171

0-315-12485-7

60396

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Thomas T. Welz

Date of Birth — Date de naissance

Dec 5 1956

Country of Birth — Lieu de naissance

Canada

Permanent Address — Résidence fixe

9305-168st
Edmonton, Alberta

Title of Thesis — Titre de la thèse

Categorization and Discrimination of WORD JUNCTURE IN ENGLISH

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Masters of Science

Year this degree conferred — Année d'obtention de ce grade

Fall 1982

Name of Supervisor — Nom du directeur de thèse

Dr. J. Hogan

Date

Oct 14 1982

Signature

Tom Welz

THE UNIVERSITY OF ALBERTA

CATEGORIZATION AND DISCRIMINATION OF WORD JUNCTURE IN ENGLISH

by

(C) THOMAS T. WELZ

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

IN

SPEECH PRODUCTION AND PERCEPTION

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

FALL 1982

Microfilming

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR THOMAS T. WELZ

TITLE OF THESIS CATEGORIZATION AND DISCRIMINATION OF WORD JUNCTURE IN ENGLISH

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED FALL 1982

(SIGNED) Tom Welz

PERMANENT ADDRESS:

9305 - 168 st

Edmonton

Alberta

DATED October 14 1982

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled CATEGORIZATION AND DISCRIMINATION OF WORD JUNCTURE IN ENGLISH submitted by THOMAS T. WELZ in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE in SPEECH PRODUCTION AND PERCEPTION.

John T. Hogan
Supervisor

....

....

....

Date June 25, 1982

ABSTRACT

Experimental studies have shown that juncture perception can be influenced by subphonemic cues. How might perception in this situation differ from the perception of phonemic differences? One way of exploring these differences is by using the experimental techniques used for testing categorical perception, where identification and discrimination tasks are conducted on a perceptual continuum. This study compares the perception between phonemic and subphonemic distinctions using mainly the VOT continuum and occasionally some durational cues. The results seem to depend mainly on how perceptually salient the cue itself is, independent (to a large degree) of weather it is cueing a phonemic or subphonemic difference. In the perceptual experiments (and also in the measurement of production data) aspiration showed up as a very strong cue, often showing discrimination well above what would be predicted from identification, while prevoicing was a weak cue as were the durational cues.

ACKNOWLEDGEMENTS

From the start to the completion, this thesis owes itself to numerous factors from my envirnoment, too many to count. However number of individuals stand out, a few of which I would like to mention here.

First, I would like to thank my supervisor Dr. J. T. Hogan and my previous supervisor Dr. T. M. Nearey. It was their interest and participation that played the major role in helping me along in my work. I would also like to acknowledge the members of my comittee Dr. W. J. Baker, Dr. A. J. Rozsypal (both for their criticisms of the stylistics, logic, and statistics of the thesis), and Dr. G. Holdgrafer.

I must also give credit to the departmental technician Allen Opperthauser, who guided me out of numerous computer and instrument related problems. People from around the department (usually fellow graduate students) also offered important aid in certain aspects of this thesis, and I thank them.

A special thanks goes out to all of those who participated as speakers and listeners in the experiments run in this study. Specifically, would like to mention those who took part in experiments III and IV, where the tasks involved were particularly long and tedious. Their names are Jerry Kubina, Kathy Look, Gary Okrainic, Devinder Pannu, and Maximilian Welz. They volunteered without the promise of material reward often given in such situations.

Finally I would like to thank my parents for their consideration and help while I was working on the thesis.

Table of Contents

Chapter Page

LIST OF TABLES

LIST OF FIGURES

1. Introduction

In a cross-language study of the voicing distinction in initial stops, Lisker and Abramson (1964) found that measurements of voice onset time (VOT) are distributed on a continuum into three major areas. The first situation is where voicing begins before the release of the burst and it is this type of stop that is described as being voiced. In the second case voicing begins simultaneously or just briefly after the release of the burst, and this is called devoiced or unaspirated. The third case does not have voicing until long after the burst, and is described as aspirated.

In traditional descriptive phonetics the voicing distinction in English was usually portrayed in terms of three allophones: voiced, unaspirated, and aspirated. The unaspirated stop is in complementary distribution with both the voiced and aspirated stops but the aspirated and unaspirated stops are grouped together as allophones of the voiceless phoneme with the voiced stop representing the voiced phoneme. However, voicing lead rarely occurs in English initial stops unless being carried through from the voiced segment immediately previous (Lisker and Abramson, 1964; and Zatlin, 1974). Though most initial voiced stops are actually devoiced, there are still some (Ladefoged, 1971) who describe these as being at least 'partially voiced'. In an experiment using tape splicing methods, Lotz et al. (1960) looked at the unaspirated stops of /s/+stop consonant clusters. When they removed the /s/ portion from a word like 'spin', they found that the stop was overwhelmingly identified with the /b/ phoneme rather than the /p/ phoneme. However, listeners from languages which have a clear case of the prevoiced category (Spanish and Thai), described the same stop as being devoiced. So in terms of phonetic similarity, the English unaspirated /p/ of 'spin' is more similar to /b/ than aspirated /p/, at least as far as the burst and VOT goes. Furthermore, comparative studies on the perception along the VOT continuum for Thai and English speakers (Abramson and Lisker, 1970), show that Thai speakers can discriminate between all three VOT positions, while English speakers can categorize only between unaspirated and aspirated conditions. Considering this and the fact that other cues have been found to be sufficient in signaling the voicing distinction for other word positions, in English voicing lead seems to carry less information as a perceptual cue than is suggested by traditional statements.

However, under certain rare conditions, prevoicing does have a distinct effect in the initial position. Nearey, Hogan and Rozsypal (1979) describe a pilot study, where, by manipulating VOT information alone, within the appropriate context, they were able to distinguish between *it's bat*, *it spat* and *it's pat*. In other words, they were able to make English speakers categorize along the VOT dimension in a manner superficially similar to Thai speakers. The difference between the two situations is that in one case, only a phonemic distinction is being determined at each crossover, while in the other case the position of juncture is also involved. Here we are offered an opportunity to compare different types of information, phonemic versus subphonemic, on a similar physical continuum.

Experimental techniques developed for tests of categorical perception provide a framework for exploring the issues raised above. Of particular interest is the question of enhanced discrimination along a region of the stimulus continuum involving word juncture and this study is an exploratory look into that question.

The next chapter looks more closely at some of the notions outlined here. Voice onset time will be defined more carefully and its real possible usefulness qualified. Some of the experimental work that has been done on juncture will be reviewed and finally, the notion of categorical perception will be considered. Discussion will focus on the demonstration of categorical perception and the various interpertations of its meaning that people have put forward.

Chapter III describes a measurement study of production. The possible cues affecting placement of juncture between *it's till*, *it's dill*, *it's still* and *it still* are examined. Duration measurements were made on sections of the speech signal corresponding to certain types of acoustic events. These measurements were made primarily to clarify the choice of parameters used in the preparation of stimuli for the perception studies.

Chapter IV outlines four perceptual experiments. In the first experiment with variation along the VOT dimension only, and where the data of the subjects had to be pooled, three categories were obtained as found in the Nearey et al. (1979) study. A large discrimination peak was found on the lag side of the VOT dimension but not on the lead side. The crossover between the categories of the lead portion of the VOT continuum was not very well defined and predicted peaks were low so it was decided to involve other cues in order to get sharper category curves for the next discrimination task. To see how other cues could be combined with VOT in order to obtain sharper identification curves, a crossed identification task was conducted for the second experiment. With a new set of stimuli and a more narrowly focused methodology, the third experiment tested for categorical perception on the lead voicing side only. Results showed that categorical perception could be demonstrated for this part of the continuum. In the last experiment discriminability is again tested along the VOT continuum but, as in experiment one, other possible cues were held constant. This experiment was done across both lead and lag areas but with a larger step size than in the previous experiments.

The last chapter discusses the results of the perceptual studies. In general, it appears that the discrimination and identification results are best interperted on a psychophysical basis where the cues (such as aspiration) associated with a certain range of positive VOT are more perceptually salient than the rest of the VOT continuum.

2. Review of Literature

2.1 Voice Onset Time

Voice Onset Time (VOT) is defined in terms of a timing relationship between laryngeal activity and supralaryngeal articulations. Specifically, it is the time between the release of the stop closure and the beginning of laryngeal phonation. This continuum is then defined in physiological terms and probably it cannot be said to have the property of a simple acoustic dimension. Even so, the measurement of it in speech production has been shown to be a useful differentiator of voicing in stop consonants. The VOT measurements distribute into three main clusters: a long voicing lead represented as negative values, a short voicing lag of around +10 msec, and a long voicing lag with values around +60 to +100 msec. This has been demonstrated with words in isolation across a number of languages (Lisker and Abramson, 1964), and also in running speech in English (Lisker and Abramson, 1967), though for the latter case there is a slight overlap of VOT values with smaller lead and lag values.

As for acoustic properties there is a number of different possible co-varying cues that are manifest acoustically in different areas of the VOT continuum. Voicing lead is generally accompanied with a low amplitude and low frequency voice bar before the burst. When voicing begins after the burst there are a number of possible cues. One is the delay of the first formant (called F1 cutback) as demonstrated by Liberman, Delattre and Cooper (1958). Since F1 has a rising transition after the burst, the frequency at which F1 begins may also be a factor. The transition itself may be a cue for voicing since a long enough delay eliminates the transition (Stevens and Klatt, 1974). Another acoustic property of voicing lag is the noise excitation of the higher formant frequencies which is called aspiration. Some (Haggard, et al., 1970 ; Fujimura, 1971) have discussed the role that perturbations of fundamental frequency play in voicing distinctions.

In positions other than word initial, different kinds of cues have been shown to be sufficient, or more appropriate than those associated with VOT, as voicing cues. Lisker (1957) demonstrated that the duration of the stop closure is sufficient to cue the voicing distinction for intervocalic stops, while it has been shown that vowel length is the most important cue for prepausal stops (Raphael, 1972). Another difference between voiced and unvoiced stops is intensity. The voiceless class of stops has a more intense plosive release and, since voiced stops are usually devoiced in initial positon, some prefer to describe the distinction as lenis/fortis (weak/strong) rather than voiced/voiceless. Wayskop and Sweets (1973) did some studies concerning this difference, demonstrating that the burst release can have a perceptual effect in VC syllables.

2.2 Word Juncture

Because of the continuous nature of the speech signal there has to be some mechanism for dividing the flow of acoustic events into words. There is no doubt that higher level processes influence the parsing of the speech signal but, since minimal pairs such as *a nice man* and *an ice man* can be distinguished out of context, there must be some perceptual cues that can indicate placement of juncture.

This section offers a quick glance at some of the studies that look into the question of what physical correlates are associated with juncture. For a closer look at such studies and also the formal investigations into juncture, see Shammass (1980).

2.2.1 Production Studies

Lehiste (1960) started her study looking for acoustical cues to morpheme boundaries but discovered that, for English, the characteristics of juncture are found mainly at word boundaries. In measuring minimal pairs differing as to placement of juncture, Lehiste identified a number of junctural cues. They included: longer durations for /s/'s in word-initial and phrase-final position; longer stop durations in word-initial position; aspiration for initial voiceless stops; glottalization or laryngealization for word-initial vowels; long durations for final vowels. The /l/ phoneme has formant differences according to its position in the word and also word final /l/'s are longer than in other positions. Nasals also vary according to word position with initial nasals being the longest.

Lehiste's results were verified for running speech by Hoard (1966). He used four speakers to produce the minimal pairs and had a listening test to pick correctly identified items for analysis. Allophonic distinctions proved to be maintained within connected speech. Segment duration cues correlated with juncture but not fundamental frequency or amplitude. Lisker (1975) measured phoneme sequences of /s/ followed by a stop with the juncture either before or after the /s/. He found final /s/'s significantly shorter than word initial /s/'s supporting previous studies.

An extensive study was made on subphonemic details in American English by Umeda and Coker (1975). They found that segmental allophonic variation plays a main role in stops but durational allophonic cues are important for fricatives. Allophonic variation for voiceless stops was determined by devoicing time, while word-initial and stress-initial stops are marked by aspiration. Voiced stops were found to differ in vocal cord vibration. For the initial position the glottal waveform is more similar to a sinusoid rather than saw-tooth waveform as in the case of final voiced stops.

Umeda and Coker found that consonant duration varies according to such factors as stress, position, and context.

As a rule they found that in a fricative context the duration is shortened, while the lengthening factors were stress, word boundaries, and pauses. For lengthening at word boundary they found that the importance of the word is a factor. The more important the word, in terms of information content, the more the consonant is lengthened.

2.2.2 Perceptual studies

We have seen what kind of physical correlates are associated with word juncture in production but how effective are they as perceptual cues? In a study using real speech, Nakatani and Dukes (1977) investigated how these cues affect identification between minimal pairs involving juncture as the minimal difference. The two minimal pairs were each spliced into four slices for areas of suspected junctural cues. Then these slices were exchanged at various locations to create 'hybrids' of the two original word pairs. In this way they tested the strength of the various cues against each other.

They found that the strongest cues were at word onset except for /r/ and /l/ which have distinct allophones for different word positions. The most important cues for boundary perception were burst, aspiration, glottal stop placement, laryngealization and the distinct allophones of initial /r/ and /l/. Duration information did not have much of an affect in the results but in their study they were often competing with stronger allophonic spectral cues.

McCasland (1977) studied in more detail the effects of segmental duration by itself and in competition with aspiration. The four minimal pairs he used were *it's till*, *it's dill*, *it's still* and *it still*. He found that the aspiration of /t/ almost always gave the response *it's till* despite the segmental values for /s/-duration or stop closure. The parsing responses of the other three choices were a result of different combinations of /s/-duration and closure duration of the stop in the second word. For a boundary to be heard after the /s/, the /s/-duration had to be short and the /d/ of *it's dill* had to be long. For the boundary to be heard before the /s/, the /s/ had to be long with the stop closure being short. An even longer /s/ was required to signal the geminate /s/.

Concerning prosodic cues for word juncture, Nakatani and Schafer (1978) eliminated the effect of segmental spectral allophonic cues to test for the experimental effects of rhythm, pitch and amplitude. In their stimuli they replaced all syllables of a three-word noun-adjective phrase with /ma/ syllables but with the stress pattern preserved as the only cue left to signal which two /ma/'s went together. The results indicated that the subjects could parse the phrases using the information from the stress pattern. With hybrid speech synthesis they studied the effects of amplitude, pitch and rhythm independently and found rhythm to be the only aspect of the three to affect parsing behaviour.

2.3 Categorical Perception

The method for the investigation of categorical perception originates from an important experiment by Liberman, Harris, Hoffman, and Griffith (1957). They set up a synthetic series of two-formant stimuli that approximated CV syllables and varied in the direction and extent of the second formant transition. The stimuli varied in acoustically equal steps along a range through which the consonants /b/, /d/ and /g/ are perceived as members of discrete categories.

After an identification task, pairs of stimuli, differing in equal steps along the range, were presented to subjects in an ABX task. In such a task, listeners are first presented with the two different stimuli and then one of them is repeated. Subjects are then asked to indicate whether the third stumulus is the same as the first or second stimulus.

