onset Time Variability	
Non-Rejection of the Null Hypotheses	
Differences in Voice Onset Time Means	
Six Variations in Voice Onset Time	
Discussion	
Voice Onset Time Variability Over Time	
Voice Onset Time Variability Due to a Change	
in Physiological State	
Variations in Voice Onset Time Type	
Sex Differences in Voice Onset Time	
Digital Instrumentation and Methodology	
Comparisons to Classical Results	

CONCLUSIONS . . . . .	71
Suprasegmental versus Segmental Differences	
Changes in Variability versus Changes in Means	
Sex Differences in Voice Onset Time	
A Closer Look at the Negative VOT Aspects	
Computer Analyses of the Oscillogram	
Clinical Implications	
APPENDICES	
A Sentences Used in the Articulation, Voice and Rhythm Screening . . . . .	76
B Subject Interview Questionnaire . . . . .	77
C General Information Provided to the Subjects at the Interview . . . . .	78
D Informed Consent Form . . . . .	79
E Sentences Used for the Collection of VOT Data, and Their Phonetic Transcriptions . . .	80
F Voice Onset Time Data Collection Form . . . .	81
G Time Required for Intoxicated Treatment, Alcohol Consumed, and PBT Level for Each Subject . . . . .	82
LIST OF REFERENCES . . . . .	83

## LIST OF TABLES

1	Subject log transformations of standard deviations of /d/ and /t/ VOTs across four treatments, and respective p values . . . .	39
2	Log transformation p values for the entire subject population, across treatments, sexes, and sentence group replications . . . . .	41
3	Means and standard deviations of /d/ VOTs across treatments, sexes, and sentence numbers .	43
4	Means and standard deviations of /t/ VOTs across treatments, sexes, and sentence numbers .	43
5	Analysis of variance for /d/ VOTs, across sentence group replications, sex, treatments, and sentence number. . . . .	44
6	Analysis of variance for /t/ VOTs, across sentence group replications, sex, treatments, and sentence number. . . . .	44
7	Analysis of variance for /d/ VOTs, across sentence group replications, sex, treatments, and sentence number, with intoxication level as a covariate . . . . .	45
8	p values of males and females for linear and quadratic covariate regression effects on /d/ VOTs. . . . .	46
9	Means and standard deviations of /d/ and /t/ VOTs, broken down by sex and treatment . . .	47
10	Analysis of variance for /t/ VOTs, across sentence group replications, sex, treatments, and sentence number, with intoxication level as a covariate . . . . .	48

11	p values of males and females for linear and quadratic covariate regression effects on /t/ VOTS. . . . .	49
12	Six possible VOT variation types, respective data codings, and interpretations. . . . .	50
13	Frequency of /d/ VOT types (for all sober treatments) for each subject, frequency of each type across all subjects, and percentage of each type across all subjects. . . . .	55
14	Frequency of /d/ VOT types (for the intoxicated treatment) for each subject, frequency of each type across all subjects, and percentage of each type across all subjects . . . . .	56



## LIST OF FIGURES

1	Typical displays used to discern VOT . . . . .	13
2	Suggested curvilinear relationship of the covariate effect . . . . .	46
3	Oscillograms demonstrating the six variations of VOT type . . . . .	52
4	Idealized versions of oscillograms of the six variations of VOT type . . . . .	53
5	Histogram showing relative frequencies of six variations of /d/ VOT types as used by males and females . . . . .	57
6	Standard deviation logarithmic transformations across four treatments, for subjects who demonstrated significant difference in VOT variability. . . . .	60
7	Male and female /d/ and /t/ VOT means across treatments. . . . .	63



## INTRODUCTION

Speech is a temporal phenomenon. The ongoing production of distinctive features, phonemes, syllables, words, phrases and longer utterances requires a timed, coordinated effort on the part of the speaker. Several of these temporal variables, for instance stress, intonation and timing; are suprasegmental and are often characteristic of specific languages.

Within the latter of these features, that of timing, fall the subgroups of rate, rhythm and the use of pauses. While rate of speech and even rhythm are relatively easy to conceptualize and measure, the pauses in our speech are another matter. Part of the problem arises from a lack of a definition of a pause. MacKay (1987) defined pauses as simply those occurring during the breathing cycle, or others which he called prolongations of sound or hesitation noises. Another perspective is that of Tiffany and Carrell (1977) who stated that the short, long and extended pauses in speech may be of several types: unfilled or silent pauses; or filled pauses consisting of adventitious and repetition pauses.

An early examination of pauses in general was conducted by Goldman-Eisler (1958). She examined hesitation pauses in ongoing speech and found a relationship between the length of the pause (in this case pauses between words) and the transition probability of the



words to follow. However, she only defined pause as silence in comparison to the occurrence of speech. In later studies (Goldman-Eisler, 1961, 1972) she evaluated the distribution of various lengths of pauses ranging from 0.25 to 8 seconds. She noted that none of her work accounted for pauses of less than 0.25 seconds. She termed these very short pauses articulatory and nonsignificant, contrasting them with the hesitation pauses in which she was in fact interested.

This lack of a definition of pause, as well as the inaccurate measure of true articulatory pauses, was rectified by Tosi (1974) with the introduction of pausometry. By defining a pause as being less than some predetermined maximum amount of acoustic energy occupying a greater than some predetermined minimum amount of time, he was able to efficiently measure pauses of all types, even those occurring within the normal production of consonants in a language.

A particular articulatory pause of this kind had been described nearly a decade earlier by Lisker and Abramson (1964) as voice onset time (VOT). This was defined as the time between the instant of stop closure release and the initiation of voice in consonant-vowel syllables beginning with the stop consonants /p, b, t, d, k, g/. The instant of release of the stop closure was labeled as time zero. If the voicing preceded the release,



it was termed negative VOT. If it lagged behind the release, it was labeled positive VOT. They thus laid the ground work for measuring VOT and found to their surprise that in English, at least, positive VOT did not always yield a voiceless sound. Rather, the amount of positive VOT was the critical distinction and was the real difference between the voiced and voiceless cognates of /p/ and /b/, /t/ and /d/, and /k/ and /g/. In expanding a definition of the voiced/voiceless distinction between stop consonant cognates, they hypothesized the existence of this temporal feature in addition to the known variations in presence or absence of voice, strength of aspiration, and the fortis/lenis dichotomy.

In their classic study, Lisker and Abramson studied VOT, as they defined it, in eleven different languages, in contexts of isolated word-initial position, and in initial and embedded-in-sentence positions. In the isolated words, they found VOT separated the voiced from the voiceless homorganic stops in all of the two phonemic category (voiced and voiceless) languages such as English. There was a less clear distinction in three category languages (voiced, voiceless aspirated, and voiceless unaspirated) and even more ambiguity in four category languages. Even within the two category languages, there was overlap of VOT values if the voiced and voiceless groups were taken together. For example, English





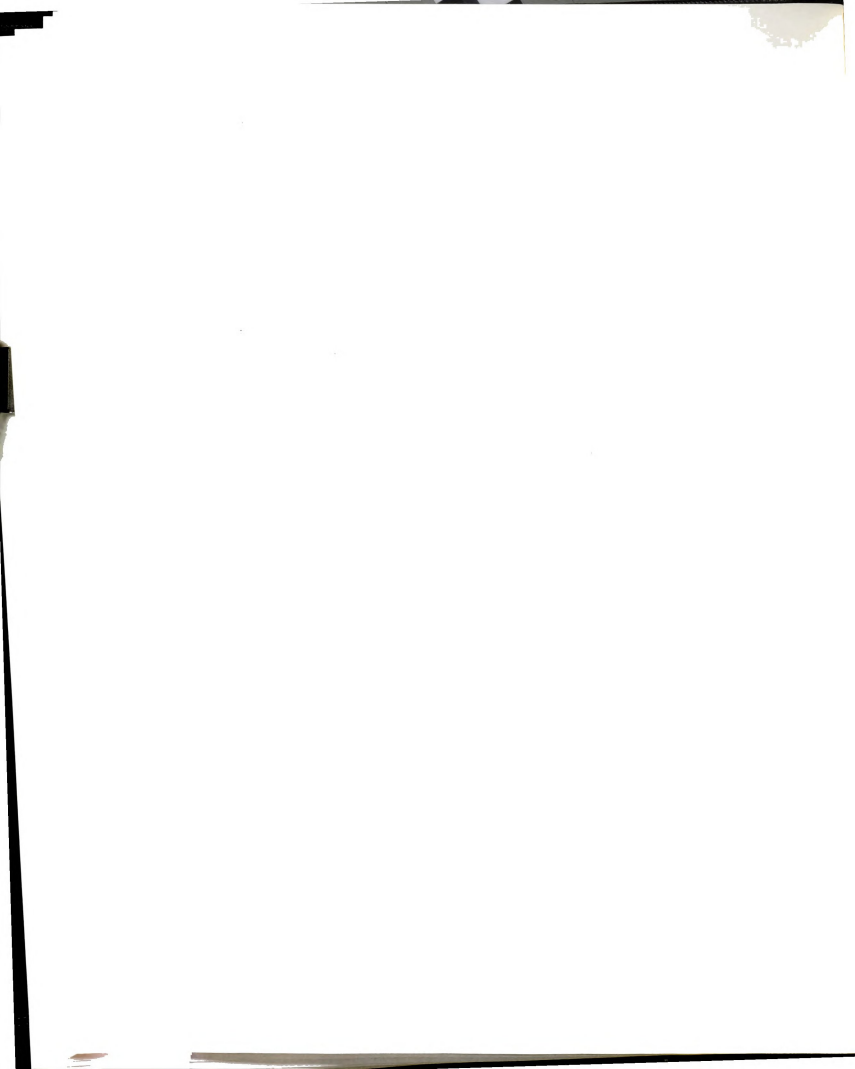
demonstrated a VOT boundary region between /p,t,k/ and /b,d,g/ of about 20 to 30 milliseconds ([ms]). However, there was no overlap within each homorganic pair, such as /p,b/. In the within-sentence experiments, they reported data congruent with isolated words, though the ranges of VOT were more compressed.

These ranges of VOT for each language were reported in their study, and in some cases there was only one speaker of the language used. Thus, it was really a reporting of a single subject's variability. In other languages there were up to four speakers, so the VOTs were more indicants of means across subjects. In discussing the variability in VOT for each subject or for a group of subjects, they noted the wide dispersion of VOT values within each phonemic group. For example, in English where four speakers were used, the /p/ VOTs in words ranged from 20 to 120 [ms]. In Cantonese, where only one subject was used the /p<sup>h</sup>/ VOTs ranged from 30 to 110 [ms]. Of the four subjects used in English, three had positive VOTs for the voiced group of phonemes (/b,d,g/), whereas one subject had all negative VOTs. Thus there certainly was variability noted across subjects. However, it was Lisker and Abramson's contention that each subject had a distinct range in which he produced VOTs for voiced versus voiceless stops. Beyond mention of the existence of such group variability and the probability of intrasubject relative

consistency, they conducted no study nor reported any results of within-subject variability.

Following this initial investigation of VOT, a large number of studies have been conducted over the past twenty-plus years. These studies have focused on either the production or perception of VOT and its value as a phonemic feature of speech. By and large, the research has been finely focused within production or perception, to closely examine VOT in various native languages, in specific communication disorders, or as a function of subject age and speech development. In total, nearly 120 investigations have been conducted since 1964 into the productive and perceptual attributes of VOT, relative to various population groups. In addition to these clinical studies, there have been a few which have attempted more accurate measures of VOT, in contrast to the spectrographic analyses employed by Lisker and Abramson (1964).

Three facts are evident following a review of the available literature. 1) There have been few investigations into variations in VOT production within the individual; 2) there have been no investigations into adult intrasubject variability in VOT over a series of trials, and; 3) as an extrapolation of these first two facts, there has been no research into the variations in VOT within the individual as a result of changes in physiological state. It is therefore the intent of this study to explore



VOT from the perspective of intrasubject variability over a series of trials, and further, to determine whether there is a change in intrasubject variability as a result of a change in physiological state.

Only a handful of studies are available which report intrasubject data. Among them, Eguchi and Hirsch (1969) reported intrasubject variability of children versus adults in VOT and other frequency measures of speech. They collected /p/ and /t/ samples during a single recording session, during which the same stimulus sentence was repeated five times. They found decreasing and stabilizing individual variability of VOT throughout early childhood, until about 7 to 8 years of age. Macken and Barton (1980) reported results of VOT development in a longitudinal study of toddlers over approximately one year. VOTs were collected on all six English voiced and voiceless plosives every two weeks. Individual means, standard deviations and distributions of VOTs were reported for each of several of the sessions in the study for each of the four children used. The authors proposed a three stage developmental model of VOT acquisition. Till and Stivers (1981) examined /t/ and /d/ VOT in three adult subjects and reported intrasubject variability and standard deviations from a one-session recording. It is noteworthy that they reported considerable variability in two and perhaps all three of their subjects and that their data were randomly selected,

being only 50% of that actually collected. Such variability in itself was not surprising. But because of the single recording session, there was no consideration of changes in this variability over time, nor with concurrent changes in some aspect of physiology. In their conclusions, Till and Stivers expressed a need for more work to be done in the area of intrasubject variability, especially regarding normal versus articulation disordered children. While Till has reportedly gathered more data in intrasubject VOT variability among adults and children, his research has not been published. Sweeting and Baken (1982) examined /p/ and /b/ VOT in young adult, middle-aged, and elderly subject groups. They reported intrasubject and group means and standard deviations obtained during a single recording session. They noted increasing variability with age although group means did not differ. The increased variability was attributed to a breakdown in temporal coordination as a function of age.

Physiology and physiological state are commonly defined as the processes, activities and phenomena associated with the cells, tissues and organs of the living body (Thomas, 1970). Changes in physiology can be brought about by the introduction of some foreign matter into the system (as with alcohol or some other drug); or as a result of the body's response to some external stimuli (as in increased adrenalin flow following a shock or scare).



With such changes actual physical, chemical or electrical differences can be noted. Tissues can increase or decrease in size, blood levels of various chemicals can be measured, or muscle tension and nervous transmissions as a function of body electrical activity can be discerned. Whether permanent or temporary manifestations, these changes constitute a difference in physiology from that measured at another time or place. Some motoric behaviors which can be altered during physiological changes include pulse rate, breaths per minute, eye blinks, and finger or limb reaction time. That temporal aspects of the voice can be altered with changes in physiological state is not new either. Witness the research into jitter and shimmer values and the use of such data as a diagnostic tool in vocal cord pathology (Iwata and Von Leden, 1970; Kitajima and Gould, 1976). Also, pauses in general have been evaluated as a function of circadian rhythm and the introduction of the hallucinogenic compound psilocybin. Daily rhythmical variations had no noted effect on pauses, whereas physiological changes from psilocybin did, as noted by Tosi and Lashbrook's; and Tosi, Fischer and Rocky's work (cited in Nakasone, 1979). In each of these studies, however, VOT was not discerned specifically. Indeed, all speech pauses as defined by Tosi in terms of maximum acoustic energy over a minimum amount of time were grouped together under the general term of pause.





Though voice changes such as shimmer and jitter are apparently evident with permanent disruptions in physiological state and general changes such as pauses are evident even with temporary physiological swings, specific changes concurrent with temporary alterations in physiology have not been explored. If certain changes in the temporal features of an individual's voice do occur with temporary alterations in one's physiological state, the use of such information can serve as a noninvasive diagnostic instrument for such physiological changes. It is the assumption of this writer that such changes will occur when a subject is intoxicated, thus subjected to a temporary and measurable change in physiological state. Further, such changes can best be examined in terms of intrasubject variability, as there is a large recognized intersubject variance in VOT, as already described in the available literature.