The listener's discrimination performance was enhanced at different regions of the continuum, noticeably at the identification boundaries. That is to say, the discrimination function increased as it approached the identification boundary and then decreased as it left the boundary. In contrast to this, most discrimination functions in psychophysical studies are either monotonically increasing or decreasing.

This established a standard test for categorical perception. The criteria were specified by

Studdert-Kennedy, Liberman, Harris and Cooper (1970) and are as follows:

1. There should be distinct labeling categories with the identification functions having an abrupt crossover on the continuum at the boundary.
2. When the stimuli being compared in the discrimination task are from the same category, the discrimination between them is at the chance level.
3. At the region of the boundary there are peaks of improved performance in the discrimination.
4. Finally, that the discrimination function can be predicted from the labeling function, where the probability of discriminating two stimuli is equal to the probability that the stimuli are identified as different.

In other words, for categorical perception the listener can discriminate only as well as he can identify.

As an example of how the discrimination function would be calculated under the above critera consider two stimuli being compared in the Liberman et al.(1957) experiment. Here the subjects were asked to identify /b/, /d/ or /g/ based on the slope of the second formant. After the identification values were graphed, the percentage identification values were used to compute the discrimination function. For example, let Pb1 represent the probability of the first stimulus being identified as a /b/ (taken from the percent identification of /b/ for that stimulus number) and Pb2 as the probability that the second stimulus will be identified as a /b/. Using the same convention for /d/ and /g/ the resulting formula (where P(D) is the predicted discrimination) would be:

$$P(D)=.5+.25((Pb1-pb2)^2+(Pd1-Pd2)^2+(Pg1-Pg2)^2)$$

This function, called the Haskins model, provided a fairly good fit to discrimination data involving consonants. However within-category discrimination was usually somewhat better than chance. Fujisaki and Kawashima (1970) added an extra component to the Haskins model. They proposed that besides a phonetic memory for phonemic category there is an auditory memory. Two signals could be compared in auditory memory to discriminate characteristics of the signal that are non-phonetic (called 'timbers'). But because auditory memory decays much faster than phonetic memory, it is phonetic memory that usually plays the dominant role in discrimination at category boundaries. It was posited that auditory memory is operative in within-category discrimination. Fujisaki and Kawashima (1970) added this as a factor to the Haskins model and found that it provided prediction curves which gave a better fit to the obtained curves. It should be noted that the added component is estimated from the obtained data and this did not provide independent evidence for such a memory.

Pisoni (1973), however, did provide some evidence for the notion of two types of memory. Vowels had been shown to be perceived more continuously than consonants (Fry, Abramson, Eimas, & Liberman, 1962; and Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969) with fairly good discrimination within categories and less dramatic peaks at the boundaries. Obtained overall discrimination was much better than the predicted discrimination and the identification curves have less abrupt crossovers. Pisoni (1973) suggested that more auditory memory is available for vowels than consonants due to such factors as having longer durations (supplying more information) and being presented as steady state signals. Employing an AB paradigm (where listeners judge the two stimuli as being different or the same) he changed the time interval between the two stimuli. Vowels show a decrease in discrimination performance as the interval increases while consonants do not. The decrease is interpreted as result of information in the auditory memory being lost as the interstimulus time interval increases. Consonants already have little representation in the auditory memory being discriminated via only phonetic memory which lasts a little longer. This same argument has also been suggested for the finding of categorical type perception with vowels of short duration. It has been suggested that rather than two distinct modes of perception there may be more a difference of degree between so called categorical and continuous perception (Pisoni 1971).

Although categorical perception was originally thought to be unique to speech stimuli, it was later demonstrated for various non-speech stimuli. Cutting and Rosner (1974) demonstrated categorical perception along a dimension with stimuli ranging in short to long rise times. Subjectively they were perceived being either plucked or bowed type sounds. Noise-buzz sequences were used by Miller, Wier, Pastore, Kelly and Dooling (1976) to demonstrate categorical perception for non-speech sounds. They varied the onset of the buzz in relation to the noise onset, with the offset of the components always ending at the same time. The labeling functions showed a sharp boundary at around 15 msec of noise lead (but with a large amount of variability between subjects) and the criteria for categorical perception, as set out by Studdert-Kennedy et al.(1970), were met.

There has been some question as to whether these effects along the continuum showing categorical perception are 'natural' or learned. Possible evidence for existence of categorical like perception has been shown for two month old infants along the VOT continuum (in aspirated portion) by Eimas, Siqueland, Jusczyk, and Vigorito (1971), and along a non-speech continuum (rise time cues) by Jusczyk, Rosner, Cutting, Foard, and Smith (1977). Kuhl and Miller (1975) demonstrated that chinchillas are able distinguish the /t/ and /d/ sounds.

Such studies suggest the possibility of 'natural' rather than learned categories, but evidence has been shown for learned distinctions as well. Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) showed that native adult English speakers could categorically perceive on a continuum ranging from /r/ to /l/ but Japanese adults, whose language does not have the distinction, could not.

Many have now been proposing that categorical perception may be more appropriately understood if characterized at the psychophysical level. Miller, Wier, Pastore, Kelly, and Dooling (1975) talk of a masked threshold effect, where a single signal component is varying in relation to a stimulus complex. Pastore, Ahroon, Baffuto, Friedman, Puleo, and Fink (1977) support similar veiws. Rather than a direct casual relationship between the abilities demonstrated by the categorization and discrimination functions, as found in models with a phonetic memory component, they prefer to have a single (but common) factor which is responsible for the two types of performance. As to the exact character of such a factor, Pastore et al. talk of internal or external limitations. An example of internal limitations would be a sensory threshold of some kind. In discussing this notion, they note that many examples of categorical perception involve a timing relation with the critical duration being at about 15 to 25 msec. Such examples includes most VOT studies including the infant study by Eimas et al.(1971) and the noise-buzz experiment by Miller et al.(1976). External limitation would involve some kind of interfering or reference stimuli.

In summary, then, there appears to be a number of qualifications to the original notion of categorical perception. The difference between categorical perception and continuous perception appears to be more a matter of degree of acoustic saliency. Also in some cases the difference can be reduced to a psycho-acoustic explanation such as sensory threshold. Keeping this in mind, what purpose can the discrimination task have in the experiments outlined in Chapter IV. First of all, consider the situation involved according to the standard phonemic analysis of English stops. With the case involving juncture, in the lead portion of the VOT continuum, where the crossover from *it's dill* to *it still* occurs, there is a change in phonemic category along with a change in juncture, while in the lag portion of the VOT continuum (with the crossover from *it still* to *it's till*), there is a change in only allophones that indicates the change in juncture. However, in the case with only a change in phonemic category, the situation is reversed. The lag part of the VOT continuum is cueing phonemic change while pre-voicing is not cueing anything.

The last case is the typical type of situation for tests of categorical perception. It is the goal of this experiment to compare the differences between these two cases and examine the subject's discrimination performance. The results will be related to the predicted performance which is determined from the identification performance. The results may say something about how the same acoustic cue affects perception in different contexts and perhaps how it operates differently at different levels of perceptual processing. Still, much of our interpertation may rest on assumptions about categorical perception and phonemic theory which we may want to consider differently in light of the results.

3. Measurement study of production

3.1 Procedure for Collection and Measurement of Data

This chapter deals with measurements made of the four, minimally contrasting, utterances used in a perception study by McCasland (1977). The four utterances are *it's till*, *it's dill*, *it's still*, *it still*, where the intensity and duration differences near juncture points are analysed.

3.1.1 Speakers

Twelve speakers were used for recording, six females and six males. All were native speakers of Canadian English except one of the females who was a native American.

3.1.2 Apparatus

The following is a list of instruments, and their technical specifications, used in this study.

1. Microphone: Sennheiser MD 421N, frequency response 30-17000 Hz. +5 dB; sensitivity .2 mV/microbar at 1000 Hz.; cardioid directionality.
2. Tape Recorder: TEAC A-7030, frequency response 50-15000 Hz. +2dB; speed 15 ips.; SNR 58 dB.
3. Audio-frequency Filter: FrøKjauer-Jensen type 400, frequency response slope 36 dB/oct.
4. Minicomputer: PDP-12A; word length 12 bits; A/D, D/A converters 10 bits; operating systems OS/8 and

Alligator.[1]

3.1.3 Recording

Each speaker was individually recorded in a sound insulated recording room. They each went through a ten stimilus list (see Appendix A) four times, where he or she repeated the token when prompted by hearing it on a master tape. The master tape was used to regulate the tempo of speaking. Each speaker also had a written list of the stimuli in front of him/her. The speakers were told to talk in a natural manner and were allowed to practice the list once before the four replications were recorded. The list consisted of three one-word items followed by the four two-word items which were of main interest in this study. It is possible for the last few items of a spoken list to be affected by a different intonation. To avoid this, another three two-word items were included at the end of the list.

3.1.4 Digital Gating

An interactive Alligator program was used to digitize the desired phrases which were then stored on magnetic tape. In the procedure, the signal coming from the tape recorder is band-pass filtered from 68 to 6800 Hz. This is to eliminate 60 Hz hum and speech components above 8000 Hz

----

[1]Developed by Stevenson and Stephens (1978), the Alligator programming system is written in OS/8 PAL 12D assembly language and is designed for psychoacoustic experimentation. The system is executable on PDP-12 computers.

before the signal is digitized. The signal amplitude was adjusted for the broadest range of quantization while still avoiding signal clipping. Only the first three replications were digitized, if one of them were bad, the fourth was used. In Figure 1 the block diagram is shown.

3.1.5 Segmentation and Measurement

Utilizing Fortran programming similar to that described by Nearey and Hogan (1979) each stimulus sentence was segmented into seven or more segments. To aid in segmentation, devices were available for playing back and observing the spectrum of desired segments of the signal. The block diagram is the same as in Figure 1. The following explains how the cursors for the beginning of the different sections were defined. Reference to Figure 2 will make the explanations easier to follow.[2]

1. /I/-vowel(I): Start of vowel /I/ in 'it'. The first cursor is set at the beginning of the waveform periodicity or a glottal stop, if the case be. The end of each segment is marked by setting the next cursor.
2. /t/-closure(T): Start of the closure of the stop /t/ in 'it'. Sometimes indicated by zero amplitude but often there is voicing carried on through part or all of the closure. A judgement has to be made as to where the vowel ends. This is usually easily detected by a change

---

[2] It should be noted that in changes from one signal type to another (for example, from vowel to stop-closure) there are natural boundaries for segmentation (Fant, 1982).

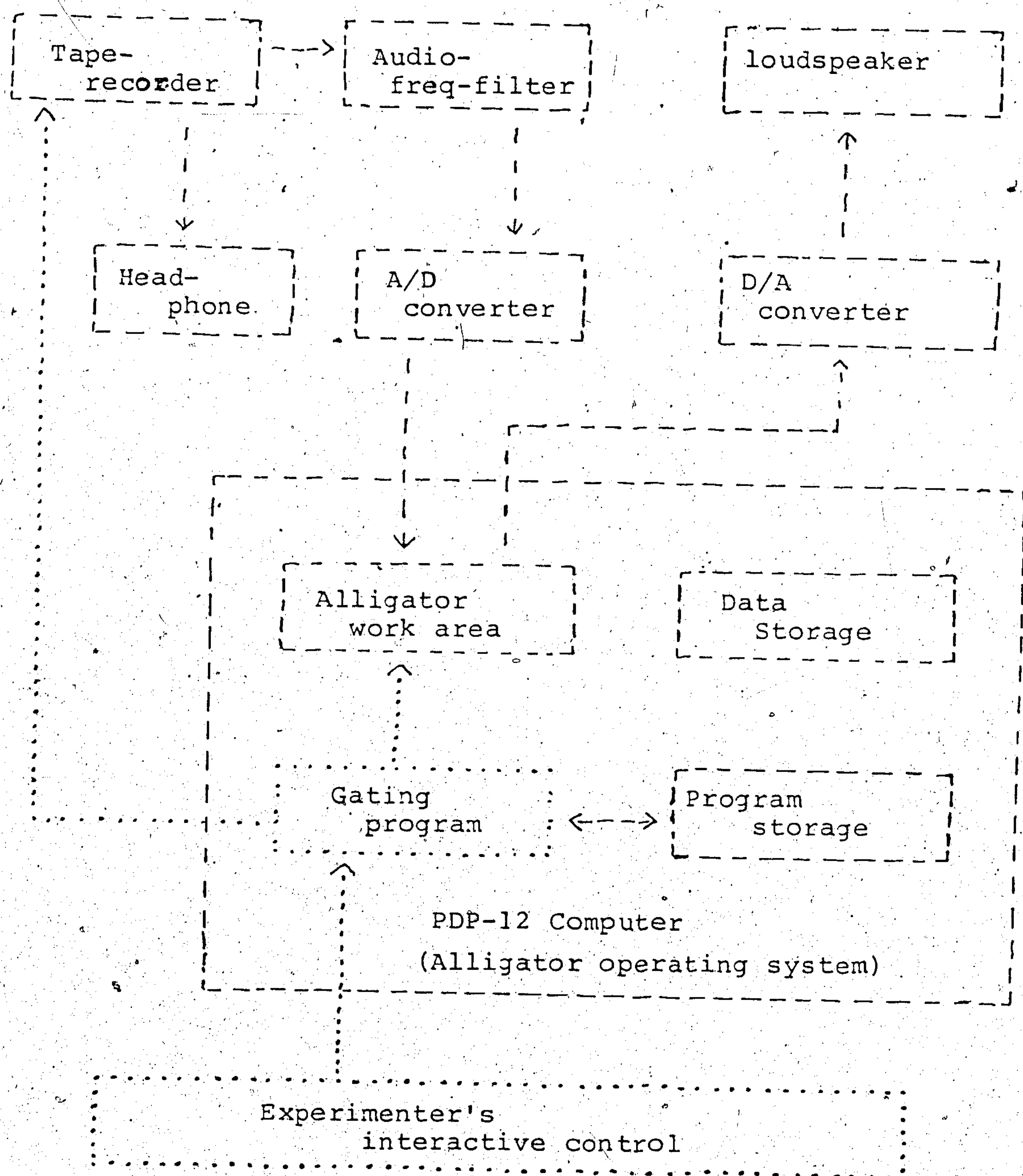


Figure 1: Block diagram of digital gating and segmentation. *Dashed arrows indicate signal flows; dotted arrows, control flows; dashed boxes, devices; and dotted boxes, controllers.