Changes in VOT are the result of exceptionally fine motor coordination involved in the production of speech. It is the opinion of this writer that such motor movements are not likely to be under the speaker's conscious control. Rather, these temporal alterations are likely to be a result of true changes in physiology. VOT can possibly serve as such a specific temporal measure of physiological state. Its attractiveness for such use is threefold: 1) The phenomenon of VOT as a temporal entity



in the voice is well established; 2) it is relatively easy to measure, via traditional broad band spectrograms or, perhaps more accurately, via digital signal processing. 3) it is an unconscious motor act, characteristic of various age, communication disorder or native language groups. Its ability to characterize the temporary physiological state of an individual is unknown.

#### Null Hypotheses

Therefore the following null hypotheses are to be tested:

1) There will be no significant difference in the variability of VOT within each subject between the sober and intoxicated treatments for /d/ or for /t/.

2) There will be no significant difference in the variability of VOT over time within each subject for /d/ or for /t/.

#### Operational Definitions

Voice Onset Time (VOT)--This is the time between the instant of stop closure release and the onset of glottal vibration, i.e. voicing (Lisker and Abramson, 1964). Spectrographically, release of the stop can be determined by fixing a point of abrupt change in the overall spectrum, sometimes referred to as a plosive spike. The onset of voicing can be marked by the first regularly spaced vertical striations which indicate the pulsing of the glottis. Oscillographically, release is defined



as a sudden and visually discernable change in the overall amplitude with a noted random or aperiodic configuration to the waveform which accompanies a stop release in a time-domain display. The onset of voice is characterized by the initiation of periodicity and an accompanying increase in amplitude representing the following vowel or, in the case of voiced stops, the initiation of voicing.

Physiology or Physiological State Change--By definition this is a change in the chemical, physical or electrical behavior of tissues, cells, or organs of the body brought on by the introduction of some foreign matter or as the result of the body's reaction to an external stimulus. It can be measured via some instrument external to the body. For the purposes of this study, an alcoholic beverage was ingested and the presence of alcohol in the blood was measured via a portable breathalyzer test which noted the degree of such a change in physiology.

Intoxication--The state of Michigan regards 0.100% blood alcohol level as a condition of legal intoxication. The range of 0.080% to 0.100% is considered as a state of impairment. The intoxicated range for this study was between 0.075% and 0.100%.

Spectrographic and Oscillographic Analyses--Visual analyses of the recorded data were accomplished by using digital signal processing. MacSpeech Lab software (Weinreb, 1986) provided frequency-time (spectrographic)



and time domain (oscillographic) output for such analyses. The critical points of VOT (stop release and voice initiation) were readily spotted using this technology. Cursors were placed at these two positions with the time domain values of each cursor available. The time difference between these cursors was also displayed and thus accurately indicated the VOT to within 1 [ms]. Figure 1 depicts two such displays and shows the oscillograms of approximately 0.5 second of ongoing speech, during which an aspirated /t/ phoneme is produced by a male speaker in Figure 1A, and by a female speaker in Figure 1B. The left cursor in each display is at the point of stop closure release. This is made even more evident in each display's upper left enlargement window, where 24 [ms] on either side of the left cursor is displayed. The right cursor is at the point of voicing initiation. This point is enhanced in the upper right enlargement window where, in this case, 24 [ms] on either side of the right cursor is shown. The time elapsed between the cursors (0.057 and 0.065 seconds respectively) is in the lower left corner of each display. The simultaneous spectrogram is also presented and was used for secondary verification only.





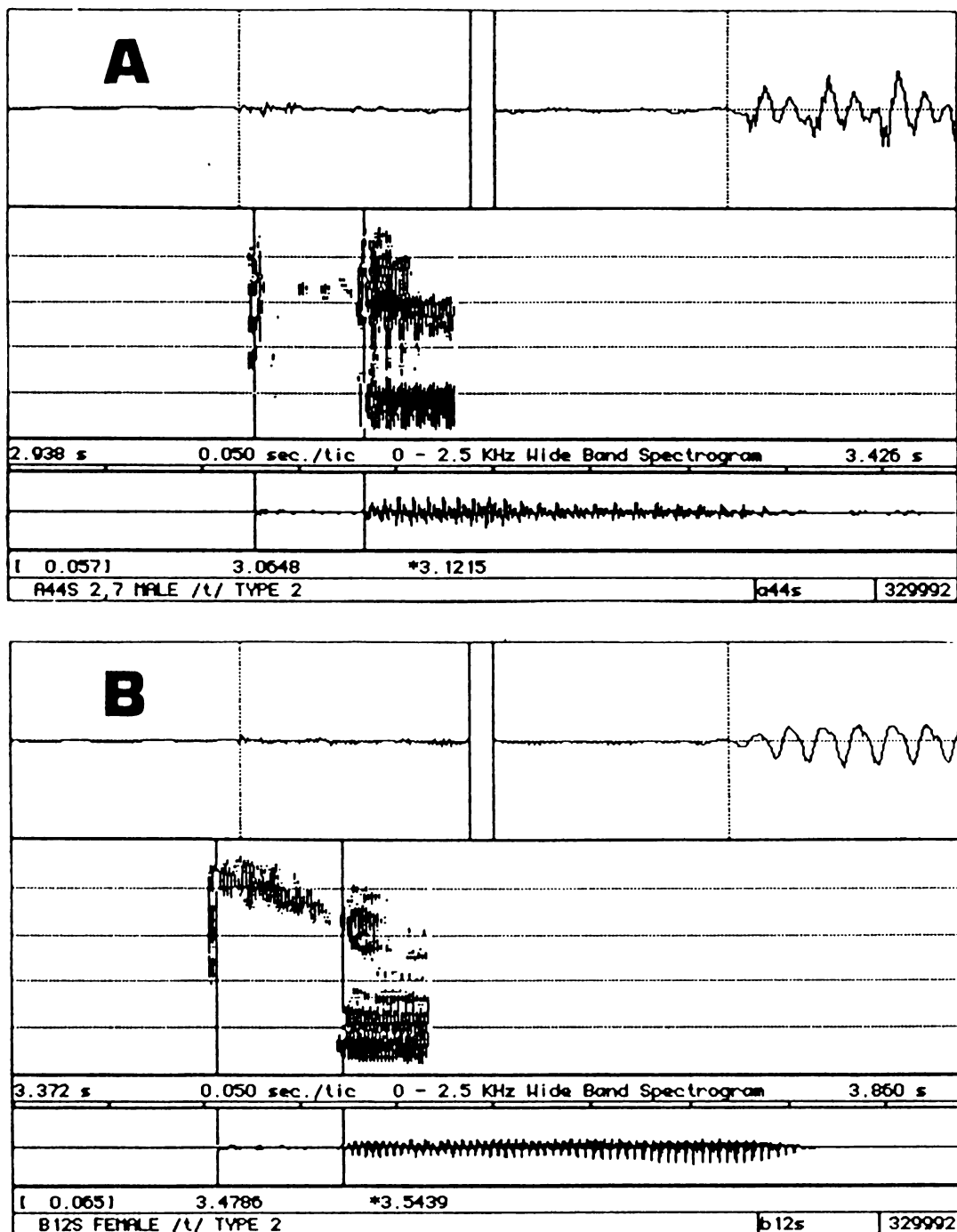


Figure 1. Typical displays used to discern VOT. Display A is that of a male speaker and display B is that of a female. Each display shows the oscillogram near the bottom of the figure, the spectrogram in the center, and the two enlarged oscillograms in the upper enlargement windows.



## REVIEW OF THE LITERATURE

Since 1964 when Lisker and Abramson presented their first report on VOT, there have been approximately 120 articles published on the subject. These articles have focused on VOT vis-a-vis a variety of other topics but generally in terms of communication disorder, foreign language, or subject age. During these past 24 years, only a few articles have appeared which have reported intrasubject variability data, and these were mentioned in the opening chapter. As stated earlier, there were no articles found which looked at variability over time (the constancy of variability, in other words) nor were there articles which considered VOT relative to a change in physiology. Such will be the areas of investigation in the present research.

However there have been articles which have taken issue with the methodology employed in measuring VOT. The suggestions made in these articles were important to the present study's research methodology.

In a follow-up to their 1964 work, Lisker and Abramson (1967) again examined VOT. Here, they evaluated VOT in words versus sentences and in stressed versus unstressed contexts. Within isolated, monosyllabic words they found that all stops possessed distinct ranges of VOT values and that VOT increased as the place of articulation moved posteriorly. The sentence portion of the study used



sentences of eight to fifteen syllables in length. The stops studied occurred as syllable-initial phonemes at various places in the sentences. Here they found a less-clear distinction in VOT values between the voiced and voiceless homorganic stops; that is, there was more overlap. However, there was also less variability than in isolated words. Whether the syllable in question was stressed in the sentences, was not considered the critical factor in determining the different VOTs noted between sentences and in words. Apparently, merely the fact that VOT was occurring in on-going speech caused such a difference to occur. Nor were these notable differences in VOT dependent on the within-sentence position of the sound, unless the word was stressed and occurred as the last word in the sentence. Then the VOT was generally longer. They concluded by stating that the most variability occurred in the /p,t,k/ phonemes. Further, the VOT was greatest in stressed, isolated words. This was partly due, perhaps, to the complications involved in testing for VOT in sentences possessing the /b,d,g/ group of phonemes, when these stops were preceded by a voiced sound such as a vowel or a voiced fricative. In such a situation, there was continuous voicing from the preceding sound through the stop and into the following vowel. The notion of measuring VOT as defined, then became problematic.



It appeared evident that in order to use VOT measures derived from sentences the phonetic context must be controlled. The context should eliminate a voiced sound preceding the stop whether the stop itself is voiced or voiceless. In this way, the reduced variability Lisker and Abramson found in sentences can be achieved, while VOT is being measured in all six stops.

Nearly twenty years later, in a pilot study involving anticipatory coarticulation, Repp (1986) used his two daughters and himself as subjects. Each repeatedly uttered words beginning with /t/ in the carrier phrase "I like the \_\_\_\_". Mean VOTs were computed and showed variations depending on vowel sound to follow the stop. These results lent credence to the notion of the need to control the phonetic environment when measuring VOT. At least the vowel following the stop should be a constant phoneme.

While making an oral presentation to a conference, Weismer (1977) mentioned the effect on VOT of the following vowel sound's tenseness and the post-vowel-consonant voicing in consonant-vowel-consonant (CVC) utterances. Longer VOTs were found for tense vowels and voiced final consonants. This too speaks to the issue of context control. If longer VOTs are required for measurement (as they may be easier to measure), tense vowels and voiced following-consonants should be used.





Weismer (1979) published his report which he spoke of in his 1977 presentation just cited. In this publication, he reported three findings which were of importance to the present study. First, he again stated that the phonetic context of the stop in question was critical to VOT. Specifically, if the following vowel was tense and the post-vocalic consonant was voiced, the VOT tended to be longer. This was done with voiceless stops in a carrier phrase. Second, he reported individual subject means and standard deviations of his admittedly small sample of three subjects. Other studies had reported ranges or isolated cases of intrasubject variance; but with only three subjects he apparently felt freer to report these results more fully. Standard deviations ranged from 8.00 to 16.00 [ms] depending on the individual and the phonetic context in question. Although such large variances were expected in VOT measures, there was no indication of the normalcy of such numbers for each subject, as the trials were not repeated over time. Third, Weismer concluded that the combined effects of vowel tenseness and final consonant voicing (the larger phonetic context in other words) had a complex effect on VOT. The differences he saw between contexts and between individuals were small but never-the-less evident. He finished by stating that physiological variations were perhaps responsible for the differences noted in his data.



Though lacking a research basis, this hint at a physiological impact on VOT, taken with his comments on phonetic environment in VOT measurements, were of primary concern in the preparation of the present work.

Port and Rotunno (1979) also examined VOT in varying phonetic contexts by altering the vowel sound which followed the stop and by changing the manner of production of the consonant which then followed the vowel. All speech samples were monosyllabic words beginning with /p/, /t/ or /k/. Vowels were varied with a constant nasal sound to follow, e.g. "pin" /pIn/ versus "pan" /pæn/. Final consonants were varied with a constant preceding vowel sound, e.g. "tin" /tIn/ versus "tipped" /tIpt/. Results of their study indicated that VOT did change as a result of both vowel quality and as a result of the nature of the final consonant, even though this consonant (or consonant cluster) was two segments removed from the stop. They reported mean times for their subjects but did not report intrasubject variability. Variability would be expected to be high for VOT measures anyway, but especially in light of the fact that single syllable isolated words were used. However, the changes as a result of context gave added direction in making efforts to remove phonetic context variances from the test stimuli.

Klatt (1975) too, found differences in VOT depending on the phonetic environment. He investigated VOT



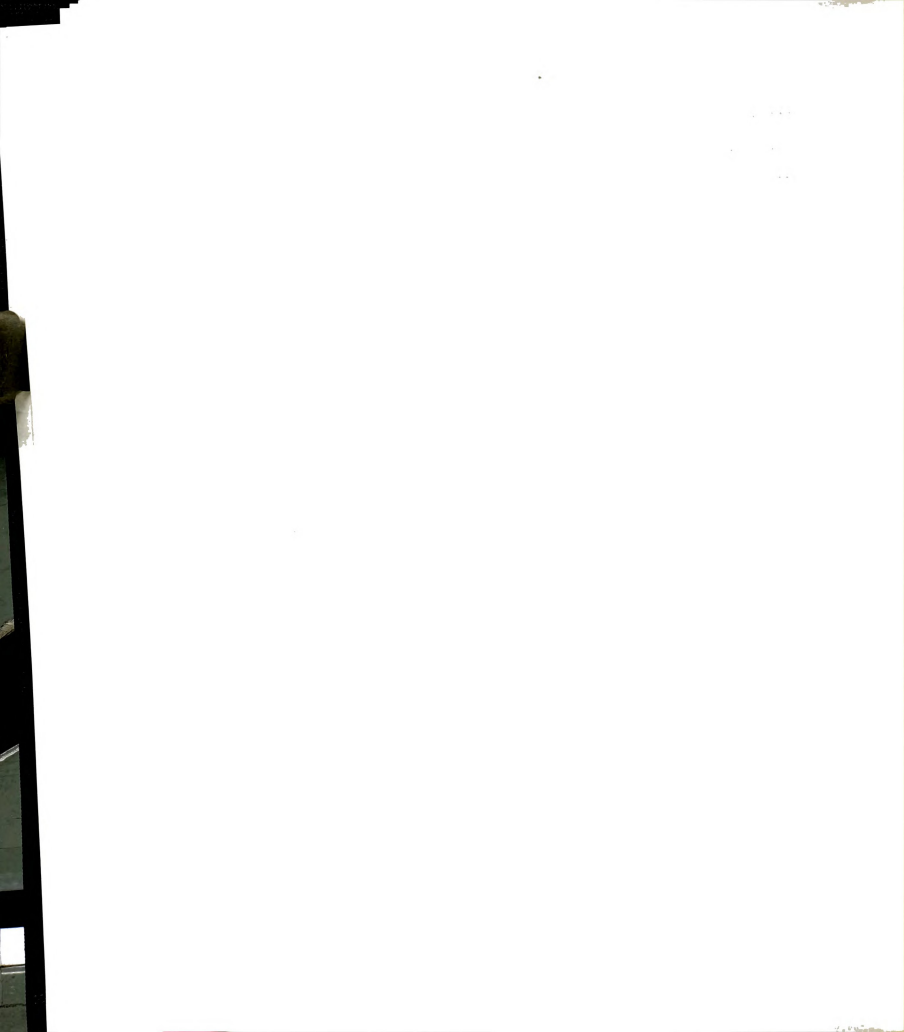
in words where stops were followed by vowels, as was the traditional approach, and in words where the stop was part of a consonant blend. He found shortened VOTs in the blends versus nonclustered consonants. This also argues for context control to eliminate such clustering, even across syllable boundaries. He also reported intrasubject means of VOT for his three subjects, though he did not report intrasubject variability. Since all recordings were done in one session, there were no variations over time nor with changes in physiological state. Standard deviations were summarily mentioned as 5 [ms] for voiced and 11 [ms] for voiceless stops but not in regard to any one individual. Klatt attributed this variability to hand measurement methodology, intrasubject variations, and vowel environment differences. The intrasubject variability was of primary importance to the present study, as it needed to be examined in controlled phonetic contexts.

In an article critical of other work investigating VOT differences in aphasics, Walsh (1983) made some interesting observations on VOT measures in general, not unlike those reported in Klatt (1975). In particular, he noted that at the lower end, VOT may be negative or zero, depending on when the glottal pulsations occur relative to the stop release. A clear delineation of these two events is impossible if the phonetic context permits a voiced sound to precede a voiced plosive. A related issue was the



carrier phrase he found being used in the critiqued studies. The phrase in question was "This \_\_\_\_\_", where the blank was filled with words initiated by various stops. When these stops were voiceless, there was the real possibility that an unaspirated consonant cluster followed the carrier phrase, rather than a voiceless stop. For example, "This top" /ʒIs tap/ becomes "Thi stop" /ʒI stap/. It thus has the same effect as Klatt's (1975) consonant clusters, but rather than intentionally being in the same syllable, there is unintentional resyllabification with a cluster resulting. This can and does affect the VOT of the plosive sound. He concluded by suggesting future researchers carefully define VOT relative to its context, VOT measurement techniques use glottograms as well as spectrograms, and the context be controlled to eliminate the accidental recombining of syllables.

Ohde (1984) looked into fundamental frequency as an acoustic indicator of stop voicing. He compared changes in VOT with changes in fundamental frequency, specifically in carrier phrases. He noted other studies which had found a reduction in VOT values when used in carrier phrases, and he was most interested in the English-phonemically-equivalent voiceless aspirated /p, t, k/ versus voiceless unaspirated /p<sup>̥</sup>, t<sup>̥</sup>, k<sup>̥</sup>/ sounds. These latter ones would be similar to the VOTs found in the clustered stops of Klatt (1975) and Walsh (1983). He found that





these unaspirated stops were nearly identical in VOT to the voiced stops /b,d,g/ (which are always unaspirated in English). This work was of primary importance to the methodology used herein for measuring VOT. Though phrases tended to create shorter and less variable VOTs, the context must remove the possibility of creating voiceless unaspirated stops if they are to be differentiated from voiced stops.

Finally, Zlatin (1974) investigated adult production and perception of VOT in four word-pairs, varying in their initial consonants of /p,b/, /t,d/ and /k,g/. (The /p,b/ cognates were used in two pairs.) She used isolated words, read one at a time by her subjects. She reported significant differences in VOT between voiced and voiceless cognates and between places of articulation within each of the voiced and voiceless groups. This was not surprising, as many earlier studies had reached similar results. But Zlatin went on to report another conclusion. Although there were individual variations in VOT from one phonetic context to another, "...individuals tended to be highly consistent in the range of VOT values for their single word productions." This is taken to mean that individual variations may not be as great as the variations noted in group data. What is normal for such individual variations is open to question and may indeed change over time or with alterations in physiology. Zlatin expressed

1228

1229

1230

a need for studies involving physiological measures, at least involving the preponderance of voicing lead versus voicing lag, in voiced stops noted in some of her subjects. But when her conclusions are coupled with the reduced variability noted by Lisker and Abramson (1967) when on-going speech was used, the determination of intrasubject variability of VOT over time and with changes in physiology, takes on increased importance in the understanding of the complexities of voice onset time.

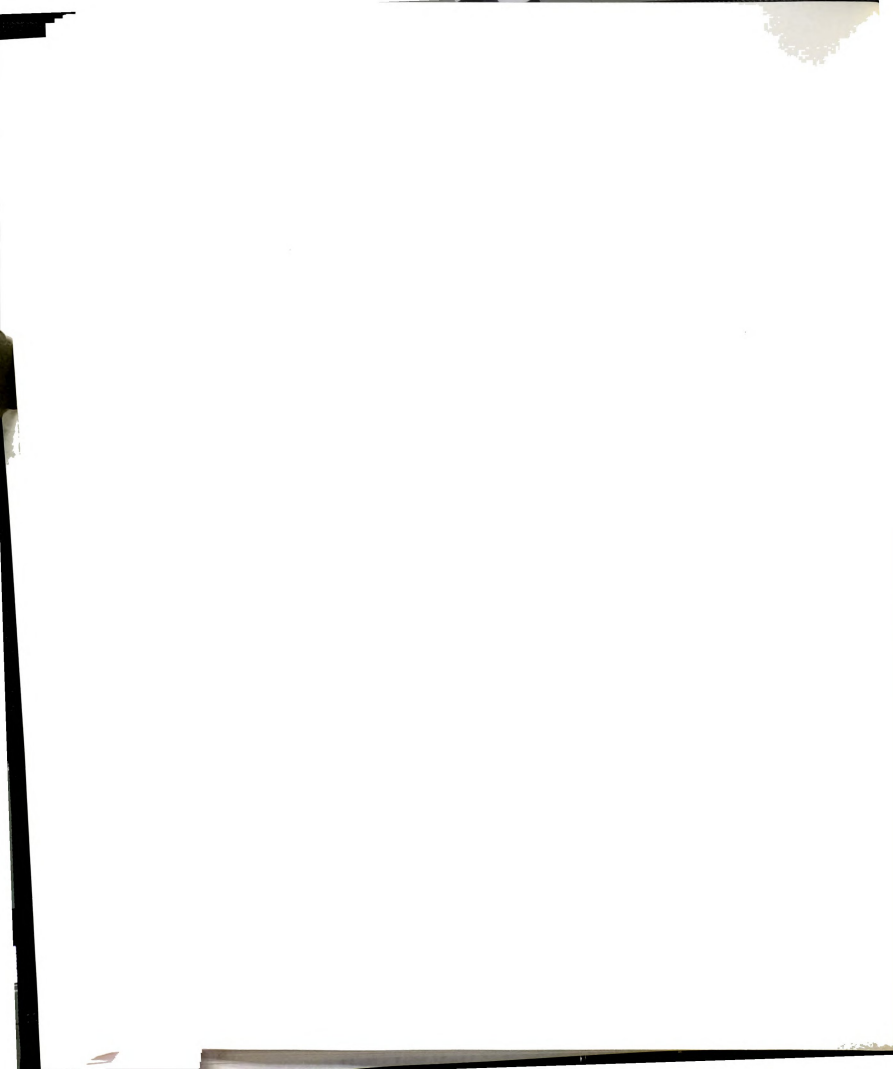


## METHODOLOGY

In order to investigate changes in voice onset time variability over time and the relationship between intrasubject voice onset time variability and changes in physiological state, the following methods were employed.

### Subjects

Sixteen subjects, eight males and eight females between the ages of twenty-one and twenty-six were used. All subjects passed hearing and speech screenings and spoke with a General American dialect. Hearing screening was performed using pure tone audiometry at 25 dBHL, through the frequencies of 500, 1000, 1500, 2000, 3000 and 4000 Hertz bilaterally, in a sound insulated booth. The General American dialect is the mode of speech used by natives or long-term occupants of the American midwest and western states. Use of this dialect, as well as the presence of normal articulation, voice, and rhythm, was determined using the vowel portion of a standard adult articulation test. Appendix A provides the sentences used in this speech screening. Hearing, speech and dialect screenings were performed by the author of the present study, a certified speech-language pathologist. None of the subjects were smokers, dependent on any prescribed medication, diabetic or known or admitted alcoholics. None were affiliated with the professions of communication disorders or sciences, thus avoiding any prior knowledge of



voice onset time. All subjects were occasional drinkers of alcoholic beverages and were willing to drink beer as an intoxicating agent. Subjects were recruited from classes or social groups on campus at Central Michigan University. They were interviewed in a one-on-one setting with the examiner in order to meet this study's requisites. The questions asked can be found in Appendix B. In this interview, all subjects received a verbal summary of the purpose and methods of the study and written general information, which can be seen in Appendix C. Prior to taking part in the study, each subject signed a consent form, which is presented in Appendix D. The subjects did not receive any monetary remuneration for their participation, although they may have received credit for such participation as part of their enrollment in college classes.

#### Voice Onset Time Stimuli

For the speech samples in this study, each subject read phrases which were seven syllables in length and contained a  $C_1C_2VC_3$  combination (where V was the vowel /u/,  $C_1$  was the voiceless fricative /θ/,  $C_2$  was either of the alveolar stops /t/ or /d/ in a syllable-initial position, and  $C_3$  was the lateral /l/). There were five phrases for /t/ and five others for /d/. The ten phrases were printed in four lists, each with a separately randomized order. Each subject said one of the





sets of ten phrases two times during each treatment, generating ten /t/ and ten /d/ tokens for each subject, during each treatment. Appendix E lists the phrases used. There were four treatments for each subject, each held on separate days. Three of the treatments were conducted when the subject was in a state of sobriety and one when in a state of intoxication. Each subject rendered a sober reading of the phrases as treatment 1. The intoxicated reading was counterbalanced among the sixteen subjects from among treatments 2, 3, and 4.

#### Data Collection

Data were collected using a form displayed in Appendix F. This form allowed for the easy and accurate coding of the sixteen subjects (A through P), treatments (1 through 4), counterbalanced versions of the sentences (1 through 4), and physiological state (sober or intoxicated). Also included was the date, beginning and ending time of the treatment, initial and final PBT readings, alcohol content consumed, elapsed time of the treatment, microphone to mouth distance, and microphone input level for each subject. For the sober sessions, the start time was the time at which the recording commenced. The end time was the time at which the subject stopped talking after recording the twenty utterances. For the intoxicated session, the start time was the time at which the subject started consuming the alcoholic beverage. The



end time was the time when the subject stopped talking following the recording.

#### Treatments and Physiological States

A portable breathalyzer test (PBT) was used to determine levels of intoxication. Subjects underwent a PBT test before all sober trials and a reading of 0.00% was required to continue. For the intoxicated session, the actual rate of beer consumption and the total amount of beer consumed varied with each subject, all in an effort to have the subject reach the minimum level of 0.075% blood alcohol level. For instance, it is recognized that the larger a person, the more beer that must be consumed by the person in a given time period to reach some specified level of intoxication. Using a standard police chart which contrasts an individual's weight with the number of 12-ounce beers needed to be consumed in a specified period of time to reach various levels of intoxication, a drinking rate and amount was prescribed for each subject. Should a subject require five 12-ounce beers in order to reach the 0.075% to 0.100% range, he or she was requested to consume this amount in the first 1.25 hours, or at a rate of approximately one beer each 15 minutes. Fifteen to twenty minutes without drinking followed, in order to allow the subject's mouth to be clear of excess residual beer, before the breathalyzer test was performed. This method was quite successful in achieving an intoxication level of between



0.075% and 0.100% at a time of 1.5 hours into the treatment, when the PBT reading was taken. If the subject failed to fall within this PBT range at this time, another PBT was taken a few moments later, so as to allow more alcohol to be absorbed and raise the blood alcohol level. This was repeated until the subject was within the desired range. If the subject was beyond the maximum level for this study, a PBT was taken approximately 1 hour later, after which the peak level of intoxication had presumably been reached and the level had begun to fall. Thus, all intoxicated speech samples were taken when the blood alcohol level was indeed between 0.075% and 0.100% as measured using a portable breathalyzer test. The PBT tests were given by officers of the Central Michigan University Department of Public Safety, who were trained and qualified technicians in the administration of such a test. Following the sober trials, each subject was free to leave. Following the intoxicated trial, the subject remained with the examiner and an assistant of the subject's choosing until a PBT reading indicated the subject's intoxication level was falling and it was considered safe to release the subject. This usually required 2 to 3 hours. The subject was then delivered to his or her residence by the examiner. Appendix G provides a summary of the time required for intoxication, amount of alcohol consumed, and PBT level for the subjects in this study.



### Traditional VOT Measures

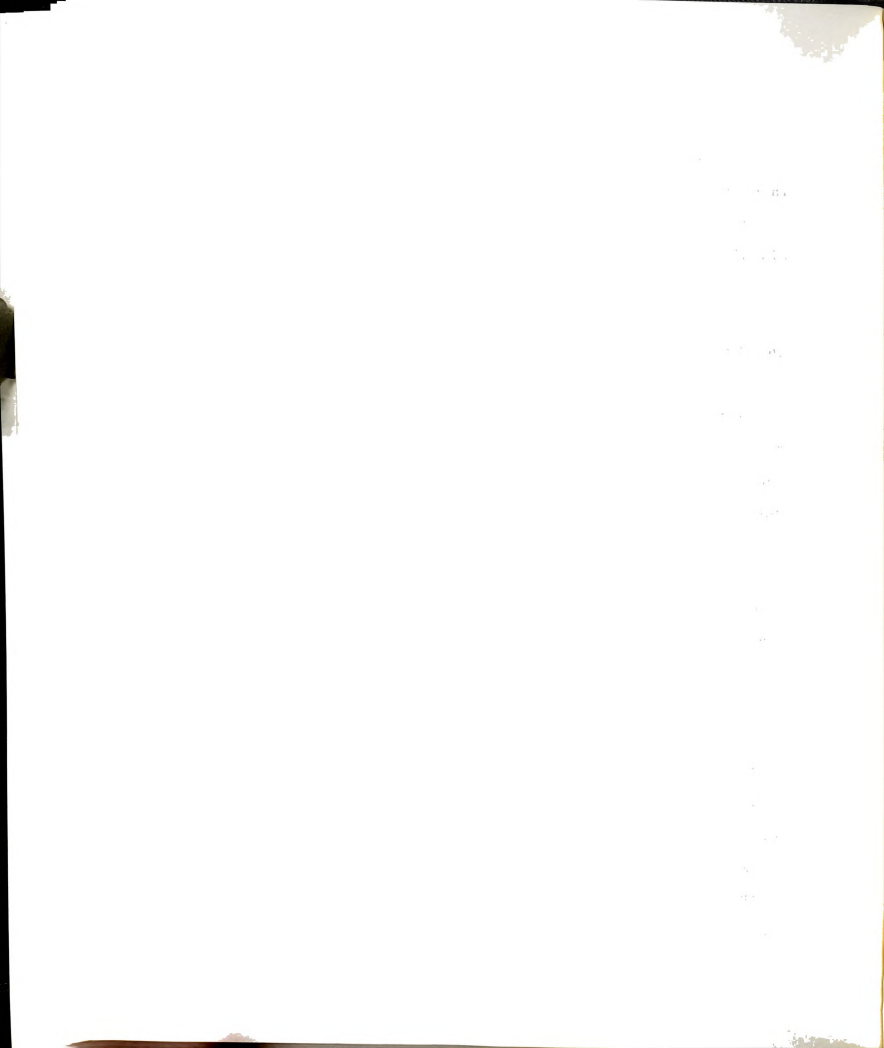
The most common manner of determining VOT, which has been widely used in previous studies, is the visual inspection of spectrograms. These so-called three dimensional displays of the ongoing speech signal reveal frequency on the vertical axis, time on the horizontal axis, and changes in sound intensity via changes in marking darkness on the print. VOT has been computed by making a linear measure of the distance from the instant of stop closure release, to the onset of voicing, and then converting this measure of distance to that of time using an appropriate algorithm. Such "hand measures" have given rise to the accepted means and ranges of VOT as reported in the literature but have been potentially flawed in two particular respects. First, their accuracy has been questionable because of a reliance on linear measurements and conversions to temporal values. Second, the often minimal intensities of the high frequencies occurring during a stop release have not always been evident in the spectrogram. Indeed, the attenuation characteristics of the filters used in such instruments require a signal of sufficient strength in order to be shown as a mark of even minimal grayness on the spectrogram. This potential lack of information was obvious in the digital spectrograms produced in this study and is assumed to have been a possibility in all earlier studies using digital or analog spectrographic instrumentation.