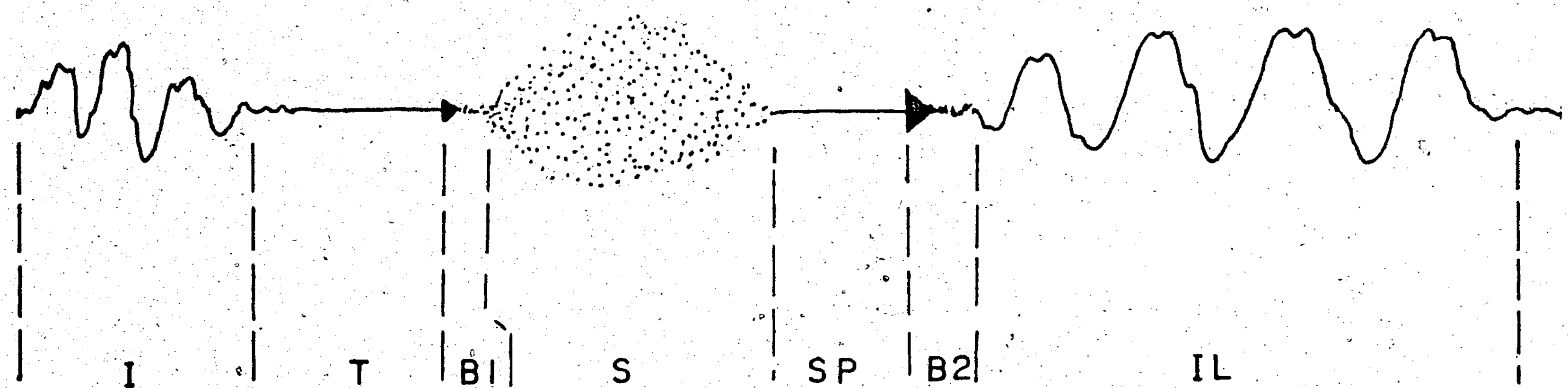


Figure 2: Signal segmentation

in the waveform with its simplification and decrease in amplitude and also the change in the spectrum.

3. /t/-burst(B1): the /t/-burst of 'it'. Usually easily detected on the waveform and gives a distinctive spectral section.
4. /s/-frication(S): gives a spectral peak at approximately 6 kHz, and appears as random points on the digital display of the waveform.
5. Silent period(SP): /s/-noise ends giving a silent period representing the closure for the next stop.
6. Voice bar(VB): marked by the onset of its waveform. This cursor was rarely used.
7. Dental burst(B2): the burst of the second stop including aspiration.
8. /Il/-nucleus(IL): Beginning of vowel /I/ in the second word. Decisions for the placement of cursors were based on studying the spectral sections for the start of F2. This segment included the /l/.

See Appendix B for the means of the measurements (these are given for the raw scores and also for the square root transformation discussed below).

3.2 Statistical Analysis: Results and Discussion

The duration measurements were analyzed by the analysis of variance (ANOVA). The Bartlett test for homogeneity was made on the raw duration data and also on two transformations of the data. The two conversions were log and square root. The square root transformation showed the least heteroscedasticity so it was used for the analysis of the duration measurements. However, it was still high enough to warrant a conservative F-test (Winer, p.206, 1971). Significant effects are reported at the .05 and .01 levels.

3.2.1 Analysis of Variance

ANOVA's were done for each of the segments except for the voice bar (VB) which only occurred five times. The design consisted of twelve speakers, as subject factor (S), fully crossed with the four sentence types, as the second factor (T), with three replications in each cell. The ANOVA's for the duration values of each segment are shown in Table 1. Speaker main effects were significant for all segments except for the dental burst in 'it'(B1). For sentence type, B1 is also the only section that shows no significant effect. The I section is significant to the .05 level while the rest are to the .01 level. None of the interactions showed significant effects. Twenty-one correlations were computed for the data points among the seven ANOVA's that were carried out. The highest correlation co-effecient was -.30 which may indicate measurement variation given the fixed boundary between two adjacent segments. Only two were significantly different from zero correlation which was also indicated in

Table 1

ANOVA's for Duration

Source	Error	df	I	T	B1	S	SP	B2	IL
S	R(ST)	11	14.36**	9.31**	1.78	12.18**	10.91**	4.57*	5.75*
T	ST	3	4.93	19.17**	1.99	21.48**	18.83**	70.06**	9.95**
ST	R(ST)	33	0.62	2.74	0.60	2.17	1.75	3.88	0.97
R(ST)		96							

scatterplots made for each of the 21 pairs. Therefore the data points in each ANOVA will be treated as independent of each other and will be discussed separately.

To test for significant differences between the four sentence types the Tukey(type a) test was used. Main effects due to speakers and also sentence effects due to the /I/-vowel segment were not analyzed. Table 2 contains a summary of the results.

For the closure duration of the /t/ in 'it(s)'(i.e. T), *it still* shows significantly longer closure than in the other three sentences. This demonstrates the difference in duration between the non-clustered /t/ in 'it' and the clustered /t/ in 'its'. In the case of the ANOVA for /s/ durations the double /s/ of *it's still* is significantly longer than in the other three sentences. Surprisingly, the word-initial /s/ of *it still* is not significantly longer than the /s/ of the two sentences with word-final /s/ only.

The closure duration of the apical stop in the second word (see the ANOVA for SP) is significantly shorter for the two sentences with 'still' in it. That is, as with T, the non-clustered stops are significantly shorter than the clustered stops. For the burst, and aspiration, of the second stop in the sentence (i.e. B2), *it's till* had significantly longer durations than those of the other three sentences, due to the aspirated /t/. Finally, for the vowel plus /l/ section (IL), *it's dill* showed significantly longer durations than the other three.

Table 2

Tukey Test

Segment: T

	1	2	3	4
sentence type:	its dill	its till	its still	it still
mean:	7.077	7.118	7.561	10.018

	1	2	3	4
1		0.127	1.57	9.195**
2			1.385	9.067**
3				7.68**
4				

Segment: S

	1	2	3	4
sentence type:	its dill	its till	it still	its still
mean:	10.853	11.377	11.877	13.738

	1	2	3	4
1		1.934	3.7785	10.65**
2			1.845	8.71**
3				6.85**
4				

Segment: SP

	1	2	3	4
sentence type:	it still	its still	its till	its dill
mean:	7.382	8.135	9.036	9.732

	1	2	3	4
1		3.175	6.98**	9.92**
2			3.806*	6.75**
3				2.94
4				

Segment: B2

	1	2	3	4
sentence type:	its dill	it still	its still	its till
mean:	4.727	5.363	5.364	8.647

	1	2	3	4
1		3.00	3.00	18.5**
2			0.005	15.5**
3				15.5**
4				

Segment: IL

	1	2	3	4
sentence type:	its till	it still	its still	its dill
mean:	15.847	16.224	16.235	16.945

	1	2	3	4
1		2.6	2.8	7.57**
2			0.02	4.97**
3				4.77**
4				

3.2.2 Summary

The analysis of the duration measurements indicates to us which elements of the signal, found in speech production, are available to distinguish the four different sentence types that were used in this study. The main factors were, aspiration, /s/-duration and closure duration for the two stops. This for the most part concurs with previous studies but an exception might be /t/-closure (T), which shows up as a strong factor in this study. McCasland (1977) only studied S, SP and aspiration so that the effect of T as a perceptual cue was not tested.

Duration of /IL/ also showed up as a significant factor. In production *it's till* had the lowest mean duration value for /IL/. This is because aspiration, reflected in the B2 measurements, takes up part of the syllable. In the sentences *it's still* and *it still*, /IL/ may be shorter due to some isochronous type of effect in the production of the utterances.

It was surprising to see that the word-intial /s/ durations are not significantly longer than the word-final /s/ durations as often reported for production (Lisker, 1965) and preception (McCasland, 1974). Having *it still* together with *it's still* in the recitation of the list may have caused some of the speakers to utilize production strategies they would not have used otherwise. Although it was not intended for *it's still* to play any part in the perceptual experiments, it was added here in the measurement study in order to get a more complete look at the role of the durational cues studied here.

4. Perception Experiments

4.1 Experiment I

As mentioned in the introduction, Nearey, Hogan and Rozsypal (1979) reported that a three way distinction could be obtained on the VOT continuum with a particular set of two-word utterances. This experiment is an attempt to replicate the same result for *it's till*, *it's dill* and *it still* and as well to test how close this distinction fits the Studdert-Kennedy et al. (1970) criterion for categorical perception. Since we are using a consonant cue (VOT), which is considered to have poor auditory memory, we expect the discrimination performance to fit the curve, as predicted by the Haskins Model, fairly well for the three category distinction. An identification and discrimination task was also done for the sentences 'its a dill' and 'its a till' in order to have an example of the two-way distinction in an environment comparable to the sentences above, rather than in isolation.

In this, and in the following experiments, gated natural speech is used except for the voice bar. Only one male speaker was used due to storage limitations of the computer. The speaker chosen had already been used in previous studies (Shammass, 1980) with satisfatory results.

4.1.1 Preparation of Stimuli

The stimulus items were prepared using interactive Alligator programming. The following describes their preparation.

4.1.1.1 Constuction of stimulus items

In this experiment all segment durations except VOT would be kept constant. The /it/ and /s/ portion were taken from one of the *it still* sentences being similar to the frame used by Neary et al. (1979). This would give the first pause (t-closure refering to Figure 2) a duration of 120 msec and /s/ a duration of 155 msec. The second stop closure (SP) was kept at a duration of 100 msec. The /Il/ portion was from a *it's dill* sentence. An 'it's a' portion was taken out of a recording of 'it's a dill', made by the same speaker. Also a voiced /d/-burst and the intended voice bar were gated from this sentence to be used in making the VOT stimulus.

For the explanation of how the stimuli differing in VOT were constructed, we shall start with the aspirated part of the continuum. Figure 3 shows (schematically) a 'dill' and a 'till' utterance by the speaker and what sections of them are to be gated out. The first eight glottal pulses, coming after the burst in 'dill', are segmented at their zero crossings before the highest peak in the waveform of each glottal pulse. As these portions were gated out and stored they were labeled DP1 to DP8. The remaining 'dill' vowel

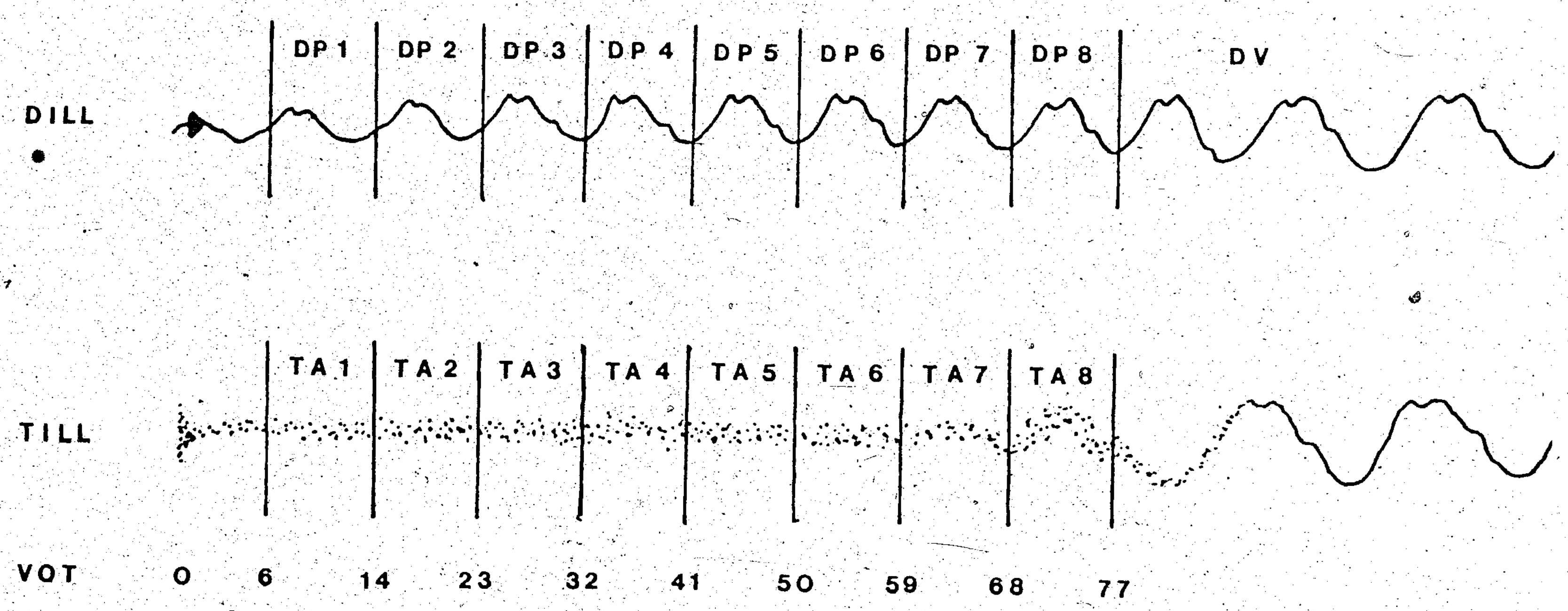


Figure 3: Preparation of stimulus items

(labeled DV) was also stored. All these pulses were nine msec in duration except for DP1 which was eight msec long. Finally, the 'dill'-burst was stored and was six msec long.

From 'till', segments of aspiration corresponding to the duration and sequence of the 'dill'-pulses were gated out and stored as TA1 to TA8 (see Figure 3). To produce the different stimuli with stops having positive VOT, the 'dill'-pulses would be removed and replaced by sections of 'till'-aspiration. For example, a CV syllable with 32 msec of aspiration would be created by queuing together the burst plus TA1 to TA3 then adding the glottal pulses DP4 to DP8 and finally adding DV.

Appendix C gives the VOT values for each stimlus item. For the stimulus item with zero VOT, the original devoiced burst was replaced by a burst that had voicing carried through it (from the 'its a dill' sentence). Added to this would be the voice bar of varying lengths to give the remaining 10 VOT stimuli. The duration of each is nine msec except for between the stimulus items with 0 VOT and +6 VOT and between items with +6 VOT and +14 VOT. These differences should not be too critical since it is a matter of conjecture to say what acoustically equivalent step sizes on the VOT continuum are anyway (Stevenson, 1979). Also, in calculating the prediction function from the Haskins model, the important thing is how the two stimuli being compared are identified. As a reminder, it should be noted that stimulus items with VOT values from -90 to 0 have a different burst than stimulus items with VOT values from +6 to +77 (see Appendix C).

It was attempted to take the voice bar from the speech signal. However, in the process of being desampled and recorded onto tape, it picked up a nonspeech like quality that caused it to be perceived as separate from the rest of the speech signal. A synthesised voice bar was produced so that it could be digitized at a high amplitude and then scaled down which solved the problem somewhat but not totally. The synthesized voice bar had the following characteristics: the waveform was a triangular function, band pass filtered at 70 to 200 Hz and varied in frequency from 105 Hz, at the start, to 93 Hz, at the end. The 100 msec voice bar was stored and to create the appropriate voiced bar durations the right amount of digital points would be removed from the front of the signal. The onset of the waveform was smoothed by multiplying it with the initial five msec of a cosine squared window. This was done to eliminate any discontinuities in the voice bar due to gating at a point above or below the zero level.

This still did not totally clear up the noise problem in the recorded stimuli. It was suggested to add another source of noise to mask it out. Following this suggestion, the lowest possible level of noise needed to alleviate the problem was determined. When the stimuli were recorded, white noise (actually a lower frequency band of noise), was added to the signal before it was filtered. The signal to noise ratio was monitored and measured to be equal to 35dB.