### Digital Signal VOT Analysis

For these reasons, a different approach was taken in this study to determine VOT. This was done through a visual inspection of digitized oscillograms of the speech signal. The MacSpeech Lab software package for speech analysis (Weinreb, 1986) was used to analyze the speech data. This software permitted an oscillographic display of the VOTs of /d/ and /t/ segments, as well as accompanying and simultaneous wide band spectrograms. The spectrogram was used only occasionally for corroboration. The most efficient means available for noting points in the speech sample corresponding to stop release and voicing onset was the use of the oscillographic waveform. The software was installed in an Apple Macintosh SE microcomputer. Input of the speech signal was via a Radio Shack model 33-1063 condenser microphone. This was plugged into a TTE model AFS411 amplifier and filter, which was interfaced to a G. W. Instruments model 411 analog-to-digital converter. The digitized signal was then fed to the computer for display and analysis. Although not used except for the auditory confirmation of the general phonetic environment under examination, a Realistic Minimus 05 speaker provided acoustic output. A digital sampling rate of 5 kHz was chosen to permit the recording of one treatment (twenty sentences, ten each containing /d/ and /t/ for VOT analysis) for each subject on one 800 K-byte microdiskette.



This sampling rate was set in the software and on the filter/amplifier. Although this is a sampling rate lower than what is normally used for speech analysis, the problem of alias frequencies associated with lower than suggested sampling rates was not a difficulty in this study. This was because of the reliance on the oscillogram display rather than the spectrogram. The oscillogram showed the changes indicative of stop release and periodicity onset independent of frequency information, alias or real. It thus permitted sampling rates lower than twice the Nyquist frequency, which is the otherwise accepted minimal level.

The portion of each utterance immediately surrounding each /d/ or /t/ segment was expanded and shown on an oscillogram of 0.5 second duration. This clearly revealed the periodicity of voicing, the aperiodicity of the duration of the release, and the abrupt change in the waveform at the point of stop closure release. Resolution was possible such that each millisecond of the speech sample could be analyzed. To further enhance this stop release characteristic, the portion of the waveform including the release was digitally amplified by a factor of three. Any temporal measure could thus be easily and accurately defined by fixing two cursors and reading the time difference between them. For example, by placing one cursor at the initiation of the release and the other at the initiation of periodicity, the VOT could be determined.



The use of the expanded and amplified oscillograms eliminated the difficulties cited earlier if using spectrographs only. That is, time resolution was obvious and accurate, down to a 1 [ms] level. There was no hand manipulation of linear measures nor conversion from one scale to another. Furthermore, as the oscillogram was a sound pressure wave, stop closure release was evident by apparent changes in the waveform, not by the presence or absence of particular frequencies which may or may not have been reproduced fully on the spectrogram. Thus by using the digital equipment along with a visual inspection of the time surrounding each /t/ or /d/ segment, voicing could be discerned by the apparent periodicity of the oscillographic waveform. The instant of stop closure release could be seen as a small but apparent spike in the wave, followed by an aperiodic wave or aperiodicity overlaid on the periodic component.

#### Types of VOT Measurements

For the /d/ phoneme, three temporal measures were recorded: 1) Pre-stop-release periodicity. This resulted from vocal fold vibration prior to the stop closure release, which has historically been referred to as negative VOT. This was recorded as zero or some value greater than zero, indicating no negative VOT or some measurable amount of a negative component of VOT.

2) Post-voicing-pre-release pause, or PVPRP. This was not



an unexpected phenomenon, but it did occur more often than anticipated, when periodicity preceded the stop release. There was a reduction of the voicing to a level where no continuous waveform was present, in other words an apparent pause, immediately prior to the stop release. In a few instances, the periodicity continued into the release and even through it to the following /u/ vowel. But this did not occur in the great majority of the cases where pre-stop-release periodicity (negative VOT) was evident. This PVPRP was recorded as zero or some value greater than zero, indicating no pause or some measurable pause located between the cessation of negative VOT and the instant of stop closure release. 3) Stop-release-to-vowel-initiation time. This time was measured from the beginning of the oscillographic spike indicating the pressure change accompanying closure release to the obvious change in the waveform where the periodicity of the /u/ vowel was assumed to start. This has historically been referred to as positive VOT. This was recorded as zero or some value greater than zero, indicating simultaneous release and voice onset or some time lapse between release and voice onset. In the event of such a lapse, the aperiodic wave signifying the duration of the release was measured as the positive component of VOT.

For the /t/ phoneme, only the time from the oscillographic spike indicating the stop closure release





until the change in the waveform showing the periodicity of the following /u/ vowel was recorded. As the /t/ phoneme is a voiceless plosive, there was no indication of pre-stop-release periodicity, or PVPRP.

#### Recording Methods

Recording of the subjects' utterances was performed in a quiet location. The subject spoke in a comfortable, natural voice with the microphone placed 3 to 6 inches from the mouth. Consistent voice intensity levels for each subject across the four treatments were maintained by the examiner fixing the microphone input control to a subject-constant level and observing a VU meter during the subject's practice of the sentences. The subject's head was fixed in head cushions on the chair. This arrangement maintained a constant mouth-to-microphone distance. Required VU meter levels in the 1 to 5 volt range assured relatively consistent speech intensity for each subject in each of the four treatments. Subjects had an opportunity to silently and verbally review each list of ten phrases prior to reading them. Subjects were permitted to ask questions regarding the pronunciation of words in each phrase. The CCVC combinations were chosen for several reasons. As stated in the review of the literature, the phonetic context can have a serious impact on the VOT of the stop in question. Specific context issues which needed to be addressed included the tenseness of the vowel sound

11.11.11

11.2

to follow the stop, the voicing of the consonant to follow the vowel, the voicing of the sound to precede the stop, the identity of the preceding consonant, the chances of artificial resyllabification of the stop, and the use of sentences versus isolated words. Therefore, the context of / $\theta$ ,stop,u,l/ was chosen. This context placed a tense vowel and a voiced consonant following the stop, conditions which had been found to increase VOT, making VOT easier to discern and measure. It put a voiceless sound before the stop, allowing for the presence of voicing lead in the voiced stop /d/. It disallowed a syllable consonant cluster incorporating the stop, as / $\theta$ t/ and / $\theta$ d/ are not phonemically compatible or typical English syllable initial clusters. It used sentences as stimuli, a context which had been shown to reduce the variability of VOT. Therefore, for reasons of longer and less variable VOTs, consistency of phonetic environment and ease of determining stop release and vowel voicing, the phrases as listed in Appendix E were chosen.

#### Statistical Analysis

Statistical analyses were performed on the data of each subject individually and on that of the subjects as a group. The independent variables were: subjects, subject sex, treatments, sentence number, and sentence replication group. Blood alcohol as measured using the PBT was used as

12. 02

13. 03

14. 04

15. 05

16. 06

17. 07

18. 08

19. 09

20. 10

21. 11

22. 12

23. 01

24. 02

25. 03

26. 04

27. 05

28. 06

29. 07

30. 08

31. 09

32. 10

33. 11

34. 12

35. 01

36. 02

37. 03

38. 04

39. 05

40. 06

a covariate when comparing the four treatments. Voice onset times for the /d/ and /t/ phonemes were the dependent variables.

Within each subject, specialized analyses of variance were performed to determine whether there were significant differences in the VOT standard deviations (variances) among all four treatments. Judgments were made as to where significant differences occurred, whether among sober treatments or between sober and intoxicated treatments. Because standard deviations were being analyzed instead of group means in the individual statistical analyses, logarithmic transformations were performed on the standard deviations of each individual's five /d/ and five /t/ productions to ensure normality (Ferguson, 1976). These transformations were done on each subject's two sentence replication groups within each treatment.

Analyses of variance of mean data were performed for the subjects as a group. Questions considered included the following. Were there significant differences in VOT means across all subjects: between treatments (with and without the covariate of intoxication level), between sexes, and between sentence replication groups, for both the /t/ and /d/ phonemes? Further, were there significant differences between the five /d/ sentences or between the five /t/ sentences?

1875

1876

1877

1878

1879

1880

1881

1882

1883

1884

1885

1886

1887

1888

1889

1890

1891

1892

1893

1894

1895

1896

1897

1898

1899

1900

Reliability is a measure of consistency and was measured through a test-retest procedure across all the data. As used in this study, it was an indication of the consistency of the methods employed to determine VOT. Over the sixteen subjects, 1280 tokens were collected, 640 each for /d/ and /t/ VOTs. The mean difference in test-retest measures for these 1280 tokens was 1.46 [ms]. For comparison purposes only this translated into a consistency of measure of less than one-third of a fundamental frequency cycle for the females and less than one-fifth of a cycle for the males, assuming average fundamental frequencies of 225 Hz and 110 Hz respectively.





## RESULTS AND DISCUSSION

The results of this study are reported in three major sections. 1) VOT variability was analyzed for both /d/ and /t/ phonemes for each subject across the four treatments, for each sex across the four treatments, for all subjects between the four treatments, for all subjects between the two sentence replication groups, and between the two sexes. 2) VOT mean differences were analyzed for both /d/ and /t/ phonemes across treatments, sentences, sentence group replications and sexes. The effect of intoxication level was investigated by using it as a covariate in the analyses. 3) As a result of the methodology employed in this study VOT was evident in six different forms, rather than the two traditional forms of positive and negative. These are described along with their frequencies of use in each subjects recordings.

For all analyses of variance, the 0.05 confidence level was used as a meter of significance. Although counterbalanced among the subjects during testing, the intoxicated treatment is hereafter given as treatment 4, so as to consistently be presented in the following discussion.

### Differences in Voice Onset Time Variability

A considerable VOT range was expected and was evident, within and across all sixteen subjects, resulting in large standard deviations. Wide ranges of VOTs have been



features of VOT research since the earliest work by Lisker and Abramson. This known intersubject and intrasubject variability of mean VOTs created problems when looking for intrasubject consistency over time, as was the purpose of the present study. A better descriptive measure of individual changes was a comparison of standard deviations. This was consistent with an observation made by Gilbert and Campbell (1978). They stated that in order to study VOT at an individual subject level, the use of data means can be problematic due to the amount of variability. Hence in the present study intrasubject variability in VOT was examined via a statistical analysis of the VOT standard deviations. This was done in order to determine if the VOTs used by each subject were "consistently variable" over time and whether the introduction of a physiological change such as that produced by alcohol consumption would result in a change in this variability. A logarithmic transformation of the standard deviations was necessary to assure normality.

The logarithmic transformations of the standard deviations of each subject--for /d/ and /t/ across all four treatments--are presented in Table 1. The p values are also listed, showing the infrequent points of significant difference in treatment standard deviations. From this table, it is evident that in only two subjects, A and C, did a significant difference in variability occur across



Table 1. Subject log transformations of standard deviations of /d/ and /t/ VOTs across four treatments, and respective p values.

Subject/ Sex	Phoneme	Treatment				p
		1	2	3	4	
A male	d	1.66	2.43	1.74	2.31	.011 *
	t	2.42	2.66	1.35	1.96	.013 *
B female	d	2.15	3.43	1.97	1.83	.285
	t	2.70	2.66	1.97	2.68	.267
C female	d	1.83	2.52	1.21	2.05	.001 *
	t	2.21	2.37	2.76	2.92	.018 *
D male	d	2.42	1.96	2.36	1.96	.837
	t	2.85	2.72	2.33	2.51	.216
E female	d	2.17	2.16	2.00	1.77	.844
	t	2.86	2.69	2.95	2.94	.982
F female	d	1.98	1.36	2.06	2.17	.029 *
	t	2.61	1.95	2.32	2.49	.398
G female	d	1.76	1.87	1.56	1.08	.041 *
	t	2.82	1.76	2.14	2.15	.084
H male	d	1.69	1.68	1.68	2.03	.554
	t	1.95	2.50	1.84	1.54	.308
I male	d	2.09	2.15	2.01	1.96	.864
	t	2.52	2.36	2.65	2.85	.763
J female	d	1.97	1.84	1.97	1.59	.693
	t	2.34	2.18	2.22	2.11	.930
K male	d	1.78	1.64	1.94	0.00	.001 *
	t	1.73	2.08	2.10	1.99	.794
L female	d	2.26	1.83	2.14	2.22	.103
	t	2.49	2.65	2.68	2.64	.961
M male	d	2.27	2.15	2.14	1.21	.621
	t	2.80	1.77	2.38	2.58	.431

\*  $p < 0.05$



Table 1 (cont'd).

---

N	d	1.34	1.29	1.62	1.33	.906
female	t	2.31	1.87	1.76	2.27	.171
O	d	2.02	1.99	1.93	2.51	.457
male	t	2.30	2.39	2.86	2.41	.068
P	d	1.70	1.03	2.32	2.08	.327
male	t	2.56	2.47	2.59	2.48	.918
males	d	1.95	1.88	2.01	1.76	.657
	t	2.39	2.37	2.26	2.29	.851
females	d	1.93	1.91	1.82	1.75	.606
	t	2.54	2.27	2.35	2.52	.220

---

the four treatments for both /t/ and /d/ VOTs. In three other subjects--F, G, and K--the /d/ VOT variances were significantly different over the four treatments. Yet a visual inspection of this table indicates that the differences found, where significant, were not between the intoxicated treatment and the sober treatments. On the contrary, differences in variability across the four treatments were widespread but these differences were seldom significant and quite random as to which one or more of the four treatments were indeed different.

A similar analysis was completed comparing standard deviations of the two sexes rather than individuals, across the four treatments. These results are also in Table 1. They indicate that for both males and females--for both /d/ and /t/ VOTs--there was no





significant difference in variability across the four treatments.

Neither were there significant differences found between sexes (across all four treatments and both replications), between treatments (across both sexes and both replications), or between replications (across all four treatments and both sexes). These data for /d/ and /t/ VOTs are listed in Table 2.

Table 2. Log transformation p values for the entire subject population, across treatments, sexes, and sentence group replications.

	Log Transformation p Values	
	/d/	/t/
Treatments	.484	.460
Sexes	.606	.258
Replications	.595	.505

#### Non-Rejection of the Null Hypotheses

Therefore, for lack of ability to meet statistical significance, the null hypotheses stated in the introduction are not rejected. Those null hypotheses are:

- 1) There is no significant difference in the variability of VOT within each subject between sober and intoxicated treatments for /d/ and /t/.
- 2) There is no significant difference in the variability of VOT over time within each subject for /d/ and /t/.



### Differences in Voice Onset Time Means

To analyze the differences in VOT means, several ANOVAs were performed. Comparisons were made of the means of VOTs for /d/ and /t/, accounting for several independent variables. The independent variables of particular importance were treatment, sex, and sentence number. Table 3 provides the means and standard deviations for /d/ VOTs across these three independent variables. Table 4 shows the data for the /t/ VOTs across the same three independent variables. Table 5 is the source table for /d/ VOTs across sentence replications, sex, treatments, and sentence number. It indicates a significant difference between sexes only. Table 6 is the source table for the /t/ VOTs across sentence replications, sex, treatments, and sentence number. It shows a significant difference between sexes, between treatments, and between sentence numbers. The reasons for the above results can be seen quite clearly by studying Tables 3 and 4. For the /d/ VOTs, there is an obvious difference in the means only between the sexes. Whereas for the /t/ VOTs, there is considerably more difference not only between sexes but between the four treatments and the five sentences. In both Table 5 and Table 6, there is no significant interaction of the variables noted. This supports the notion that the /d/ and /t/ VOTs of both sexes were consistently different across the four treatments and that the /d/ and /t/ VOTs of



Table 3. Means and standard deviations of /d/ VOTs across treatments, sexes, and sentence numbers.