4.1.1.2 Arrangement of the stimulus items

In the identification task of each VOT condition, that is, in the two category and three category conditions, the 20 stimuli were used with five presentations on each, making 100 test items. Before making each identification the subjects listened to the presentation twice. There was a tone after every ten pairs presented.

For discrimination a 4IAX task was used. In this task two pairs of stimuli are presented one after the other. One of the pairs has the same stimuli, while the other pair has different stimuli. The listeners' task is to tell which one is different. This arrangement puts less of a load on memory than the ABX task (Pisoni, 1971) and, since we are using stimulus items which are longer in duration than usual, it would be appropriate to keep the load on memory as low as possible. The prediction formula is the same as that for the ABX task (Pollack and Pisoni, 1971), and is shown in section 3.1 (replace the place of articualtion categories for the sentence categories used here).

A pilot study was conducted to determine the optimum step-size, where it is not large enough to give too good a discrimination within categories but allows us to demonstrate discrimination between categories. A step size of three was decided on, meaning that the stimulus pairs being compared were usually 27 msec apart on the continuum.

Also this gave us 17 pairs to be compared along the continuum. For comparing a stimulus pair in 4IAX there are eight possible combinations. These eight arrangements were used as the eight trials for each stimulus item which gave a total of 136 comparisons to be made in the discrimination task. The 136 comparisons were randomized and, with an Alligator program, they were recorded onto tape. For both discrimination tests the time interval between the stimuli in each pair was 50 msec. The time between the two pairs was 200 msec and the interval between each group of pairs was one second.

4.1.2 Listeners

There were ten subjects, eight of whom were taking an introductory phonetics course. Nine of the listeners were native speakers of Canadian English. The tenth was a native speaker of American English and a trained phonetician.

4.1.3 Apparatus

1. Power Amplifier: Braun AG Type CVS 250
2. Tape Recorder: Teac A-7030.
3. Headphone Sets: Telephonic TDH-49, frequency response 30 to 6000 Hz +3 dB.
4. Audio-Frequency Filter: Rockland 1524-01, slope of frequency response: 24 dB/oct.
5. 33108 Function generator: Hewlett-Packard
6. 1382 Random noise generator: General Radio Company

4.1.4 Procedure

Listeners first did identification on the 2-way distinction (namely, *it's a dill vs it's a till*). They were given an answer sheet and a key which indicated letters encoding the response type. They where asked to make their responses after hearing the second repetition. After completing this, they did the 4IAX task on the 2-way distinction. They were told to mark either a '1' or a '2' for whether they heard the second or first pair as different. They were encouraged to guess if they could not tell which pair was different. After this, identification and discrimination was done in a similar manner on the stimuli with the three-way distinction (that is between *it's dill*, *it still*, and *it's till*). Each identification task was six minutes long while the discrimination tasks were twenty minutes each.

4.1.5 Results and Discussion

Since there were only eight discrimination trials and five identification trials for stimulus item per listener the data had to be pooled together. Figure 4 shows the identification and discrimination functions for the *it's a dill vs it's a till* task. The number of trials for each stimulus number is indicated by 'N', which is a composition of ten subjects each listening to five presentations for the identification task and eight presentations for the discrimination task. The crossover between the two stimuli

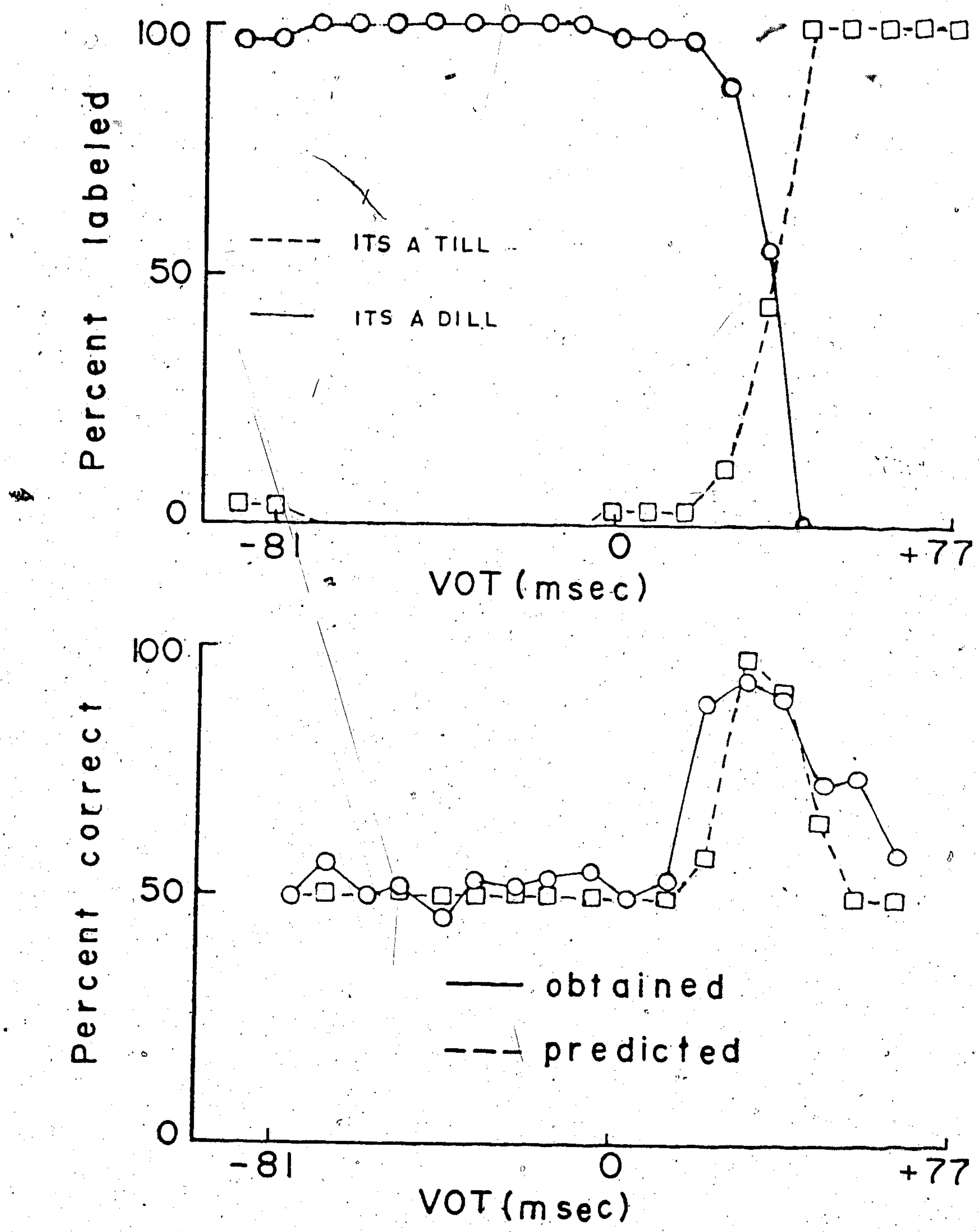


Figure 4: Identification and Discrimination for the two-way distinction in Experiment I.

is at about 32 msec of positive VOT. The discrimination peak is at +27 msec (where items with +23 msec and +40 msec of VOT are presented), as was predicted from the labeling function. A chi-squared test for goodness of fit revealed a significant difference, at the .05 level, between the observed and the predicted discrimination curves (see Appendix F). The significant difference is attributable to mainly two points on VOT continuum. The largest difference is the comparison between stimulus items with +6 msec VOT and +32 msec VOT. While 58% discrimination is predicted, the actual discrimination is at 88%. The other comparison is between +44 msec VOT and +68 msec VOT. Since they are labeled within the same category 100% of the time, discrimination between them should be at the chance level; however, discrimination turns out to be close to 75%. In these two cases it is possible that they may have been discriminated via an auditory memory rather than a phonetic memory. If so, this memory seems useful in only certain regions of the continuum hinting at some kind of psychoacoustic explanation.

Figure 5 shows the curves for the 3-way distinction. While the data are noisier and the crossovers are not as sharp, the three categories are obtained as was found in the Nearey, Hogan and Rozsypal (1979) study. There are, however several points on the identification function indicating a selection of a category considered well beyond its extreme boundary value on the VOT continuum. One might speculate

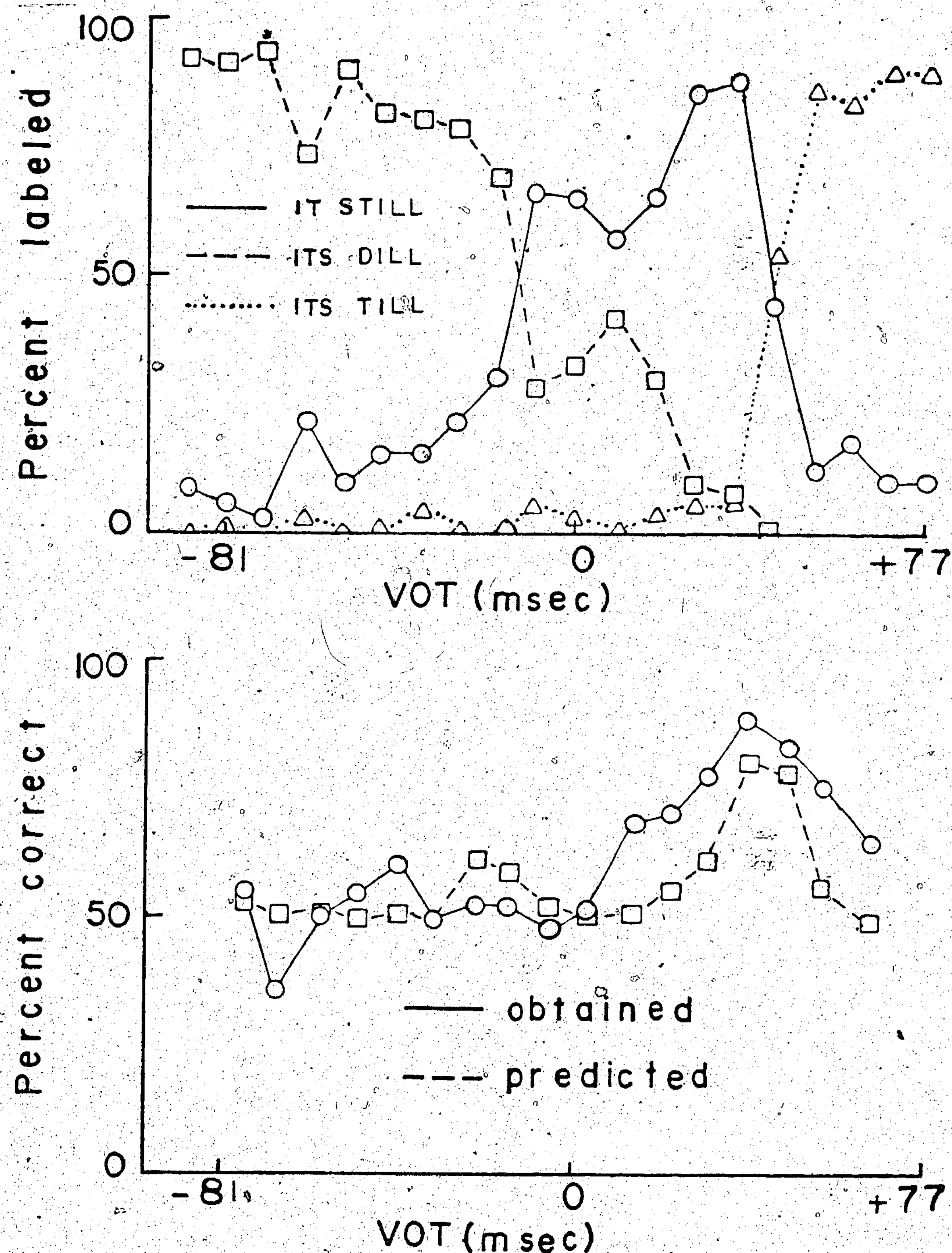


Figure 5: Identification and Discrimination for the three-way distinction in Experiment I.

that many of these are errors due to lack of concentration when the stimulus was not attended to or a response other than the intended one was accidently given. This may have been because of added confusion with the extra category and changes in word boundary but another factor may have been fatigue and boredom, since this task was done after the identification and discrimination for the two-way distinction.

The crossover between *it's dill* and *it still* is approximately at -14 msec. It is not that well defined and the slope of the curves are not particularly steep compared to those of the previous experiment. One thing that affected this crossover was the change in direction for the stimuli with 0 and +6 msec VOT. At first, this was thought to be due to subject differences, perhaps two groups with their crossovers at different locations. However, later Experiments III and IV, with individual subjects revealed that the individual subjects also showed this trend. For a possible explanation, the noted change in direction may have been due to a change in the burst between the stimulus items with the VOT values of 0 and +6 msec, however, the change in trend already starts with 0 msec VOT rather than after it. There do not seem to be any other problems with the way the stimulus was set up. This part of the continuum may in some way be peceptually unstable for native English speakers. The crossover between *it still* and *it's till* is almost just as sharp a crossover as in the task with just two categories. This crossover is at approximately +40 msec VOT.

As for the discrimination function, the chi-squared test of goodness of fit showed a significant difference between the observed and predicted curves (see Appendix F). The peak in the lag region is at 36 msec. The predicted peak is at the same point but, for the most part, obtained performance showed better than predicted in the positive VOT region. Although it may not be a significant difference as in the two-way discrimination, it is consistent with the possibility that some discrimination being made is independent of how the stimuli are labeled. Many of the subjects commented on how the /t/ in 'till' stood out the most saliently when it occurred in a discrimination pair. As for the negative VOT range, the obtained discrimination is represented by an erratic curve, which seems to indicate little relationship to the predicted curve. The predicted peak was at -22 msec VOT while the highest peak for obtained discrimination was at -40 msec VOT, though it never reached higher than 60%. At first it might seem that the available categories are not being used to discriminate with but the peak of the predicted discrimination (62%) is barely significantly above chance[3] and obtained peaks are even lower.

---

[3] For percentage value to have the limits of its confidence interval (at the .05 level) above the 50% level; more than 64% is required when N is equal to 50 and more than 60% is required when N is equal to 100.

This experiment was an attempt to compare the phenomenon of categorical perception in the two different situations (i.e. the two boundary crossovers), but the boundary between *it's dill* and *it still* did not produce a strong enough distinction (given the step-size used) to test for categorical perception. This may have been due to the cue used. Perhaps the voice bar was too weak a cue to produce a strong enough distinction. On the other hand it may be something about the distinction between *it's dill* and *it still* itself that is less salient than the traditional phonemic distinctions that have been tested (i.e. perhaps the discrimination between these two sentences takes place at a different level of processing). Experiment I tried to cue the distinction by cueing a phonemic change in the dental stop, but perhaps allophonic differences of the /s/ and the stop closure differences are the cues necessary to make this distinction, since they turned out to be more common in the production study. This category boundary should be explored more closely along with these other cues. Higher percentages of correct discrimination are needed for this distinction but, to get larger predicted peaks the slopes have to be steepened. Towards obtaining this end, the next experiment is an identification task looking at how these cues might be combined, while Experiment III will utilise the cues in an identification and discrimiantion experiment.