Variable	Mean	Standard Deviation
Population	15.04	8.76
Treatment 1	15.39	8.84
Treatment 2	15.05	8.99
Treatment 3	14.51	7.89
Treatment 4	15.21	9.30
Males	12.88	8.81
Females	17.21	8.16
Sentence 1	15.81	8.65
Sentence 2	13.91	7.77
Sentence 3	16.13	9.21
Sentence 4	14.28	9.48
Sentence 5	15.07	8.50

Table 4. Means and standard deviations of /t/ VOTs across treatments, sexes, and sentence numbers.

Variable	Mean	Standard Deviation
Population	73.79	18.03
Treatment 1	76.06	18.67
Treatment 2	71.39	17.04
Treatment 3	72.26	18.37
Treatment 4	75.45	17.74
Males	65.96	14.53
Females	82.62	16.85
Sentence 1	75.01	20.86
Sentence 2	69.93	17.90
Sentence 3	77.02	17.35
Sentence 4	74.41	17.87
Sentence 5	72.59	15.21



Table 5. Analysis of variance for /d/ VOTs, across sentence group replications, sex, treatments, and sentence number.

Source	SS	DF	MS	F	p
Replication	19.95	1	19.95	.27	.605
Sex	2997.23	1	2997.23	40.29	.000 *
Treatment	69.32	3	23.11	.31	.818
Sentence Number	465.32	4	116.33	1.56	.183
Sex X Treatment	202.03	3	67.34	.91	.438
Sex X Sentence Number	72.17	4	18.04	.24	.914
Within Cells	41660.87	560	74.39		

\*  $p < 0.05$

Table 6. Analysis of variance for /t/ VOTs, across sentence group replications, sex, treatments, and sentence number.

Source	SS	DF	MS	F	p
Replication	640.00	1	640.00	2.53	.122
Sex	49914.23	1	49914.23	197.70	.000 *
Treatment	2562.28	3	854.09	3.38	.018 *
Sentence Number	3669.73	4	917.43	3.63	.006 *
Sex X Treatment	1089.86	3	363.29	1.44	.230
Sex X Sentence Number	811.12	4	202.78	.80	.523
Within Cells	141384.25	560	252.47		

\*  $p < 0.05$

both sexes were consistently different across the five sentences.

In order to contrast the conditions of sobriety and intoxication the three sober treatments were statistically weighted and combined to form a single control group. They were compared to the intoxicated treatment using a t Test. For the /d/ VOTs, there was no





significant difference found ( $p$  value = 0.777). This was consistent with the earlier results listed in Table 5. For the /t/ VOTs, there was also no significant difference found ( $p$  value = 0.178), which was contrary to the results reported in Table 6. In other words, although Table 6 indicated a difference between treatments for the /t/ VOTs, it was not considered a difference between sober and intoxicated treatments. What difference existed was a result of inconsistency across all treatments for the /t/ phoneme as well as the /d/ phoneme.

The next step was to repeat the previous analysis, but with level of intoxication as a covariate. Statistically, the intoxication level was of two distinct ranges, 0% and 0.075% to 0.100%. Table 7 is the source table for /d/ and shows significance between sexes and

Table 7. Analysis of variance for /d/ VOTs, across sentence group replications, sex, treatments, and sentence number, with intoxication level as a covariate.

Source	SS	DF	MS	F	p
Regression	1005.35	1	1005.35	13.82	.000 *
Replications	19.95	1	19.85	.27	.601
Sex	3514.06	1	3514.06	48.32	.000 *
Treatments	1074.18	3	358.06	4.92	.002 *
Sentence Number	465.32	4	116.33	1.60	.173
Sex X Treatment	507.06	3	169.02	2.32	.074
Sex X Sentence Number	72.17	4	18.04	.25	.911
Within Cells	40655.52	559	72.73		

\*  $p < 0.05$



treatments, with a significant covariate regression effect. This intoxication level covariate effect was further analyzed to determine whether the noted regression was linear, as Table 7 would suggest, or quadratic in nature. Table 8 lists the p values for males and females for the regression coefficients, for both linear and quadratic effects. As all values were significant, a curvilinear relationship as illustrated in Figure 2 is suggested. At first glance this purports a difference between sober and intoxicated treatments. But this is not supportable when

Table 8. p values of males and females for linear and quadratic covariate regression effects on /d/ VOTs.

	Linear Effect	Quadratic Effect
Males	.0133 *	.0086 *
Females	.0078 *	.0101 *

\*  $p < 0.05$

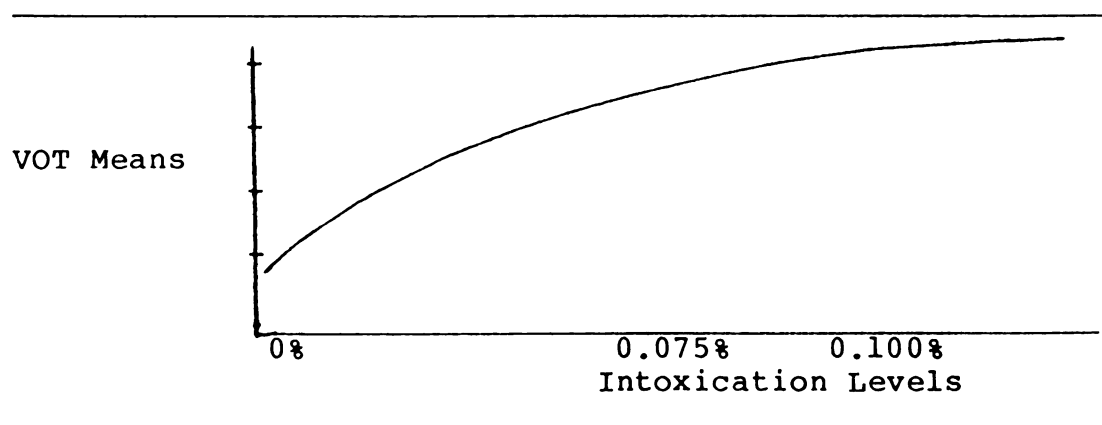


Figure 2. Suggested curvilinear relationship of the covariate effect.



comparing sex by treatment means. Table 9 presents the means and standard deviations of the /d/ as well as the /t/ tokens in a sex by treatment listing. It is obvious upon a visual inspection of the /d/ portion that treatment 4, the intoxicated treatment, was not greatly different from the three sober treatments. Furthermore, the weighted analysis

Table 9. Means and standard deviations of /d/ and /t/ VOTs, broken down by sex and treatment.

Variable	Mean	Standard Deviation
/d/ Population	15.04	8.76
/d/ VOT Males	12.87	8.81
Treatment 1	13.54	9.02
Treatment 2	13.34	8.74
Treatment 3	12.55	8.16
Treatment 4	12.09	9.37
/d/ VOT Females	17.21	8.16
Treatment 1	17.25	8.30
Treatment 2	16.76	8.97
Treatment 3	16.48	7.14
Treatment 4	18.34	8.14
/t/ Population	73.79	18.03
/t/ VOT Males	64.96	14.53
Treatment 1	66.09	15.04
Treatment 2	61.99	12.18
Treatment 3	62.92	14.60
Treatment 4	68.84	15.32
/t/ VOT Females	82.62	16.85
Treatment 1	86.04	16.55
Treatment 2	80.80	16.02
Treatment 3	81.59	17.02
Treatment 4	82.06	17.61



reported earlier supports this lack-of-a-difference observation. Apparently, any variance attributed to level of intoxication was no more than that attributable to sex or perhaps treatment variances. For the /d/ VOTs at least, level of intoxication proved to be an inadequate explanation of noted variances.

Table 10 is the source table for the /t/ VOTs with intoxication level as a covariate. It shows a significant

Table 10. Analysis of variance for /t/ VOTs, across sentence group replications, sex, treatments, and sentence number, with intoxication level as a covariate.

Source	SS	DF	MS	F	p
Regression	74.89	1	74.89	.30	.586
Replications	640.00	1	640.00	2.53	.112
Sex	49157.33	1	49157.33	194.46	.000 *
Treatment	2096.66	3	698.89	2.76	.041 *
Sentence Number	3669.73	4	917.43	3.63	.006 *
Sex X Treatment	860.59	3	286.86	1.13	.334
Sex X Sentence Number	811.12	4	202.78	.80	.524
Within Cells	141309.36	559	252.79		

\*  $p < 0.05$

difference between the sexes, treatments and sentence number, a finding which is consistent with the information in Table 6. There was no regression effect noted here. Indeed, the question of whether linear or quadratic effects were at all evident was disregarded by the regression p values given in Table 11 for males and females. None of these are significant and suggest a fairly random dispersion of the /t/ VOTs. Furthermore, as done earlier





Table 11. p values of males and females for linear and quadratic covariate regression effects on /t/ VOTs.

	Linear Effect	Quadratic Effect
Males	.6414	.8786
Females	.3044	.2863

with the /d/ means, a visual inspection of the /t/ section in Table 9 indicates that treatment 4 was not different from the other three sober ones.

All of these group results, which demonstrated a lack of significant difference in variability over time or between any treatments, were enhanced by the general lack of significant variability differences noted earlier in the intrasubject variability results.

#### Six Variations in Voice Onset Time

The methodology employed in the present study allowed for a detailed analysis of the oscillographic waveform. As a result, six variations in VOT type rather than two, are reported here. A modification of voiced phoneme positive VOT is also suggested. These six variations are listed in Table 12 and reflect the presence or absence of three features of VOT: pre-stop-release periodicity, post-voicing-pre-release pause (PVPRP), and stop-release-to-vowel-initiation time, all in the various combinations seen in this study's recorded data. Where a "0" is listed in the table for each of these, the

100  
100

100  
100

100

100

Table 12. Six possible VOT variation types, respective data codings, and interpretations.

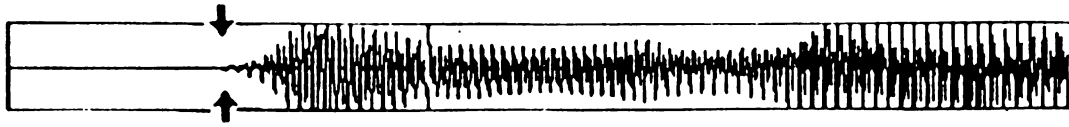
Type	Data Coding	Interpretation
1	0, 0, 0	No negative component, no PVPRP, no positive component. Periodicity begins with the stop closure release.
2	0, 0, >0	No negative component, no PVPRP, positive component. Periodicity begins after the stop closure release.
3	>0, 0, 0	Negative component, no PVPRP, no positive component. Periodicity begins before the stop closure release and continues through it.
4	>0, >0, 0	Negative component, PVPRP, no positive component. Periodicity begins before the stop closure release, stops before it, and resumes with the release.
5	>0, 0, >0	Negative component, no PVPRP, positive component. Periodicity begins before the stop closure release, stops during the release, and resumes after it.
6	>0, >0, >0	Negative component, PVPRP, positive component. Periodicity begins before the stop closure release, stops before it, and resumes after it.



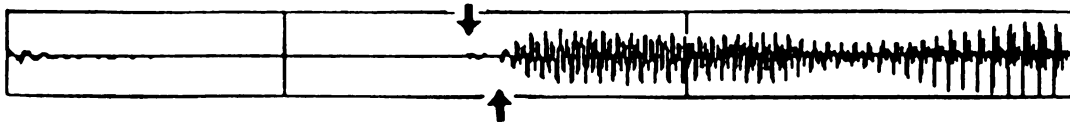
phenomenon did not occur for any measurable time. Otherwise  $>0$  (greater than 0) is listed, meaning the phenomenon was of measurable duration in milliseconds. All six variations were noted among the 640 /d/ tokens analyzed. Only variation number 2 was evident in the /t/ VOTs. The modification in definition for voiced phoneme positive VOT is seen in variations 5 and 6. In both of these situations, there was a measurable time after the stop closure release prior to voicing initiation. But in both instances, there was voicing prior to the release as well. Historically this has been termed negative voice onset time. But to do so was to ignore the cessation of voicing and duration of the release prior to voicing re-initiation. More sensibly, the positive VOT component should be analyzed, just as in variation 2, where no pre-release voicing occurred. Indeed, this positive component proved to be a more reliable and consistent feature than any occasional examples of negative VOT. Furthermore, it was often evident even when a negative component was also present. Figure 3 presents actual oscillographic data depicting these six VOT types. Figure 4 provides idealized versions of these six types to better illustrate them. Arrow markers below the zero line on each figure indicate where voicing initiates or re-initiates. Markers above the oscillograms indicate the points of stop closure release.



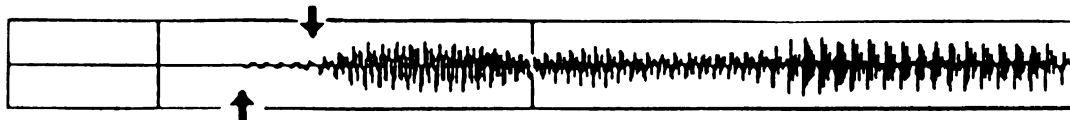
TYPE 1 0, 0, 0



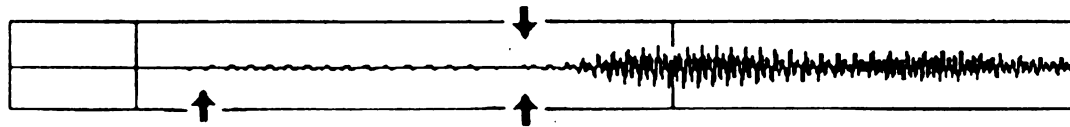
TYPE 2 0, 0, >0



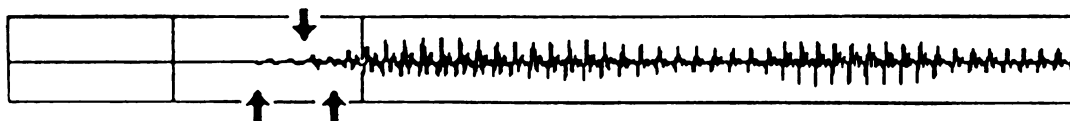
TYPE 3 >0, 0, 0



TYPE 4 >0, >0, 0



TYPE 5 >0, 0, >0



TYPE 6 >0, >0, >0

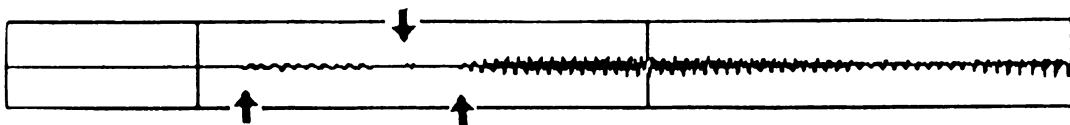


Figure 3. Oscillograms demonstrating the six variations of VOT type. Arrow markers indicate stop closure release (↓) and points of periodicity (↑).





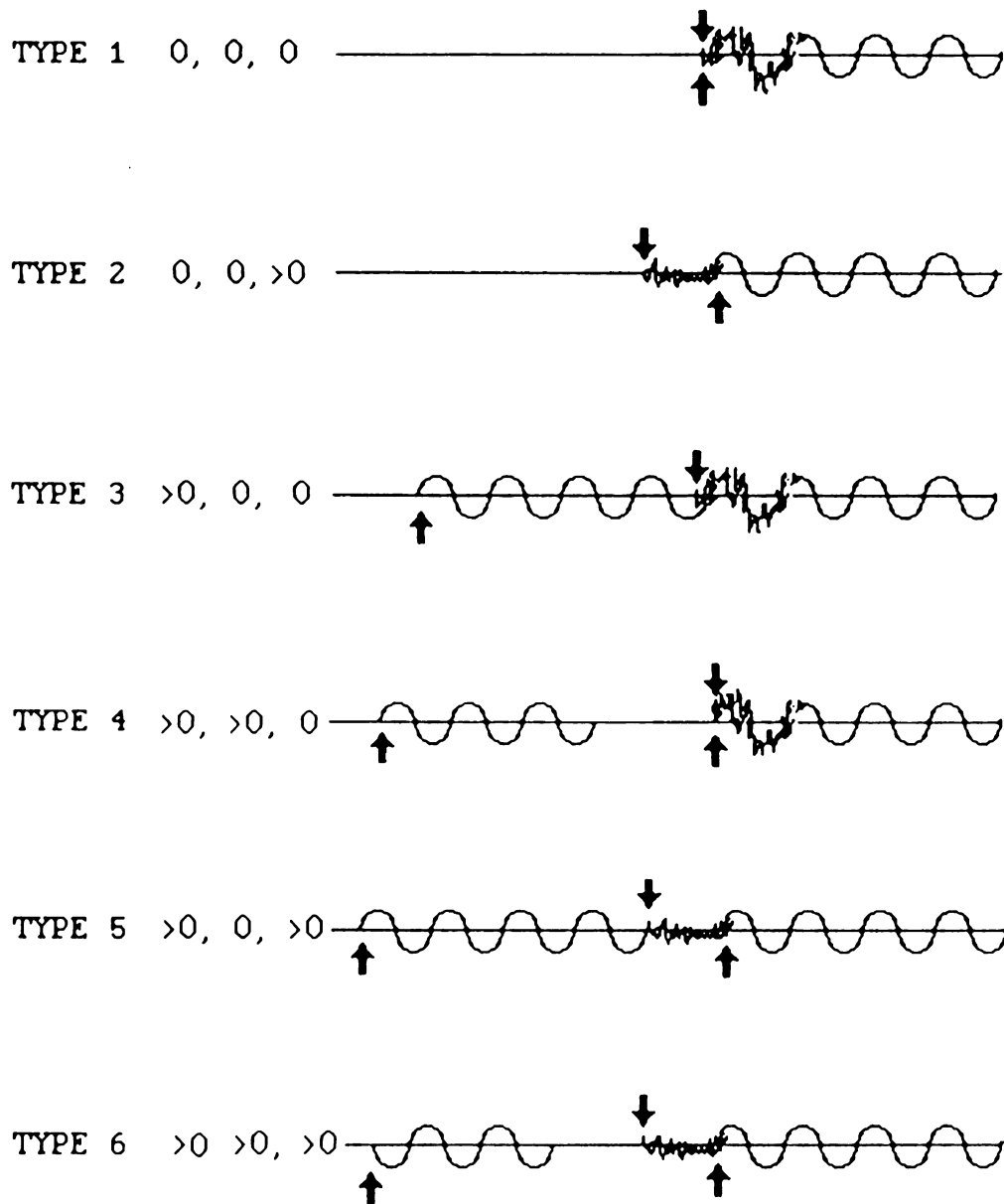


Figure 4. Idealized versions of oscillograms of the six variations of VOT type. Arrow markers indicate stop closure release ( $\downarrow$ ) and points of periodicity ( $\uparrow$ ).



Table 13 lists the frequencies of each type of the /d/ VOTs for each subject in the sober treatments and the percentages of the total for all the subjects these frequencies represent. Table 14 lists similar data for the same VOT types but in the intoxicated treatment. These two tables allow comparison of the sober and intoxicated data. As can be seen in these tables, the most common variations were types 2, 6 and 3, in that order, in both sober and intoxicated treatments. This means 57% of the time sober and 48% of the time intoxicated, there was stop-release-to-vowel-initiation time only (commonly thought of as positive VOT). There was pre-stop-release periodicity, PVPRP, and stop-release-to-vowel-initiation time (where voicing is reinitiated after the release) 25% of the time sober and 32% of the time intoxicated. In only 10% of the time sober and 13% of the time intoxicated did pre-stop-release periodicity only occur (commonly thought of as negative VOT).

Figure 5 is a histogram of the six /d/ VOT types by sex of subject. It indicates, as do Tables 13 and 14, that the three predominant types were variations 2, 6 and 3 respectively, for both sexes. However it also indicates that of those three types, the two with stop-release-to-vowel-initiation time--types 2 and 6--were used more often by women than by men. Furthermore, there was male use of all six types when sober and intoxicated,



Table 13. Frequency of /d/ VOT types (for all sober treatments) for each subject, frequency of each type across all subjects, and percentage of each type across all subjects.

Subject	/d/ VOT Type Sober					
	1	2	3	4	5	6
	0,0,0	0,0,>0	>0,0,0	>0,>0,0	>0,0,>0	>0,>0,>0
A	0	24	1	1	2	2
B	0	21	2	2	2	3
C	0	14	1	0	0	15
D	2	9	7	2	1	9
E	0	1	1	3	1	24
F	0	19	0	0	0	11
G	0	25	0	0	0	5
H	0	26	0	0	0	4
I	0	5	7	1	7	10
J	0	25	0	0	0	5
K	2	25	3	0	0	0
L	0	11	11	0	1	7
M	0	13	2	0	8	7
N	0	30	0	0	0	0
O	0	25	0	0	0	5
P	0	1	13	1	2	13
Frequency of Each Type Across All Subjects						
	4	274	48	10	24	120
Percent of Each Type Across All Subjects						
	1%	57%	10%	2%	5%	25%
Total Sober /d/ VOT Tokens = 480						



Table 14. Frequency of /d/ VOT types (for the intoxicated treatment) for each subject, frequency of each type across all subjects, and percentage of each type across all subjects.

Subject	1	2	/d/ VOT	Type	Intoxicated	6
	0,0,0	0,0,>0	>0,0,0	>0,>0,0	>0,0,>0	>0,>0,>0
A	1	5	2	0	0	2
B*	0	5	1	0	0	3
C	0	5	0	0	1	4
D	0	9	0	0	0	1
E	0	2	0	0	1	7
F	0	4	0	0	0	6
G	0	10	0	0	0	0
H	0	4	0	0	1	5
I	0	1	0	1	1	7
J	0	7	0	0	0	3
K	0	0	10	0	0	0
L	0	6	3	0	1	0
M	1	3	2	2	1	1
N	0	10	0	0	0	0
O	0	5	1	0	0	4
P	0	1	1	0	0	8

\* bad sample, only 9 tokens readable

### Frequency of Each Type Across All Subjects

2	77	20	3	6	51
---	----	----	---	---	----

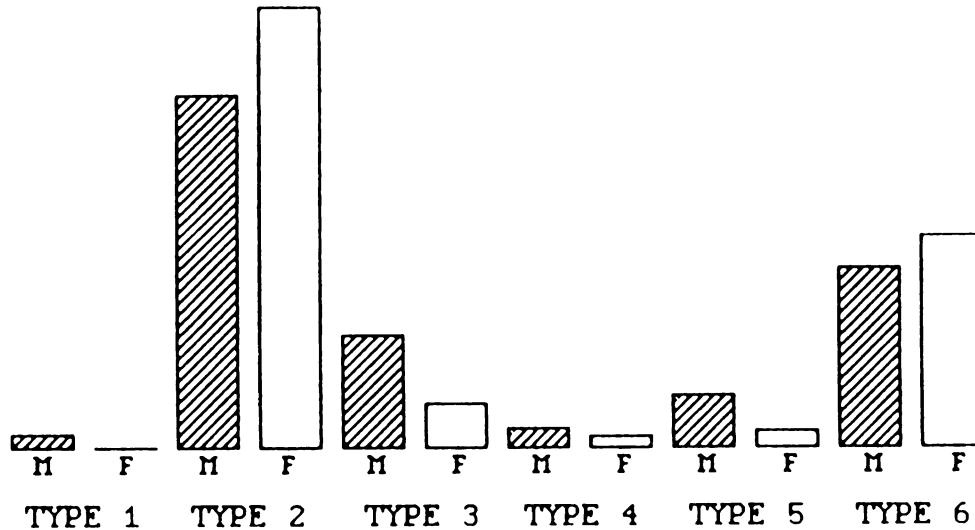
### Percent of Each Type Across All Subjects

1%                      48%                      13%                      2%                      4%                      32%

Total Intoxicated /d/ VOT Tokens = 159








---

Figure 5. Histogram showing relative frequencies of six variations of /d/ VOT types as used by males and females.

whereas no female used type 1 at any time. The use of a greater diversity of VOT types by the males and the preponderance of types 2 and 6 by the females were obvious differences between the sexes. Sex differences were also noted in the analyses of the data means. Since the positive component of VOT was the only VOT feature analyzed statistically, the obvious female reliance on this feature is one explanation of why such a statistical difference occurred.



## Discussion

Voice Onset Time Variability Over Time--There were two principle questions posed by this study. First, is VOT, though known to be variable within and between individuals, at least "consistently variable" in individual subjects' speech? That is, if tested over a series of trials, will the variability of each trial be different from that of the other trials. The results of this study supported the contention that there is no significant difference in intrasubject variability over time, for either the /t/ or /d/ phonemes. There were two subjects, one male and one female, who did demonstrate a significant difference in variability in both phonemes and three others who were significantly variable among trials in the /d/ phoneme. But because of this small number of subjects who reached significance in this test, it is concluded that adult subjects do not typically alter the variability of their VOTs over time. They are, in other words, "consistently variable". As for phoneme-specific tendencies, the greater number of significant differences in the /d/ phoneme indicated there seemed to be a greater likelihood for the /d/ variability to change over time, more so than that of the /t/. But this was also attributable to the greater /t/ VOT variability within treatments, making between-treatments variability harder to differentiate.



Voice Onset Time Variability Due to a Change in Physiological State--The second principle question was this. Assuming there is no significant difference in variability over time for each subject, will there be a change in variability due to some physiological change? The ingestion of an alcoholic beverage provided a measurable physiological change, and the variability of VOTs across the four treatments was considered. However, in this case, a close inspection of the data wherein the few significant differences were found revealed that differences in variability were not due to differences between the fourth (intoxicated) and other treatments. Figure 6 graphically illustrates the treatment standard deviation log transformations for the significantly different intrasubject results given earlier in Table 1. As can be seen, there is no evidence of consistently different variabilities in the intoxicated treatment. In other words, although results indicated two subjects had significantly variable VOTs for /d/ and /t/ and three others showed significance for /d/, there was no suggestion that the differences found were between sober and intoxicated conditions. Any of the four treatments was as likely to affect differences in variability as any other. Therefore it is concluded that the physiological change brought on by the drinking of enough alcohol so as to reach



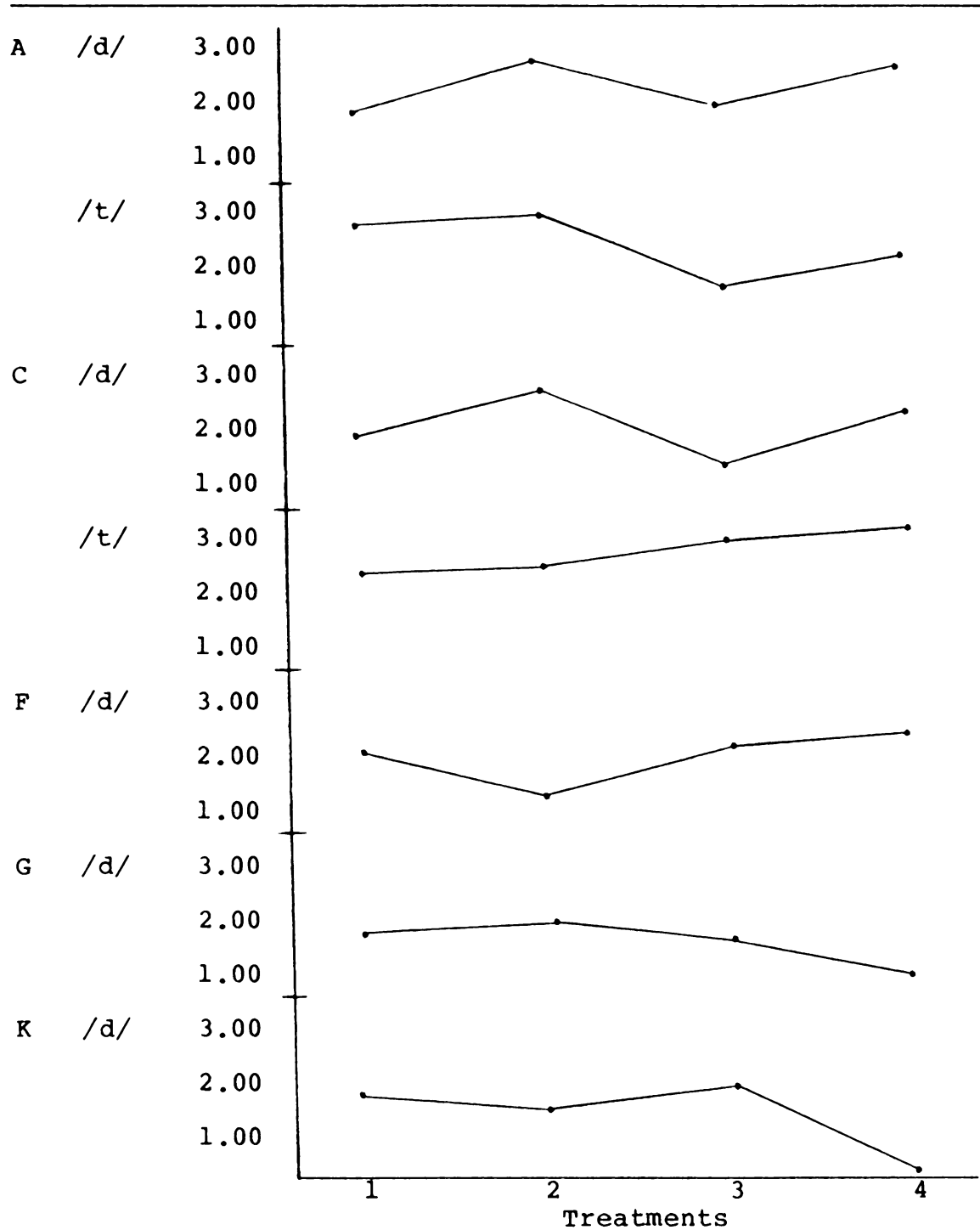


Figure 6. Standard deviation logarithmic transformations across four treatments, for subjects who demonstrated significant difference in VOT variability.





a point of impairment, was not sufficient to produce a change in VOT variability within individual subjects. Or put another way, a change in variability of VOT was not a sufficient indicator of the presence of a blood alcohol level of 0.075% to 0.100%.

Variations in Voice Onset Time Type--The traditional classification of VOT into two categories, positive and negative, was not supported by the methodology used or the results obtained in this study. Positive and negative components did co-exist and may require special consideration.

The presence of the PVPRP which often separated these two positive and negative components, was not surprising. Klatt (1975) stated that during negative VOT the vocal folds may cease vibration just prior to the stop closure release, but normally the voicing continues through the release. What was surprising in the present study was the frequency of this pause phenomenon, thus precluding voicing through the release. Furthermore, a review of the data collection forms revealed that 20% of the sixty-four treatments contained anecdotal notations of inconsistent pre-release-periodicity. It was not unusual for voicing to appear, cease, and reappear during the time between the end of the preceding /θ/ phoneme and the release of the /d/ stop.