4.2 Experiment II

This experiment manipulates some segmental duration values (/t/-closure of the first word, /s/-frication, and the closure for the dental stop of the second word)[4] in order to see how the recognition of VOT operates under different conditions.

4.2.1 Preparation of Stimuli

To keep the number of stimulus items down to a reasonable amount, six levels of VOT were chosen. They were at -54, -27, 0, +23, +50, and +77 msec of VOT. There were two levels of each /t/-closure (55 and 99 msec), /s/-frication (105 and 150 msec), and silent period (55 and 99 msec). When these are fully crossed they make up 48 stimuli. These were randomized with four replicates to make up a total of 192 stimuli. The /s/-duration was manipulated by queuing different durations of the middle /s/-portions between a beginning /s/-section and an end /s/-section.

When presented to the listener, each item was repeated twice with a 500 msec interval between them; the interval between stimulus presentations was 1800 msec. A tone was played after every ten items. As in the previous experiment noise was added to the recording.

----

[4]See Figure 2.

4.2.2 Listeners

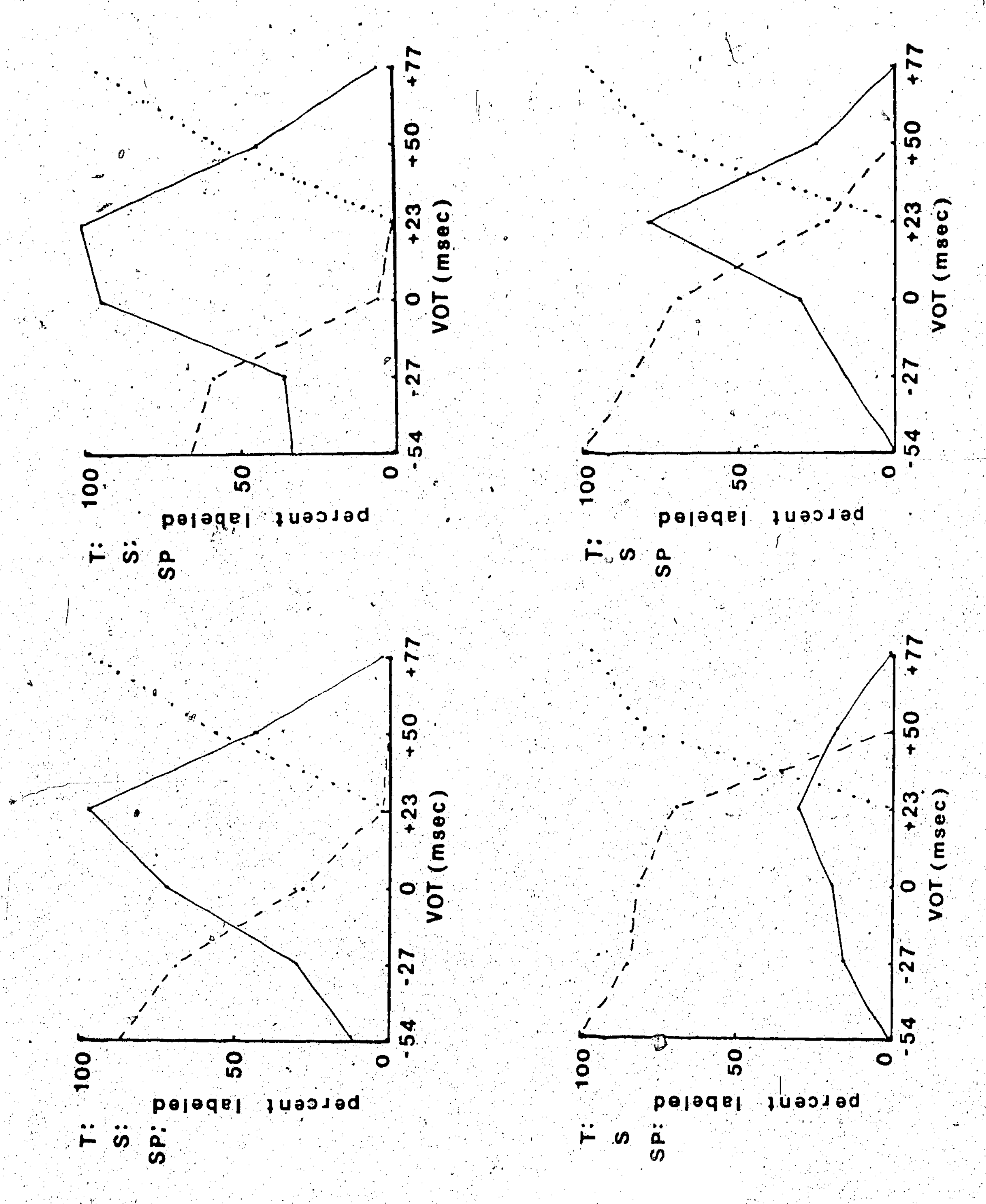
Eight subjects, four females and four males, were used in this study. Six were native speakers of Canadian English and two were speakers of American English, of which one was a trained phonetician. The last also participated in the previous experiment.

4.2.3 Apparatus and Procedure

The apparatus is the same as in Experiment I. Appendix D shows the instruction sheet given to the listeners which describes the procedure. As indicated in Appendix D naturalness judgements were also collected on the items as to the category in which they were identified. It was originally thought that these judgements might add some useful information, but later the results proved to be intractable for analysis.

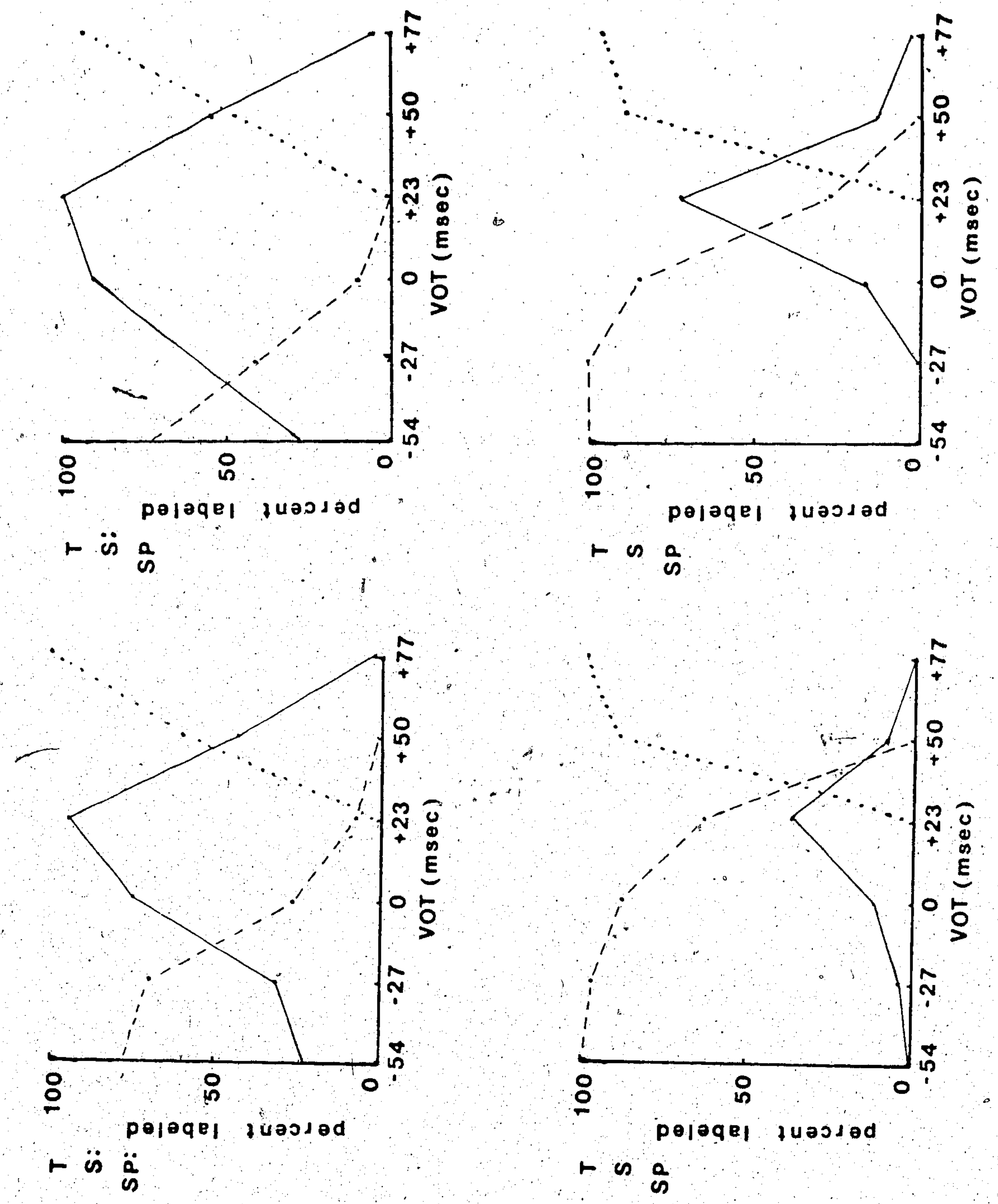
4.2.4 Results and Discussion

Figures 6 and 7 shows the results of Experiment II. The graphs show the identification curves along the VOT continuum for the eight different conditions of segment duration. In the labeling of the graphs, ':' indicates the longer duration value. The /s/-duration has the greatest effect. When the /s/-duration was long, the VOT values of 0 and +23 msec showed up strongly as *it still* responses. The next strongest duration cue is the silent period of the second stop closure (SP). Its effect is most noticeable



---- its dill —— it still ······ its till

Figure 6: Identification for Experiment II



---- its dill —— it still ······ its till

Figure 7: Identification for Experiment II

when the /s/-duration is at the shorter value. With both a short /s/-duration and long SP there is no majority response for *it still* along the VOT continuum. The two levels of duration for the first stop closure (T) had very little effect on the identification scores, although this was a difference that showed up strongly in the production study in Chapter III. These results are also inconsistent with the production study in that /s/-duration proved to be the most potent cue distinguishing initial and final /s/'s, while it was insignificant for this in the production study. However, these perception results are consistent with other studies on juncture where the significance of /s/-duration is indicated[5]. For Experiment III the most useful cues to use along with VOT will be /s/-duration and the duration of the silent period after /s/.

Another thing to note from the results is that the category along the VOT continuum most resistant against the changes in segment duration is *it's till*. Stimuli with +77 msec of VOT show practically no change at all and was identified for the most part as *it's till* 100% of the time. Only /s/-duration has some effect on stimuli with a VOT of +50 msec, while the rest of the VOT continuum values are not identified as *it's till* at all. As with the results from the production study and Experiment I, positive VOT appears to be a very powerful cue that separates *it's till* from the other sentences.

---

[5]See Section 2.2

4.3 Experiment III

In Experiment I, categorical perception involving voicing lead may have not shown up for several reasons, such as pooling of subjects and the lack of steep enough slopes in the identification curves at the crossovers. In this experiment we attempt to make the identification curves steeper at the crossover; this would lead us to expect higher discrimination peaks. Also we collected more data per listener.

4.3.1 Preparation of Stimuli

So that more data for each stimulus can be collected, the area of the VOT continuum studied in this experiment has been limited to from -54 to +23 msec VOT. We were looking therefore at the *it's dill* versus *it still* boundary only. To make the slopes steeper at the crossover, other durational cues have been added to reinforce the VOT cue. For -54 msec of VOT we have the longest /t/-closure but the shortest /s/-duration, all of which cue *it's dill* (as was shown in Experiment II). For each step towards the positive VOT values, /s/-duration increases by five msec, while stop duration closure decreases by five msec until the VOT value of +23 msec where we have the shortest stop closure but the longest /s/-duration, all for indicating *it still*. Appendix E shows the different values for each stimulus item.

4.3.2 Listeners

Five listeners were used, two females and three males. All are native speakers of Canadian English. One of them participated in Experiment I.

4.3.3 Procedure

Instead of recording the material on tape for presentation, this experiment was done interactively with the PDP-12 computer, in which an Alligator program would send the stimulus to a remote listening station and also collect the responses of the listeners. The program would wait until all listeners would give a response before presenting the next stimulus item.

Instructions are the same as those in Experiment I and some practice was allowed in the first session. Listeners came for three to five sessions (the number of trials per stimulus is given as 'N' on the figures showing the results for each subject). During the session subjects first did one identification task, which had 10 trails per stimulus item, and then did the discrimination task, which had 16 trials for each comparison. Some listeners did another discrimination task in the same session. White noise was not added to these stimuli since the problem occuring with the tape recorded items did not occur here. The identification task took about six minutes and the discrimination task took about eighteen minutes. It should be noted the stimulus conditions will not be quite comparable to those of Experiment I.

4.3.4 Results and Discussion

Figures 8 to 12 show the identification and discrimination curves for each subject. Most of the subjects (JK,KL,MW) have their crossover at approximately -13 msec VOT. The crossover for DP (Figure 8) is at -9 msec VOT and for GO (Figure 11) it is at -16 msec VOT. For all five subjects the chi-squared test of goodness of fit showed no significant difference between obtained and predicted discrimination (see Appendix F). But even though the slopes of the identification curves are a little sharper than in Experiment I, the discrimination curves peak at approximately only 75% correct for both obtained and predicted. As a result, it could be argued that the first of the four conditions set by Studdert-Kennedy et al. (1970, see Section 2.3), requiring sharp sudden crossovers, has not been met. One of the reasons for this situation is the gradualness of the slope of the identification curve corresponding to VOT values -9 msec and 0 msec. This problem is less severe than in Experiment I, due to the effect of the supporting durational cues. In any case, the fit between obtained and predicted curves (especially for the listeners DP and JK) demonstrates categorical perception where labels are being used for discrimination between categories while discrimination within categories is poor.