The actual temporal values for pre-stop-release periodicity and PVPRP were quite random across and within subjects. Only the positive VOT values, with or without evidence of pre-stop-release periodicity or PVPRP, were relatively stable. Therefore, as did Klatt (1975), only positive VOTs (VOTs greater than or equal to zero) were used in the group or intrasubject statistical analyses. Thus, the smallest possible value for VOT was 0, which was only seen in the /d/ phoneme. This was evident in variations 1, 3 and 4 as seen in Table 12. This condition represented either pre-stop-release periodicity carrying through the stop closure release, or the co-occurrence of voicing with release, whether or not it had occurred before the release of the stop. Variations 2, 5 and 6 were indicative of traditionally positive VOT, although in variations 5 and 6 there was also the co-occurrence of a traditionally negative VOT aspect. Only variation 2 was seen in the /t/ phoneme tokens. Because this is voiceless stop, only a positive component is possible. This fact is consistent with all previous studies.

Sex Differences in Voice Onset Time--The only consistent finding in this study's results was the significant difference in VOT means between males and females, for both /d/ and /t/ phonemes. This is illustrated in Figure 7, where /d/ and /t/ VOT means across the four treatments are shown. As can be seen, there was



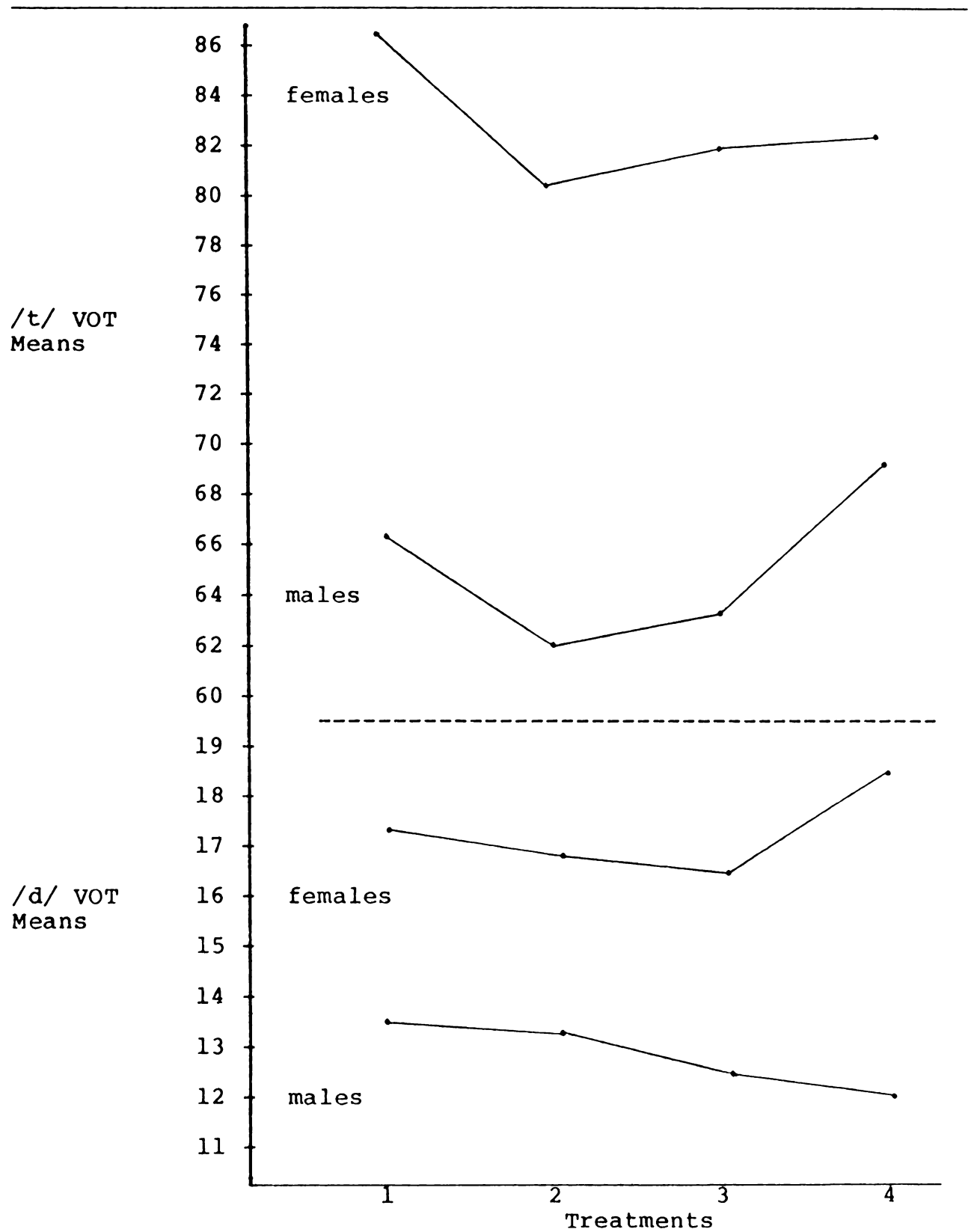


Figure 7. Male and female /d/ and /t/ dOT means across treatments.



no crossing of the lines, which reveals clearly different male and female patterns in the production of voiced and voiceless phoneme VOTs. Results indicated a female reliance on greater than zero VOTs for the /d/ tokens, whereas males used all six types of VOT. This may account for differences in the /d/ phoneme at least from a statistical perspective. But the /t/ tokens also demonstrated a sex difference in the same direction. That is females had significantly greater VOTs for /t/ as well as /d/. Reasons for this sex difference are unknown, but it is a new finding. Previous literature has found VOT differences between native language groups, communicatively disordered groups, and age groups (Lisker and Abramson, 1964; Bond and Korte, 1983; Healey and Gutkin, 1984; Ohde, 1985; Flege and Eefting, 1986; and Robbins, Christensen and Kempster, 1986). These are but a few of the many studies in these areas of interest. Throughout these investigations there has been no mention of a clear demarcation between males and females. As all subjects in the present study were in the same age range and of similar linguistic backgrounds with no communication disorders, there was little on which to speculate as to the reasons for this noted difference. However, upon delving into other VOT research, there is evidence of a known tendency for positive VOT to become less when the point of oral cavity constriction is more anterior, i.e. /p/ and /b/ VOTs





are less than /k/ and /g/. If this reduction in VOT is correlated to a lengthened distance between the point of constriction and the glottis, then otherwise similar groups which are different only in constriction-to-glottis length should be different, with those having the longer length demonstrating the shorter VOTs. Such was the case in this study, where the males, who would be expected to have a greater average supraglottal length, had the shortest VOTs for both /d/ and /t/.

Another feature evident in Figure 7 is the overall linearity of the two sexes within each phoneme. Generally speaking, treatment 1 had the highest mean VOTs for /d/ and /t/ in both sexes. Treatments 2, 3, and 4 were somewhat different, but significantly so only in the /t/ stimuli. An interesting divergence of the male and female means is seen in the /d/ phoneme data. It appeared that when intoxicated (treatment 4), the two groups tended to become even more dissimilar. To a small degree, the opposite effect was noted in the /t/ sound. Because the statistical analyses did not support a difference in VOTs due to changes in physiology, the notable variance was considered not beyond chance probability.

Digital Instrumentation and Methodology--Digital signal processing allowed for the oscillographic waveform to be analyzed millisecond by millisecond. Using these methods, the ease with which a determination could be



made of the initiation of the stop closure release and the advent of periodicity determined the accuracy and consistency of such measures. Since each measure was performed twice on each of two different days, a test-retest scenario was possible. For the great majority of the /d/ or /t/ tokens, the critical points in time were easily located. For a few of them, there was more difficulty, especially in detecting periodicity onset. A discernable, cyclical fluctuation in amplitude above and below the zero line on the oscillogram became the deciding feature. When comparing test to retest data, it was apparent that the retest information was usually very close to that of the original test data. However, it was also the product of over 1200 previous analyses of the /d/ and /t/ tokens done during the test data collection. Therefore, the data finally used for analysis were generally the retest data. In the event there was a difference between test and retest data of more than 10 [ms], a third review of the waveform was made.

The use of visual inspections of digitized speech samples to determine VOT proved to be a consistent and accurate methodology. In particular, reliance on the oscillogram rather than spectrographic information had several advantages. It permitted a saving of computer memory space in digital form as the sampling rate could be below twice the Nyquist frequency, with no loss in temporal



accuracy. It provided a clearer delineation of the VOT measure. The small but visible spike accompanying stop closure release and the sinuous wave representing periodicity were easier to pinpoint than the analogous frequency information in the spectrogram. Programmed algorithms such as that in the software used in this study can allow quick and precise determinations of the VOT duration. Test-retest reliability demonstrated an average closeness-of-measure of 1.46 [ms]. As mentioned earlier, this represents a fraction of a cycle of a subject's fundamental frequency. Since evidence of periodicity required a complete cycle on the waveform, the determination at this degree of consistency of the beginning of such a cycle and its temporal relationship to the stop closure release was considered to be quite acceptable. This was especially so when considering Smith, Hillenbrand, and Ingrassano's (1986) study which compared oscillographic to spectrographic techniques. Although they reported a typical difference of 5 to 10 [ms] in measured VOTs when comparing these two techniques they also regarded this difference as insignificant. A comparison to analog spectrograms was not possible with the present study's data. However 5 to 10 [ms] differences in measurement techniques should not be lightly disregarded. As an average test-retest difference of 1.46 [ms] was possible using the visual inspection of digital information, it was an obviously viable methodology.



Comparisons to Classical Results--Voice onset

time has been regarded in the previous literature as being of two broad types, negative and positive. In previous studies, measurements were made to determine whether voicing or periodicity initiated before or after the stop closure release. In particular, Lisker and Abramson (1964, 1967) made several observations which are worthy of comparison to the present study. First, they noted that in the four English speakers they tested, each tended to use primarily positive or primarily negative VOT in the /d/ phoneme, and each always used positive VOT in the /t/ phoneme. Such individual differences in /d/ VOT type were not random but were thought to be speaker-distinctive, at least in isolated words. Lisker and Abramson gave mean values for the positive and the negative /d/ VOTs separately. These authors stated the VOTs were in two distinct ranges, with virtually no overlap in any one speaker. The present study did not support this observation. For example, as was evident in Tables 13 and 14, only one subject, N, used only one type of VOT consistently, both sober and intoxicated. All of the other subjects used a variety of types, with subject D using all six variations, at least when sober. Second, Lisker and Abramson reported that across their subjects, positive VOT was used 83% of the time, and negative VOT was used 17% of the time. The present study prohibits such a strict

dichotomous comparison, because of the co-existence of positive and negative components of VOT in variations 5 and 6. The reasons why such a dichotomous comparison cannot be made are evident if a more traditional interpretation of these data is used. For example type 2 may be said to represent positive VOT and types 3 and 4 may represent negative VOT (with or without the PVPRP). Under these conditions the percentages are 57% and 48% positive VOT (sober and intoxicated treatments), and 12% and 15% negative VOT (sober and intoxicated). Thus, the existence of positive VOT only was apparently less prevalent than reported by Lisker and Abramson. Yet if the positive component is considered, regardless of the presence or absence of the negative component, as seen in types 2, 5 and 6, then the percentages rise to 87% sober and 84% intoxicated positive VOT which are more in line with Lisker and Abramson's percentages. As another example, the negative VOT percentages of 12% and 15% at first appear to equate with the earlier studies. But if the negative component is considered regardless of the presence or absence of the positive component, as seen in types 3, 4, 5 and 6, then these percentages rise to 43% sober and 57% intoxicated negative VOT. Thus, when making such a comparison, frequencies and percentages of VOT types can appear at once to be similar to, and yet different from earlier studies. Third, Lisker and Abramson found





reduced frequency of negative VOT in sentences versus single words. Although there were no single words here with which to compare, the high percentage of the negative component just mentioned in this study's sentence stimuli does not bear out the earlier observation. Likewise, Lisker and Abramson found reduced VOT variability in sentences when compared to single words. Again such a comparison was impossible here. But a high rather than low degree of variability in sentences is reported here, with a /d/ VOT population standard deviation of 8.76 and a /t/ VOT population standard deviation of 18.03. (These data were presented in Table 9, and demonstrated a pronounced degree of variability in the production of VOT, more so in the /t/ tokens.) Fourth, the observation of more variability in the voiceless /t/ phoneme's VOT was the same as that made by Lisker and Abramson in their studies. Fifth, the present study found virtually no overlap in VOTs within this homorganic pair. Ranges of approximately 12 to 18 [ms] on average for the /d/ and 62 to 86 [ms] on average for the /t/ were quite distinct and remained so across all sixteen subjects. These ranges were in accord with Lisker and Abramson's work, but the fact that they remained so separate was in contrast to the Lisker and Abramson observation of increased VOT overlap between homorganic stops in sentences.



## CONCLUSIONS

A number of conclusions and implications for future work can be drawn from the present study.

### Suprasegmental versus Segmental Differences

The influence of alcohol on the motoric abilities of individuals is well documented. Indeed it is not unusual for law enforcement officers to rely on aberrations in gait, eye-hand coordination, or even speech patterns when making initial observations of a person suspected of being intoxicated. When a person is so intoxicated as to affect his speech, the listener may notice a slowing of rate and some dysarthric-like misarticulations. Temporal changes in rate are suprasegmental whereas the timing feature of VOT is a characteristic of the segment itself. Apparently the segmental features are more resistant to change, even under the severe condition of intoxication, than are the suprasegmental features of speech. VOT resistance to the effects of intoxication force a conclusion that future research into the effects of physiological change on speech should be directed toward temporal features of the voice, such as shimmer or jitter. These vocal variations discount the added variable of supraglottal articulation and will perhaps be more enlightening.



### Changes in Variability versus Changes in Means

One method incorporated in the present study which will be of use in future research into speech or voice variations as a function of physiological change is the use of variability data, rather than mean data. Even such finely variable measures as shimmer and jitter produce intrasubject variations in mean values which can cloud results. To overcome these fluctuations in mean values, it is suggested that analyses of fluctuations in variability be incorporated. It is more important, when attempting to discern an individual's physiological state using variable measures such as those encountered in the voice, that the amount of normal individual variability be established first. Changes in this variability--though not that of VOT--may yet prove to be a reliable, non-invasive indicator of changes in physiological state.

### Sex Differences in Voice Onset Time

Any further investigations of VOT should consider the differences attributable to sex noted in this study. When only one sex of subject is used, investigators should be aware of the male tendency to use a greater variety of VOT types, incorporating more negative components than females. When both sexes are used the VOTs must be kept separated by sex. There were consistently significant differences between the sexes in this study. This result casts doubt on previous research that used male and female



subjects. Future researchers should provide for these differences in their methods and statistical analyses.

#### A Closer Look at the Negative VOT Aspects

The present study only used the positive component of VOT (stop-release-to-vowel-initiation time) for statistical analysis. This was the only one of the three features which was at all stable. The other two features (pre-stop-release periodicity and PVPRP) did not provide usable data for the present study. Yet the possibility of realizing practical results from investigating these two features does exist. Future research into individual and group tendencies regarding the durations and frequencies of occurrence of these two negative components should be conducted.

#### Computer Analyses of the Oscillogram

The analysis of time domain information by computer is becoming an increasingly popular practice. As was mentioned in the introductory chapter, a relatively new procedure has been used recently to measure pauses of all types, even articulatory pauses such as VOT (Tosi, 1974). This methodology defines pause in terms of acoustic energy as a function of time. Blumstein et al. (1980) used a computer controlled analysis of the waveform to determine VOT, although they did not define the program parameters. An approach taken by Guillemin and Nguyen (1984) used waveform data and convolution functions of succeeding time





segments to determine VOT and fundamental frequency. The visual analysis of the digitized data as performed in the present study suggests caution in using these types of measures to determine VOT. The presence of the post-voicing-pre-release-pause as well as the positive component of VOT, both of which can be present in the /d/ phoneme as seen in VOT type 6, can conceivably create a confounding circumstance in the use of pausometry or related analyses. Unless these two features are more clearly distinguished in the time domain measure, the numbers purporting to represent VOT (time of stop closure release to periodicity) could be suspect. Whether both need to be studied separately is open to question, but when the data are analyzed, one must be certain of which feature is being measured.

#### Clinical Implications

The clinical methods of speech and language pathology are in a constant state of flux. As research into the nature and characteristics of speech and language progresses, so too do our methods of treatment. The desk top computer is a device which is having an ever-greater impact on the profession. Such an impact is evident in the present study. The software used to determine VOT in this research was unavailable only three years ago. The changes in this technology over the next three years, or even the next decade will demonstrate an increasing reliance on the



computer by the speech-language pathologist. Such a reliance will be manifest in evaluations, diagnoses and even rehabilitation. To the extent that machines will be making clinical judgments for the clinician, these machines must be programmed to account for the rich diversity in human speech. Examples of this variety are the six variations in VOT outlined in this study, and the apparent differences in VOT between the sexes. This information needs to be added to the "data bank" of human speech characteristics already available.

In conclusion, voice onset time has been the subject of a large body of research for over 20 years. In that time, a vast array of aspects of this temporal phenomenon have been investigated. The revelations of the present study--sex differences in VOT, a variety of types of VOT and the coexistence of positive and negative components, and the lack of significant change in intrasubject variability as a function of time or physiology--add more information to that already accumulated. With the present results in mind, the research should continue.



## APPENDICES



## APPENDIX A

Sentences Used in the Articulation, Voice  
and Rhythm Screening.\*

- 
1. If you sleep you may dream.
  2. Don't sit on that pin.
  3. The spread is bright red.
  4. The cat may scratch.
  5. Some prefer to use a gun.
  6. He had a scar by his left ear.
  7. Is there a clock in the car?
  8. He's quite tall and rather bald.
  9. The shoes will wear out soon.
  10. Please look at this book.
  11. He was frozen in the snow.
  12. There were cows all around.
  13. Please make me a sailboat.
  14. Have some ice cream and apple pie.
  15. That boy makes too much noise.
- 

\*(Templin and Darley, 1969)



## APPENDIX B

## Subject Interview Questionnaire.

NAME: \_\_\_\_\_ AGE: \_\_\_\_\_

PHONE: \_\_\_\_\_ MALE / FEMALE

MAJOR OR PROFESSION: \_\_\_\_\_

- |    |   |      |      |
|----|---|------|------|
| 1  | ARE YOU A CIGARETTE SMOKER?   | YES  | NO   |
| 2  | ARE YOU DIABETIC?   | YES  | NO   |
| 3  | ARE YOU TAKING ANY PRESCRIBED MEDICATION?   | YES  | NO   |
| 4  | ARE YOU TAKING ANY UNPRESCRIBED OR ILLICIT MEDICATION?  | YES  | NO   |
| 5  | HAVE YOU EVER BEEN DIAGNOSED AS ALCOHOLIC OR RECEIVED TREATMENT FOR ALCOHOLISM OR DRUG ABUSE?   | YES  | NO   |
| 6  | HAS EITHER PARENT, ANY GRANDPARENT, OR ANY OF YOUR AUNTS OR UNCLES BEEN DIAGNOSED AS ALCOHOLIC, OR RECEIVED TREATMENT FOR ALCOHOLISM OR DRUG ABUSE? | YES  | NO   |
| 7  | DO YOU DRINK BEER?  | YES  | NO   |
| 8  | HAVE YOU EVER BEEN INTOXICATED?   | YES  | NO   |
|    | 9 WHEN WAS THE LAST TIME? _____   |      |      |
|    | 10 HOW OFTEN DO YOU DRINK TO INTOXICATION? _____  |      |      |
| 11 | HAVE YOU EVER FELT YOU SHOULD CUT DOWN ON YOUR DRINKING?  | YES  | NO   |
| 12 | HAVE PEOPLE ANNOYED YOU BY CRITICIZING YOUR DRINKING?   | YES  | NO   |
| 13 | HAVE YOU EVER FELT BAD OR GUILTY ABOUT YOUR DRINKING?   | YES  | NO   |
| 14 | HAVE YOU EVER HAD A DRINK FIRST THING IN THE MORNING TO STEADY YOUR NERVES OR TO GET RID OF A HANGOVER?   | YES  | NO   |
| 15 | DO YOU HAVE A HEARING LOSS?   | YES  | NO   |
| 16 | DO YOU HAVE A SPEECH DISORDER?  | YES  | NO   |
| 17 | WHERE WERE YOU BORN? _____  |      |      |
| 18 | WHAT IS YOUR PRESENT HOMETOWN? _____  |      |      |
|    | 19 HOW LONG HAVE YOU LIVED THERE? _____   |      |      |
| 20 | DO YOU HAVE HEALTH INSURANCE?   | YES  | NO   |
|    | 21 COMPANY: _____   |      |      |
| 22 | SPEECH/VOICE SCREENING  | PASS | FAIL |
| 23 | HEARING SCREENING   | PASS | FAIL |



## APPENDIX C

General Information Provided to the Subjects  
at the Interview.

---

As a subject in this research project, your speech will be evaluated for features of the voice while you are sober and while you are intoxicated. Your voice will be recorded in digital form on a computer disc, as you read 10 prepared sentences. There will be 3 sober readings, each done on a different day, and 1 intoxicated reading. The examiner will provide beer to drink for the intoxicated reading and intoxication will be determined by a portable breathalyzer test (PBT). You will drink enough beer to have a blood alcohol level of between .075% and .100% as indicated by the PBT. Following the sober readings you will be free to leave the experiment site. You must bring a friend with you on the day of the intoxicated reading, who will remain with you under the supervision of the researcher. You will later be returned to your residence by the examiner.

You may review the sentences quietly before reading them, and ask any questions about the pronunciation of the words or where stress is to be placed in the sentences. When ready, you will read them in a normal, comfortable voice. The sober readings will require approximately 30 minutes of your time. The intoxicated reading will require approximately 2 to 4 hours, for both you and the friend who accompanies you.

Information collected during this study will be used to give us a better understanding of normal and abnormal voice functioning, although the procedure itself may not have any direct benefit to you per se. You are not guaranteed any benefits. There are no known or anticipated risks or hazards associated with this project other than those associated with the drinking of an alcoholic beverage. The examiner and your friend accompanying you will remain with you until you are again sober in order to reduce these risks. Information collected during this study may be used for publication purposes, but your name will not be used in any publication or correspondences, nor will it be possible to identify you in publications or correspondences. You are free to withdraw from the project at any time without prejudice to you.

Please feel free to express any questions or concerns you may have pertaining to the project outlined above.



## APPENDIX D

## Informed Consent Form.

---

I, \_\_\_\_\_ freely and voluntarily consent to serve as a subject in a scientific study of the human voice wherein I will provide speech samples while I am sober and while I am intoxicated. The study is being conducted by Bradford L. Swartz.

I understand that the purpose of this study is to determine what changes occur in unconscious features of the voice while intoxicated.

I understand that I will not be exposed to any experimental conditions which constitute a threat to my hearing or speech, nor to my physical or psychological well-being. In one of the treatment conditions, I will become intoxicated by drinking beer, as measured by a portable breathalyzer. I understand that if I am injured as a result of my participation in this research project, Michigan State University and Central Michigan University will provide for emergency medical care if necessary (arrange for transportation to an emergency medical care facility), but these and any other medical expenses must be paid from my own health insurance program. If there are any questions, please contact Bradford L. Swartz, at 774-7296.

I understand that data gathered from me for this experiment are confidential, that no information uniquely identified with me will be made available to other persons or agencies, and that any publication of the results of this study will maintain anonymity.

I engage in this study on my own free will, with no payment to me for my personal time, and without implication of personal benefit from the experiment. I understand that I may cease participation in the study at any time without prejudice to me.

I have had the opportunity to ask questions about the nature and purpose of the study, and I have been provided with a copy of this written informed consent form. I understand that upon completion of the study, and at my request, I may obtain additional explanation about the study.

SIGNED: \_\_\_\_\_ DATE: \_\_\_\_\_

WITNESS: \_\_\_\_\_



## APPENDIX E

Sentences Used for the Collection of VOT Data,  
and Their Phonetic Transcriptions.

- 1    What did Beth do last weekend?  
     /wət dɪd beθ du læst wɪkənd/
- 2    It's not beneath Dooley's chair.  
     /ɪts nɒt biːniθ duːliz tʃeə/
- 3    I saw a moth do long flights?  
     /aɪ sɔ ə mɒθ du lɒŋ flaɪts/
- 4    Has Faith Doolan been here yet?  
     /hæz feəθ duːlən biːn hɪə jet/
- 5    It was the ninth due last week.  
     /ɪt wəz ðə naɪnθ du læst wɪk/
- 6    He said earth tools are needed.  
     /hi sed əθ tuːlz ɑː niːdɪd/
- 7    The fourth tulip is yellow.  
     /ðə fɔːθ tuːlɪp ɪz jelo/
- 8    I've lived up north too long now.  
     /aɪv lɪvd əp nɔːθ tu lɒŋ naʊ/
- 9    You said Heath Tulidge is here?  
     /ju sed hiθ tuːlɪdʒ ɪz hɪə/
- 10   This month two little boys go.  
     /ðɪs mənθ tu lɪf! bɔɪz go/





## APPENDIX F

## Voice Onset Time Data Collection Form.

		/d/	/t/
SUBJECT	_____	1 _____, _____, _____	
TREATMENT	_____	2 _____, _____, _____	
VERSION #	_____	3 _____, _____, _____	
STATE (I or S)	_____		4 _____
SOBER # _____ of 3			5 _____
DATE	____/____/____		6 _____
TIME START	____:____	7 _____, _____, _____	
PBT START	_____ %		8 _____
TIME END	____:____		9 _____
PBT END	_____ %	10 _____, _____, _____	
ALCOHOL CONSUMED	_____ oz	11 _____, _____, _____	
TIME ELAPSED	____:____	12 _____, _____, _____	
MIC INPUT LEVEL	_____	13 _____, _____, _____	
MIC-TO-MOUTH	_____ "		14 _____
			15 _____
			16 _____
		17 _____, _____, _____	
			18 _____
			19 _____
		20 _____, _____, _____	



## APPENDIX G

Time Required for Intoxicated Treatment, Alcohol Consumed,  
and PBT Level for Each Subject.

Subject	Time Required for Intoxication in Hours:Minutes	Alcohol Consumed in Ounces	PBT Level in Percent
A	1:49	56	0.092
B	1:48	52	0.088
C	1:36	40	0.078
D	1:42	59	0.077
E	1:33	45	0.086
F	1:36	48	0.099
G	1:18	43	0.097
H	1:41	57	0.076
I	2:46	43	0.090
J	2:06	59	0.092
K	4:27	53	0.096
L	2:14	48	0.090
M	1:55	60	0.075
N	3:09	48	0.084
O	1:45	53	0.075
P	2:01	71	0.092
$\bar{X}$	2:05	52.19	0.087



## LIST OF REFERENCES



## LIST OF REFERENCES

- Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: a voice-onset time analysis. Brain and Language, 9(2), 153-170.
- Bond, Z. S., & Korte, S. S. (1983). Children's spontaneous and imitative speech: an acoustic-phonetic analysis. Journal of Speech and Hearing Research, 26(3), 464-467.
- Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. Acta-Oto-Laryngologica, Supplementum 257, 5-43.
- Ferguson, G. A. (1976). Statistical Analysis in Psychology and Education. New York: McGraw-Hill.
- Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. Phonetica, 43(4), 155-171.
- Gilbert, H. R., & Campbell, M. I. (1978). Voice onset time in speech of hearing impaired individuals. Folia Phoniatica, 30(1), 67-81.
- Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. Language and Speech, 1, 226-231.
- Goldman-Eisler, F. (1961). The distribution of pause durations in speech. Language and Speech, 4, 232-237.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. Language and Speech, 15, 103-113.
- Guillemin, B. J., & Nguyen, D. T. (1984). Microprocessor based speech processing system. Journal of Speech and Hearing Research, 27(2), 311-317.
- Healey, E. C., & Gutkin, B. (1984). Analysis of stutterers' voice onset time and fundamental frequency contours during fluency. Journal of Speech and Hearing Research, 27(2), 219-225.
- Iwata, S., & Von Leden, H. (1970). Pitch perturbations in normal and pathologic voices. Folia Phoniatica, 22, 413-424.





- Kitajima, K., & Gould, W. J. (1976). Vocal shimmer in sustained phonation of normal and pathologic voices. Annals of Otolaryngology, 85, 377-381.
- Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 18(4), 686-706.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word, 20, 384-422.
- Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. Language and Speech, 10(1), 1-28.
- MacKay, I. (1987). The Science of Speech Production. Boston: Little Brown.
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: study of voice onset time in word-initial stop consonants. Journal of Child Language, 7(1), 41-74.
- Nakasone, H. (1979). Cross cultural study on pause characteristics in on-going speech by U.S.A. and Japanese students. Unpublished master's thesis, Michigan State University, East Lansing.
- Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. Journal of the Acoustical Society of America, 75(1), 224-230.
- Ohde, R. N. (1985). Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. Journal of the Acoustical Society of America, 78(5), 1554-1561.
- Port, R. F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. Journal of the Acoustical Society of America, 66(3), 654-662.
- Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. Journal of the Acoustical Society of America, 79(5), 1616-1619.
- Robbins, J., Christensen, J., & Kempster, G. (1986). Characteristics of speech production after tracheoesophageal puncture: voice onset time and vowel duration. Journal of Speech and Hearing Research, 29(4), 499-504.



- Smith, B. L., Hillenbrand, J., & Ingrassano, D. (1986). A comparison of temporal measures of speech using spectrograms and digital oscillograms, 29(2), 270-274.
- Sweeting, P. M., & Baken, R. J. (1982). Voice onset time in normal aged population. Journal of Speech and Hearing Research, 25(1), 129-134.
- Templin, M. C., & Darley, F. L. (1969). The Templin-Darley Tests of Articulation. Iowa City: University of Iowa.
- Thomas, C. L. (editor) (1970). Taber's Cyclopedic Medical Dictionary. Philadelphia: F. A. Davis.
- Tiffany, W. R., & Carrell, J. (1977). Phonetics: Theory and Application (2nd ed.). New York: McGraw-Hill.
- Till, J. A., & Stivers, D. K. (1981). Instrumentation and validity for direct-readout voice onset time measurement. Journal of Communication Disorders, 14(6), 507-512.
- Tosi, O. (1974). Pausometry: measurement of low levels of acoustic energy. In World Papers in Phonetics, Phonetics Society of Japan (pp. 129-144). Tokyo: The Phonetic Society of Japan.
- Walsh, T. (1983). Voice onset time as a clue to the nature of Broca speech errors. Brain and Language, 19(2), 357-363.
- Weismer, G. (1977). Some context effects on VOT. Proceeds of Meeting, Journal of the Acoustical Society of America, 62(S1), S78
- Weiser, G. (1979). Sensitivity of VOT measures to certain segmental features in speech production. Journal of Phonetics, 7(2), 197-204.
- Weinreb, G. (1986). MacSpeech Lab 2.0. Cambridge, MA.: GW Instruments.
- Zlatin, M. A. (1974). Voicint contrast: perceptual and productive voice onset time characteristics of adults. Journal of the Acoustical Society of America, 56(3), 981-984.



MICHIGAN STATE UNIV. LIBRARIES



31293005694520