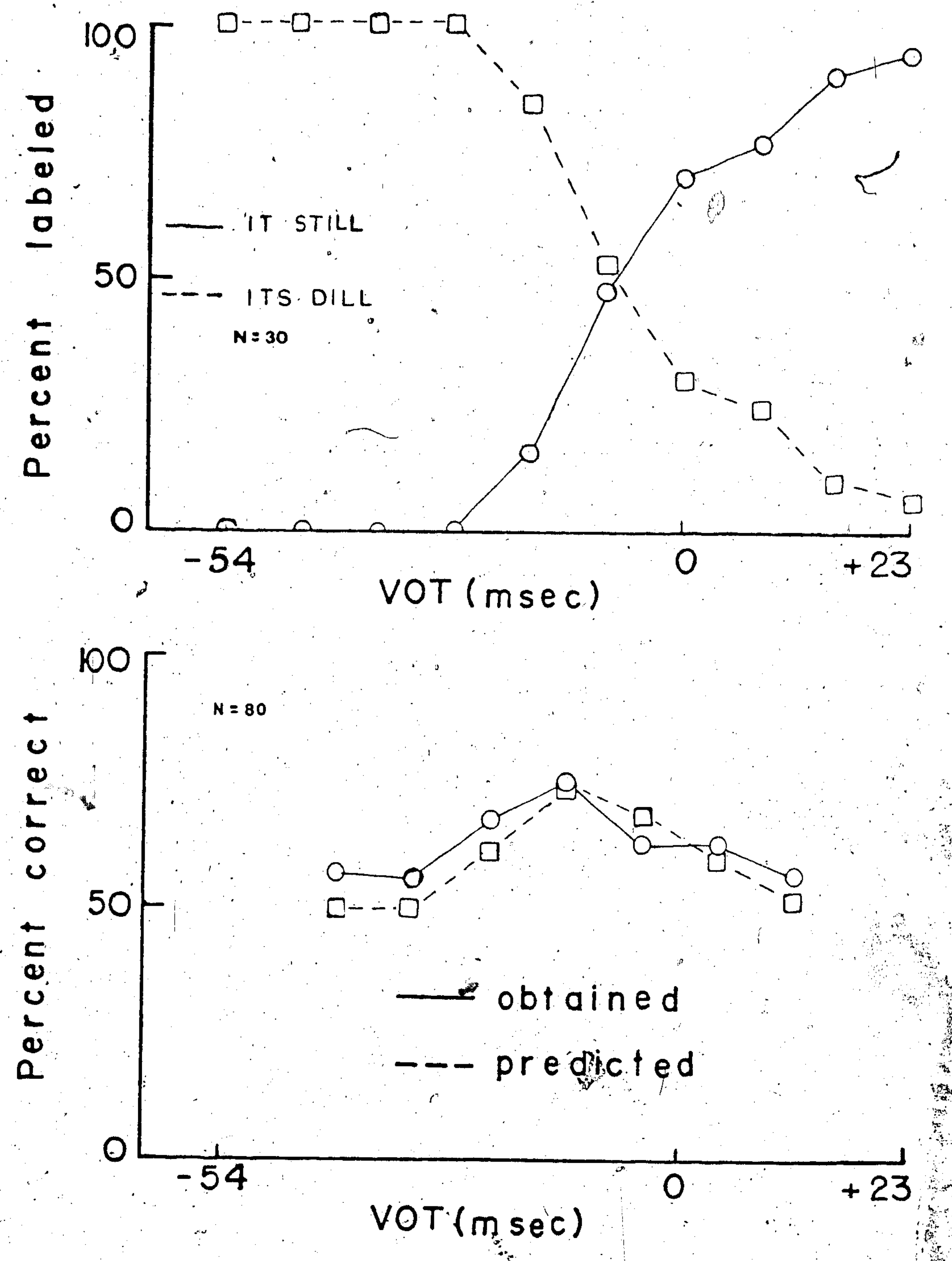


Figure 8: Identification and Discrimination for the two-way distinction in Experiment III (listener: DP).

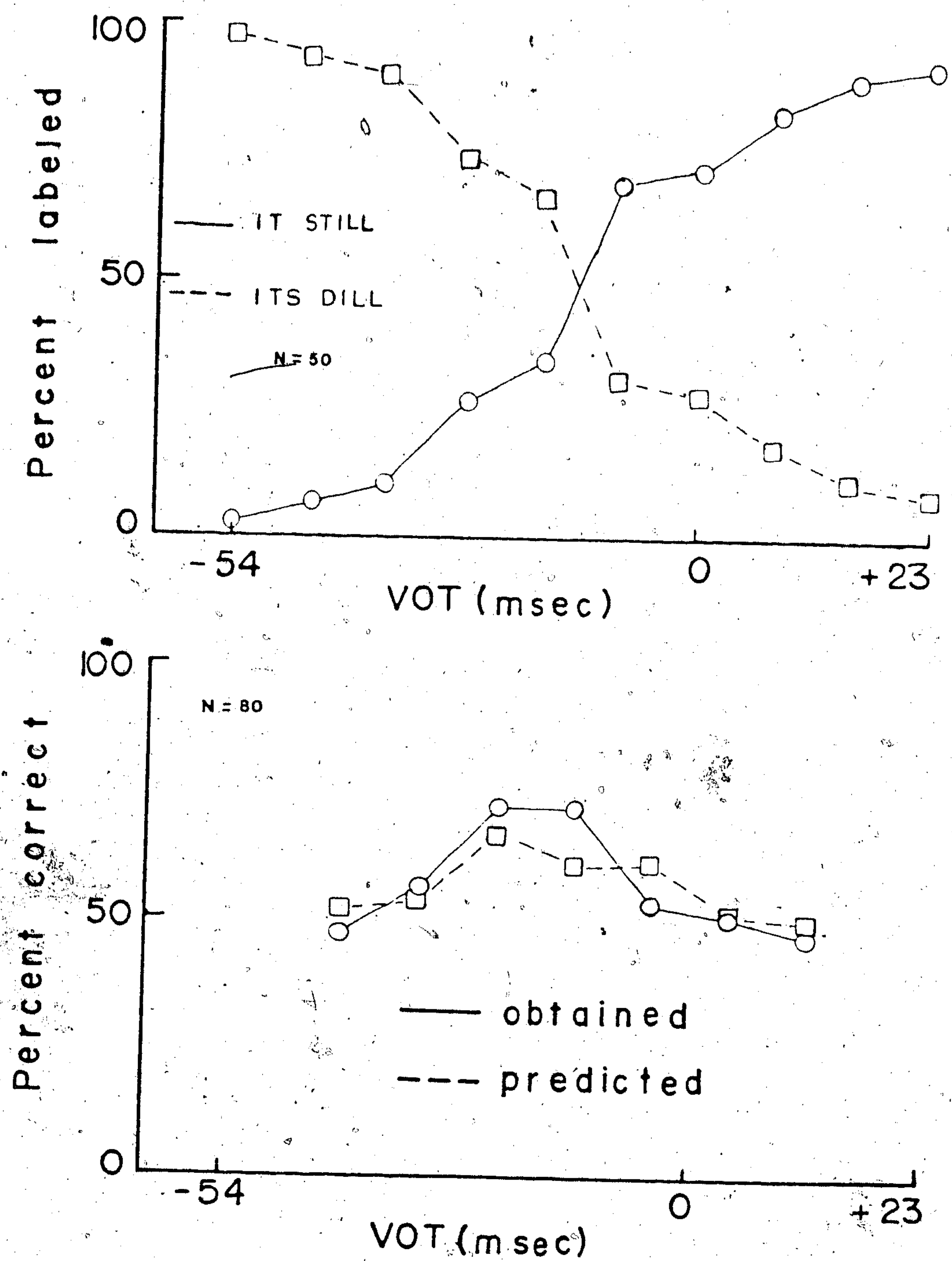


Figure 9: Identification and Discrimination for the two-way distinction in Experiment III (listener: MW).

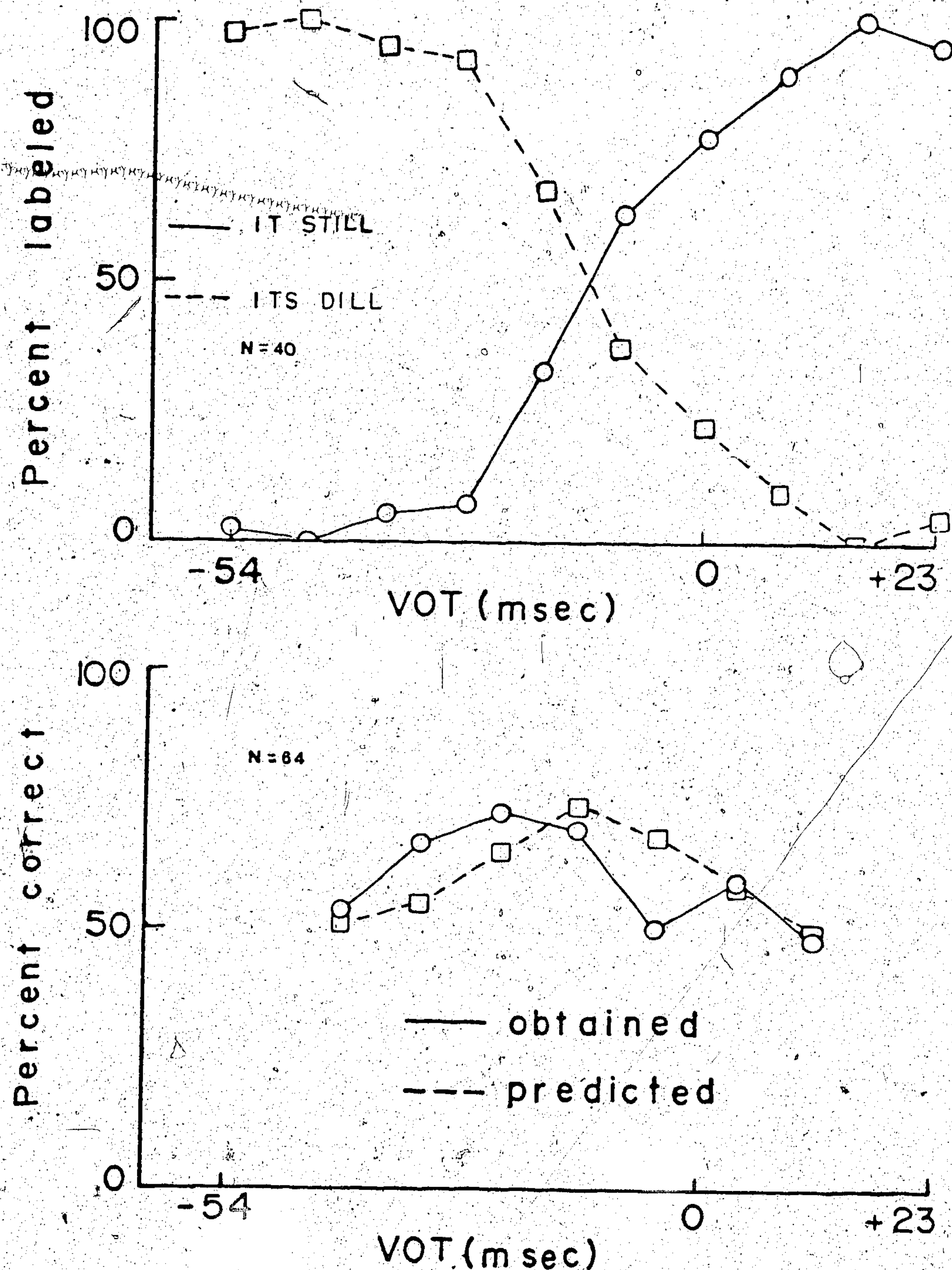


Figure 10: Identification and Discrimination for the two-way distinction in Experiment III (listener: KL).

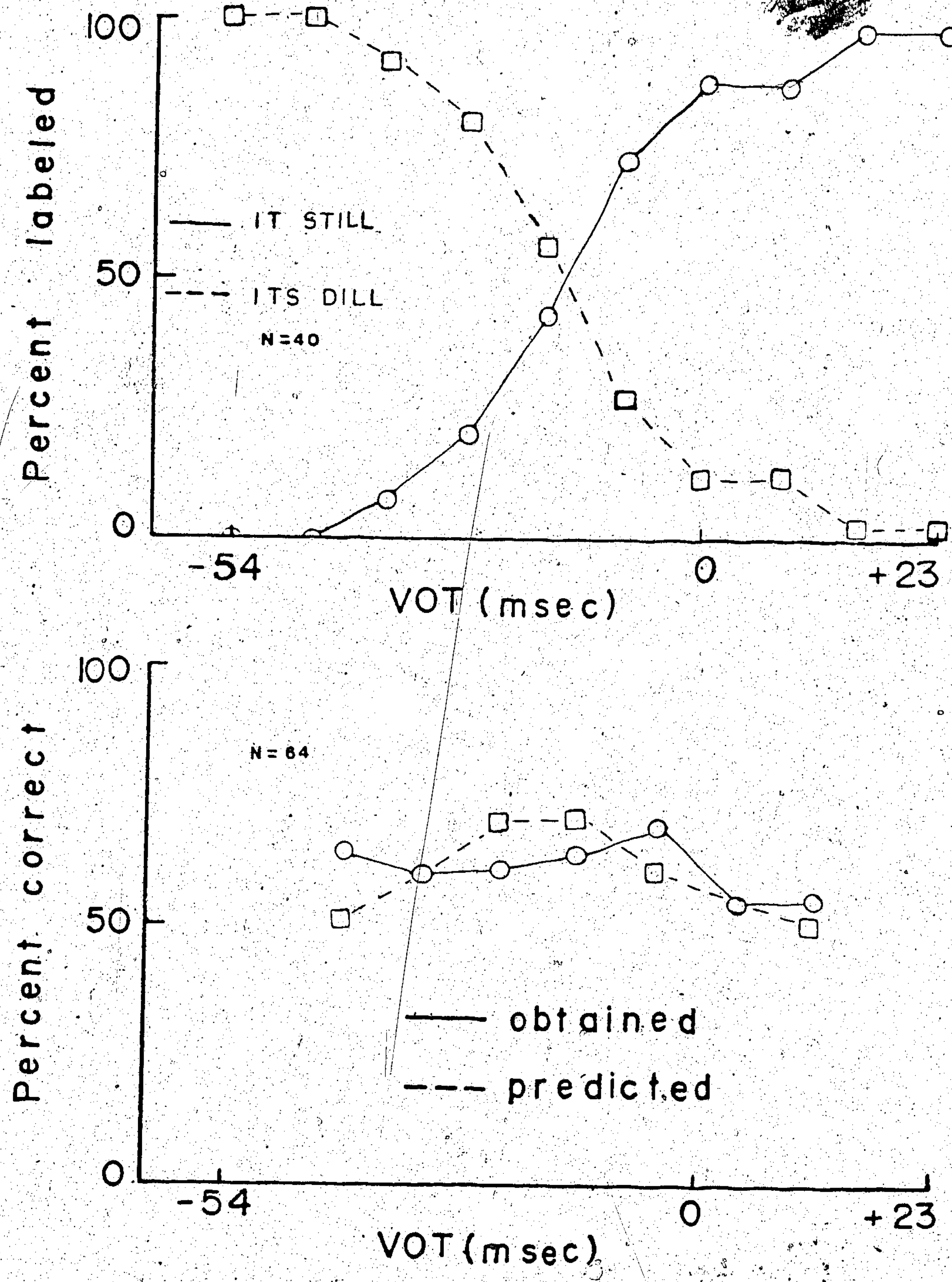


Figure 11: Identification and Discrimination for the two-way distinction in Experiment III (listener: GO).

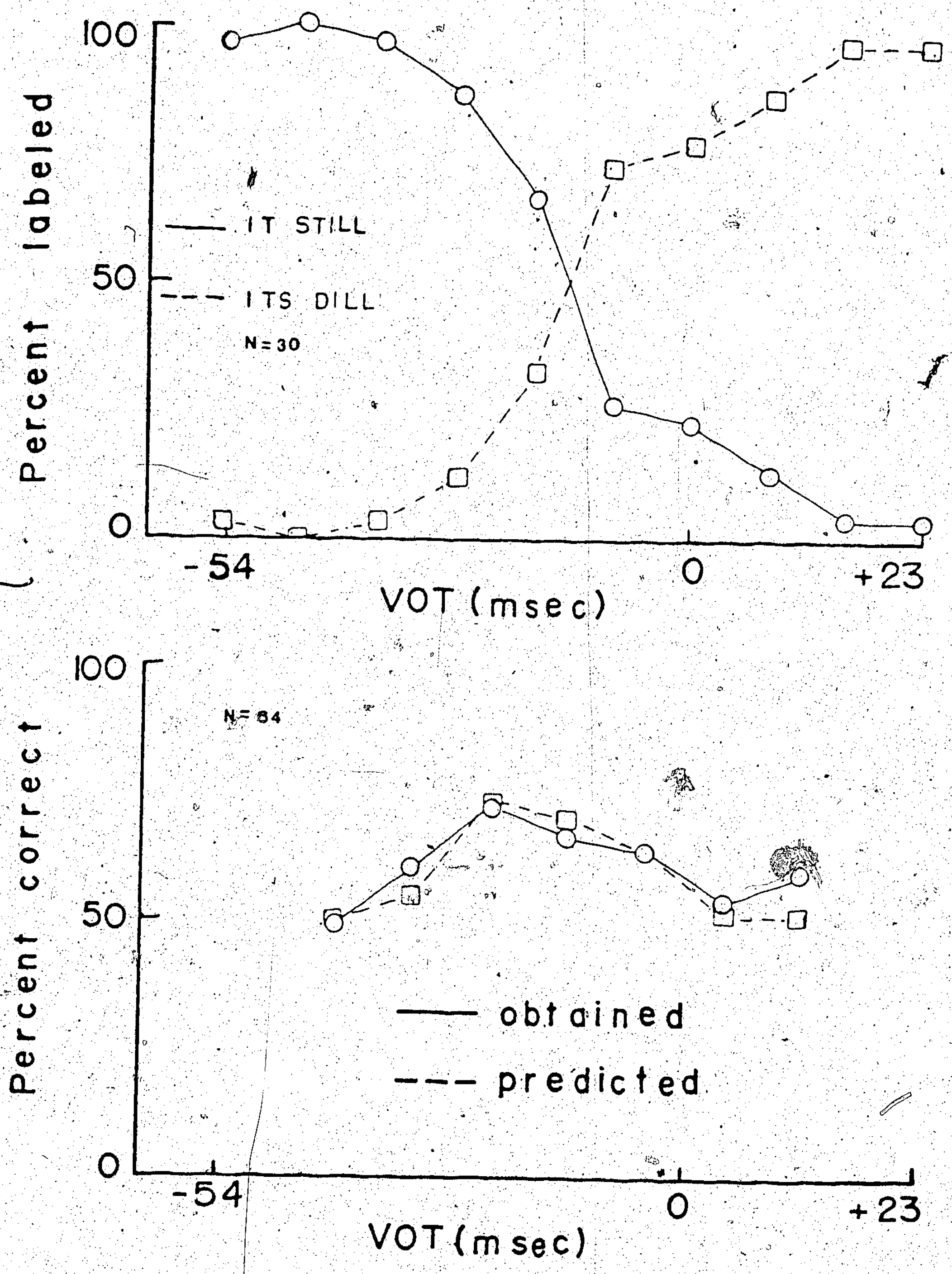


Figure 12: Identification and Discrimination for the two-way distinction in Experiment III (listener: JK).

It would be easier to interpert the results if the slopes were sharper and also if there were more stimulus points, and this would have been desirable, however collecting such an amount of data would have been difficult. All of the subjects from this experiment, and many from Experiment I, commented on the difficulty of the task. They found it very tedious yet requiring a great deal of attention. Because of this the number of stimuli was kept as low as possible. Increasing the range to include more within category comparisons would have made the task even more boring, allowing the listeners attention to drift away. Also, some subjects complained of auditory fatigue effects where changes in vowel quality and place of articulation (bill instead of dill) were percieved.

Having demonstrated categorical perception for the boundary between *it's dill* and *it still* with a combination of allophonic and durational cues, the next experiment was a short study to see if two of the subjects showing the categorical perception in this experiment would do the same for a continuum differing in VOT only, with all other cues being held constant. Experiment IV was more comparable to Experiment III in terms of signal conditions than Experiment I was, since Experiment IV was done interactively (and therefore without having white noise added) as was Experiment III. With the step-size being increased Experiment IV again tried to compare between the two different boundary conditions found in the lead and lag regions of the VOT continuum.

4.4 Experiment IV

For this experiment the full VOT range from Experiment I was used but in order to keep the number of stimuli down to a managable level, only every second stimulus item from VOT values -81 to +77 msec is used (see Appendix C). Also, as in Experiment I, only the VOT dimension was varied but not the other cues as was the case in Experiment III. The percentage of correct discrimination in Experiment I was fairly close to chance level for the negative VOT region and, since we do not have the extra complementary cues as in Experiment III, the step-size for this experiment was increased from three steps to four steps (to approximately 36 msec).

The procedure was the same as that for Experiment III and the two subjects who gave the best fit of obtained to predicted discrimination values from that study were used here. They were DP and JK but it should be noted, as it was not realized until too late, that DP has some second language experience with Punjab in which prevoicing plays a role in contrasting different types of stops.

4.4.1 Results and Discussion

Results from this experiment are shown in Figures 13 and 14. The crossovers for DP are at -22 and +45 msec of VOT for the two boundaries and for JK they are at -16 and +50 msec. In the chi-squared test of goodness of fit DP showed no significant difference between the obtained and predicted discrimination curves, while JK did show a significant difference (Appendix F). Comparison pairs where positive VOT values are involved show very good discrimination performance well above the predicted values. When the chi-squared test is performed on just the part of the curves that are in the positive VOT range[6] the test reveals a significant difference between the obtained and predicted values for both subjects, while the negative region showed no significant effect for either subject. One pair especially above the predicted values is the comparison between stimuli with VOT values of +6 and +41 where obtained values are at 86 and 79 percent for DP and JK, respectively, while the corresponding predicted values where 68 and 58 percent, respectively. Note that for the discrimination curve in the environment involving only the two way distinction in Experiment I, the peak was, as predicted, at the comparison of +14 msec and +41 msec VOT but there was

----

[6]The positive VOT values seemed to show above predicted discrimination while negative VOT values seemed to be slightly below predicted, indicating possible differences in the way discrimination is being made. The differences between predicted and obtained values seemed larges[illegible] the positive VOT range, and chi-squared test were done [illegible]ately on the two different regions (see Appendix F).

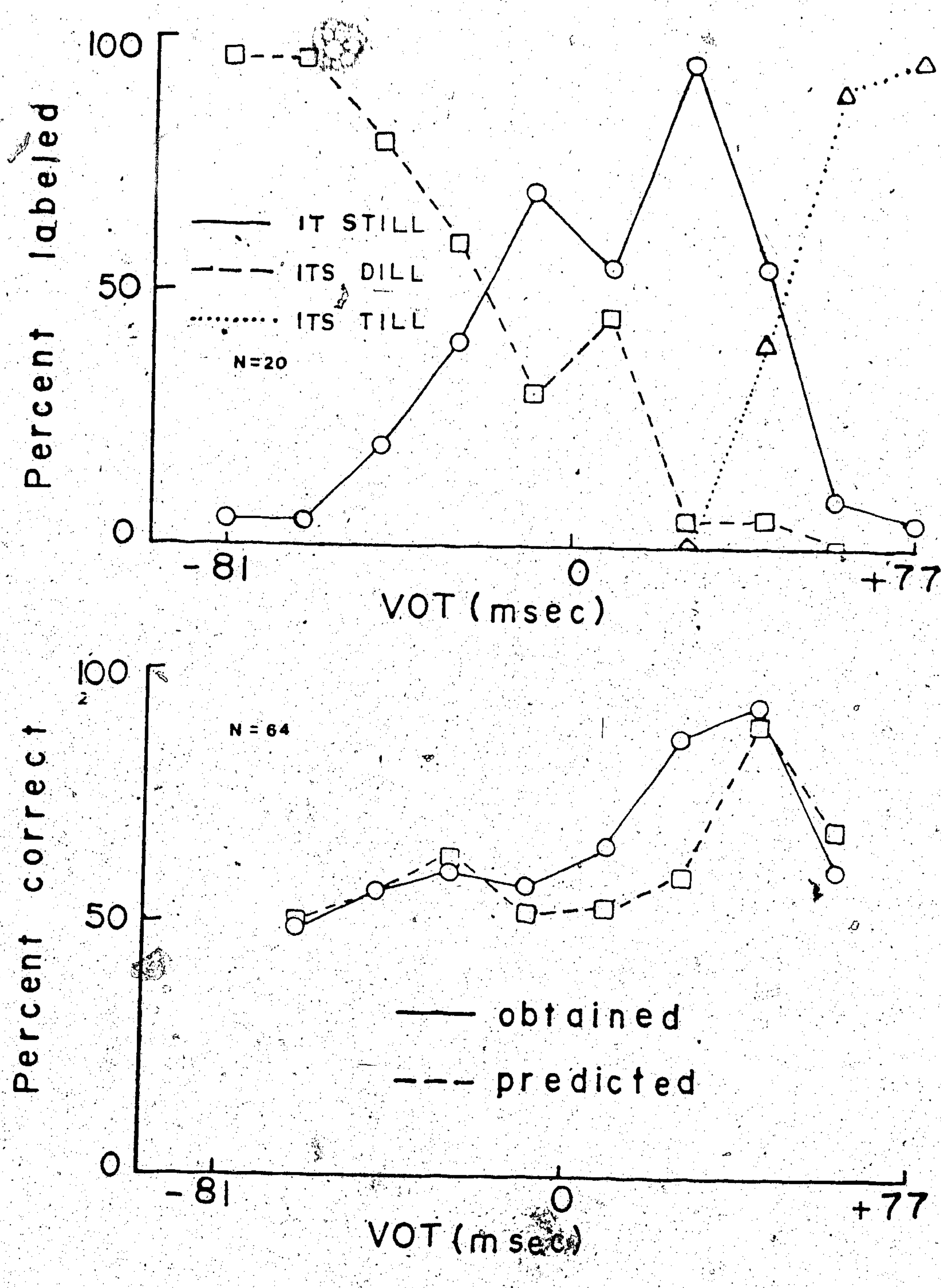


Figure 13: Identification and Discrimination for the three-way distinction in Experiment IV (listener: DP).

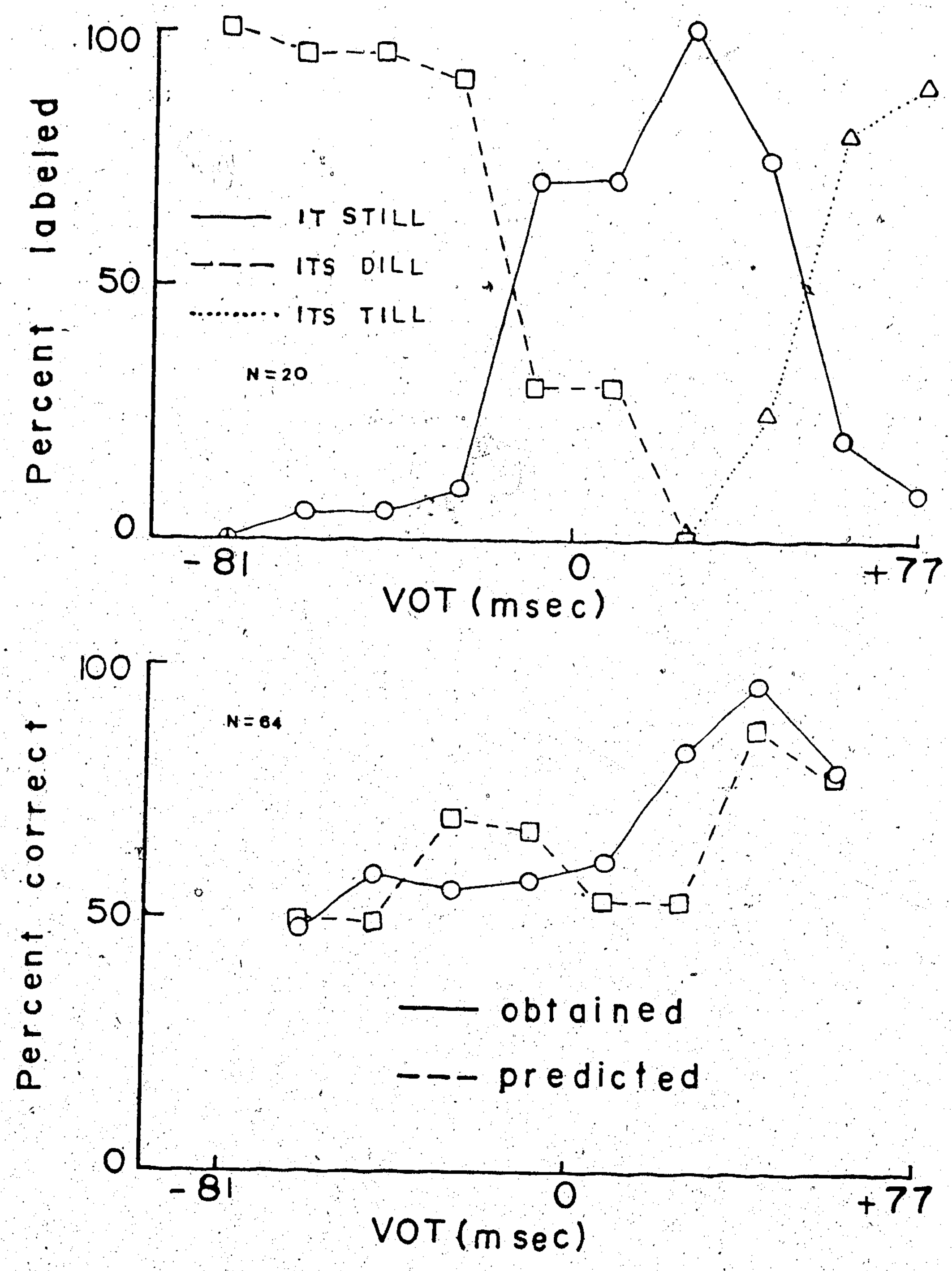


Figure 14: Identification and Discrimination for the three-way distinction in Experiment IV (listener: JK).

also very good discrimination between +6 msec and +32 msec VOT although it was not predicted.

We have found in Experiment IV enhanced discrimination of two signals, which would usually be on different sides of a /d/ and /t/ phoneme boundary in a normal discrimination experiment with this same VOT continuum. In the particular context of the three-way experiment the stimuli with +6 and +32 msec VOT are now both found to be members of the unaspirated /t/ category (i.e. the *it still* category). This is due to the fact that the task requires the identification of three categories instead of two and consequently the crossover values for the categories are changed. Thus these two stimuli are no longer separated by a category boundary. Perhaps the extra enhanced discrimination, which is independent of the category boundary, is due to some psychoacoustic factor (such as sensory threshold)[7] that is operative in determining phoneme boundaries in the normal phonemic category condition and which is still in effect in the more complicated condition of Experiment IV.[8] This tuning to a specific factor in the signal component may be a strategy easily adopted by a listener in the task conditions of these experiments where he or she is hearing similar

----

[7] See section 3.3 for reference to Pastore et al.(1977), where a critical duration of 15 to 25 msec. is mentioned. The pairs of stimuli which discriminate well, as being discussed here, are comparing items on either side of +25 msec VOT.

[8] Another notion that could be argued along these lines are detectors tuned to the output in the enviroment of natural speech.

signals over and over again, and having much of the top-down processes trivialized and also hearing auditory or speech distortions (such as adaptation effects) of everything but the changing components. This may cause the listeners to attend to more acoustic differences than is usual in the speech mode. As mentioned, the two-way distinction in Experiment I also had a point well above the predicted values. This may still be explained in terms of sensory threshold where a critical duration of aspiration, for example, is involved.

As for the discrimination on the negative end of the VOT continuum, the chi-squared test of goodness of fit did not show any significant differences. However, it should be noted that none of the points of obtained discrimination in the negative part of the VOT range have a correct percentage of discrimination performance that has the lower limit of its confidence interval above the chance level.[9] A situation that appears to demonstrate the lack of using categories in the negative VOT region but of "auditory discrimination" in the positive VOT range is Figure 14 showing the data from JK. The predicted values for both boundaries are not that far different from each other, where it is 70% for the negative VOT area and 78% for the positive region. However the difference between the obtained discrimination at these points is much greater, 57 and 92 percent respectively. The

---

[9]As mentioned before, the correct discrimination would have to be above 64% when N is equal to 100.

results indicate that the subject is not using the labels with much success in the negative VOT range but is discriminating more than just labels in the positive VOT range.

In conclusion then, the results from Experiment IV appear to be, to a large degree, due to discrimination of signals in auditory memory as well as through phonetic labels. For stimuli with positive VOT, there is much information available, such as F1 cutback, formant transitions and aspiration, that can contribute to the better than predicted discrimination. As for stimuli with negative VOT, there is just the voice bar which is just not a good enough cue to give us distinct enough categories.[10] As a result the boundary between *it's dill* and *it still* is very unstable.[11]

---

[10] In the course of listening to ordinary speech the frequency of occurrence of the voice bar is intermittent at best. This was also reflected in the measurement study. Hence, listeners expectancies for the voice bar cue would not be high.

[11] For example, in Experiment IV, the indentification task was done in two sessions. For DP the boundary for *it's dill* and *it still* differed by 22 msec between the two sessions.

5. Summary and Discussion

By manipulating the same signal components in two different contexts, we tried to look at how the perception of juncture distinctions might be similar or different from the perception of ordinary phonemic distinctions in identification and discrimination tasks. Perceptual experiments were preceded by a study of production data, where the items *it's till*, *it's dill*, *it's still* and *it still* were measured for the duration of different signal components in order to see what might play an important role in distinguishing the four different utterances. Delay of voicing in stop consonants came up as the most significant cue. It is the only spectral cue that appeared and it distinguished *it's till* from the rest of the sentences which were distinguished by different combinations of /s/ duration and pause duration. The voice bar did not turn out to be that common, occuring in only five of the 36 *it's dill* tokens.

Prevoicing is rare in English, unless being carried through from a previous vowel context, and it does not appear to serve an important function in cueing linguistically relevant distinctions, at least for stops. Identification tasks, such as the 'dill' vs 'till' task in Experiment I, have demonstrated this where, on the VOT continuum, prevoicing was well within the category for voiced stops. However a second identification task between the categories *it's dill*, *it still* and *it's till* produced an extra distinction along the VOT continuum. There was a new boundary in the prevoiced range, although it is far from being as sharp and well defined as the boundary in the aspirated range. Also there was little demonstration of enhanced discrimination at this new boundary, but this may have been due to the pooling of the data across subjects and of the slopes at the crossovers being too shallow. At the other boundary, involving *it still* and *it's till*, there was good discrimination and it was even better than predicted from identification. This boundary involved a change in juncture but supposedly none in phonemic category.

The intention of Experiment III was to try and make the crossover boundary between *it's dill* and *it still* sharper and to see how well we could predict the discrimination from the identification results. To do this it was decided to control other cues to reinforce the category change being marked by VOT. Experiment II tested how VOT operated in different combinations of the three segmental durations T, S, SP (that is, the two stop-closures and the /s/-duration, see Figure 2). The results showed that /s/-duration and the duration of the stop-closure of the second word (SP) were, along with VOT, the most important duration cues in cueing the difference between *it's dill* and *it still*. In Experiment III somewhat sharper boundaries were obtained and none of the five subjects showed a significant difference between the obtained and predicted curves while two of the subjects showed particularly good examples of discrimination via categories. The last experiment again tested the whole VOT range without extra cues but the step-size was made one step larger. Only two subjects were used. In the results, positive VOT showed better discrimination than would be perdicted from identification, while the negative VOT range appeared to show discrimination poorer than predicted.

The case that involved a change in allophones was the stronger, more sharply deliniated boundary with good discrimination among stimuli near the boundary. The other category boundary which had a change in phonemes was not as well defined and had poor discrimination between the stimuli from different sides of the boundary. These results might seem surprising described in this manner, though it should be admitted that the one boundary (between *it's dill* and *it still*) is cued in an uncommon manner, namely by the voice bar. The interpretation of the results are complicated by the way the experimental conditions allow the cues to function differently from their traditionally designated role. But, this was more of an exploratory study of discrimination between categories involving distinctions in word boundary and, it appeared that the categories distinguished by a phonemic difference were not being fully used, while at the other boundary (between *it still* and *it's till*), more than just categories were being used (that is, some of the discrimination appeared to be due to comparison of timbers in auditory memory as proposed by Fujisaki and Kawashima, 1970). The effects may be largely attributable to some underlying psychoacoustic basis. An "explanation" of this type is often given for contrasts along the VOT continuum where it is proposed that there is a sensory constraint (a critical duration of about 15 to 25 msec) that can be capitalized on as a general perceptual strategy. This would give improved discriminating ability at the +20 msec VOT region but also around the -20 msec VOT range.[12]

However, a psychoacoustics basis is not the whole explanation behind the results. Not all languages have their boundaries at plus or minus 20 msec of VOT. Thai, for example, has its boundary between aspirated and unaspirated stops at around +40 msec. Also, for English, the boundary between aspirated and unaspirated stops differs for place of articulation. While the boundary between word initial /t/ and /d/ is around +20 msec of VOT, it is at about +40 msec for velars. This difference reflects an actual difference between dentals and velars in speech production indicating the effect language experience also has in the way perceptual cues operate.

Whatever the reason, the cues that distinguish aspirated stops from unaspirated stops in English are so strong that it may have overshadowed many of the effects that we were looking for in the comparison of distinctions made phonemically to distictions made nonphonemically.

----

[12] The average of negative VOT values for crossovers in Experiment III, where other cues were involved, were at about -13 msec while for Experiment IV they were around -20 msec.

However, this may not necessarily be because of the experimental setup as was discussed in the results of Experiment IV. In Experiment III it was demonstrated that the *it's dill* and *it still* boundary relied on other segmental allophonic cues such as /s/ and pause. It is probable that the segmental durations (indicating allophonic differences in /s/ and whether a stop is initial or medial) are the important cues as indicated from the measurement studies.[13] Perhaps phonemic categorization does not play that important a role in speech perception. Most categorical perception studies with speech are done with a contrast that involves only one word and, therefore, any meaningful contrasts in categories will involve a change in phonemes. As a result, any change in category caused by manipulation of subphonemic detail is not independent of a change in phonemic detail.[14] However, when placement of word juncture plays a crucial role, we can demonstrate categorical perception with acoustic cues that do not usually mark phonemic distinctions. It appears that we are even able to change the phonemic category of the stop with

----

[13] An experiment like that of Experiment III on durational cues alone would have been useful in analysing their strength relative to the voice bar.

[14] Some attempts have been made to get native English speakers to distinguish between words with prevoicing and those without, but with words in isolation. That is, they tried to get listeners to distinguish a third, but nonlinguistic, category along the VOT continuum. Strange and Jenkins (1978) claimed that it is very difficult to do this. On the other hand, Aslin and Pisoni (1980) found it very easy to teach the distinction to English speakers, and feel that it depends on getting the listener to attend to the difference.

changes in /s/-duration. This was demonstrated in Experiment II where, if you look at Figure 6 and 7, and compare the two graphs with a short /t/-closure (T), and short silent period (SP), but different /s/ durations, there is a change in category at 0 msec VOT. This illustrates how the allophonic variation of one phonetic segment can change the phonemic category of an adjacent segment. We might wish to question, as Klatt (1979) does, whether the recognition of phonemes or phonetic segments are an intermediate step in the perception of words.[15]

It might be easier to understand the results of the preceptual experiments in this study if we ignore the factors of juncture and phonemic category and analyse the results in terms of how the second word as a whole was being perceived. Perhaps the reason why /t/-closure (T) did not play such an important role in Experiment II is because it was only the second word that was being attended to. In the tasks given to the listeners in the perceptual experiments, they may have attended mainly to the second word since all three distinctions required were made in the second word and only two distinctions were made in the first word. Asking for the identity of the first word might produce different results, such as having the effectiveness of the pause cues (T and SP) reversed. From Experiment III it is apparent

---

[15] Phonemes would still have psycholinguistic relevance in terms of speech production and an indirect role in speech perception as this would be important for such things as language acquisition.

that both the spectral and durational cues play a role in influencing the listeners categorization. It can be seen that the change in direction discussed in the results of Experiment I is still there, indicating that changes along the voicing continuum are having an effect. However, the prominence of the change in direction is being countered by the duration cues. The effect of categorical perception demonstrated for at least some of the subjects in Experiment III is not a result of a distinction between any one phonemic or allophonic contrast but between the percepts that result from the intergration of several subphonemic cues.

Bibliography

Abramson, A.S. & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In *Proc. Sixth International Congress of Phonetic Sciences*, Prague, Academia, 1970, 569-573.

Aslin, R.N. & Pisoni, D.B. Some developmental processes in speech production. In *Child Phonology Volume 2: Percetion* Eds. G.H. Yeni-Komshain, J.F. Kavanagh, and C.A. Ferguson. New York, Academic Press, 1980, 67-96.

Cutting, J. & Rosner, B.S. Categories and boundaries in speech and music. *Percept. and Psychophys.*, 1974, *16*, 564-570.

Eimas, P.D., Siqueland, E.R., Juszyck, D. & Vigorito, J. Speech perception in infants. *Science*, 1971, *172*, 303-306.

Fant, G. Analysis and Synthesis of Speech Processes. in *Manual of Phonetics* Ed. B. Malmberg. Amstrerdam North-Holland Publishing Company, 1968, 173-277.

Fry, D.B., Abramson, A.S., Eimas, P.D. & Liberman, A.M. The identification and discrimination of synthetic vowels. *Lang. and Sp.*, 1962, 5, 171-189.

Fujimura, O. Remarks on stop Consonants - synthesis experiments and Acoustic Cues. In *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fisher-Jorgensen*, Eds. L.L. Hammerich, R. Jakobson, and E. Zwirner. Copenhagen, Akademisk Forlag, 1971.

Fujisaki, H. & Kawashima, T. Some expeiments on speech perception and a model for the peceptual mechanism. *An. Rep. Eng. Res. Inst.*, 1970, *29*, University of Tokyo, 207-214.

Haggard, M., Ambler, S., Callow, M. Pitch as a voicing cue. *Jour. Acoust. Soc. Am.*, 1970, *47*, 613-617.

Hoard, J. Juncture and syllable structure in English. *Phonetica*, 1966, *15*, 96-109.

Jusczyk, P.W., Rosner, B.S., Cutting, J.E., Foard, C.F., and Smith, L.B. Categorical perception of non-speech sounds by two-month-old infants. *Percept. and Psychophys.*, 1977, *21*, 50-54.

Klatt, D. Speech perception: a model of acoustic-phonetic analysis and lexical access, *J. of Phon.*, 1979, *7*, 279-312.

Kuhl, P.K. & Miller, J.D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, *170*, 69-72.

Ladefoged, P. *Preliminaries to Linguistic Phonetics*. Chicago, The University of Chicago Press, 1971.

Lehiste, I. An acoustic-phonetic study of internal open juncture. *Phonetica*, 1960, *Supp. 5*.

Liberman A.M., Delattre P.C. & Cooper F.S. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1958, *1*, 153-167.

Liberman, A.M., Harris, K.S., Hoffman, H.S. & Griffith, B.C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psych.*, 1957, *54*, 358-368.

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, *33*, 42-49.

Lisker, L. The English stops after /s/ at word boundary; a three-way contrast. Unpublished manuscript. 1965.

Lisker, L. & Abramson, A.S. A crosslanguage study of voicing in initial stops - acoustical measuremnets. *Word*, 1964, *20*, 384-422.

Lisker, L. & Abramson, A.S. Some effects of context on voice onset time in English stops. *Language and Speech*, 1967, *10*, 1-28.

Lotz, J., Abramson, A.S., Gerstman, L.J., Ingemann, F., Nemser, W.J. The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: A tape cutting experiment. *Language and Speech*, 1960, *3*, 71-77.

McCasland, G. English stops after /s/ at medial word boundary. *Phonetica*, 1977, *34*, 218-228.

Miller, J., Wier, C., Pastore, R., Kelly, W. & Dooling, R. Discrimiantion and labeling of noise-buzz sequencies with varing noise-lead times: An example of categorical perception. *J. Acoust. Soc. Am.*, 1976, *60*, 410-417.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J. & Fujimura, O. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Percept. and Psychophys.*, 1975, *18*, 331-340.

Nakatani, L. & Dukes, K. Locus of segmental cues for word juncture. *Jour. Acoust. Soc. Am.*, 1977, *62*, 714-719.

Nakatani, L. & Schafer, J. Hearing 'words' without words: prosodic cues for word perception. *Jour. Acoust. Soc. Am.*, 1978, *63*, 234-245.

Nearey, T., and Hogan, J. Normative study of English initial consonants. *Humanities and Social Science Research*, Grant No. SSHRC-410-79-0312, 1979.

Nearey, T., Hogan, J. & Roszypal, A. Speech signals, cues, and features. In *Perspectives in Experimental Linguistics*. Ed. G. Prideaux. Amsterdam: John Benjamins, B.V. 1979.

Pastore, R.E., Ahroon, W.A., Baffuto, K.J., Friedman, C., Puleo, J.S. & Fink, E.A. Common-factor model of categorical perception. *J. Exp. Psych: Human Percept. and Perform.*, 1977, *3*, 686-696.

Pisoni, D.B. On the nature of categorical of categorical perception of speech sounds. Unpublished PhD dissertation, University of Michigan, 1971.

Pisoni, D.B. Auditory and Phonetic memory codes in the discrimination of consonants and vowels. *Percept. and Psychophys.*, 1973, *13*, 253-260.

Pollack, I., and Pisoni, D. On the comparison between identification and discrimination tests in speech perception. *Psychonomic sci.*, 1971, *24*, 299-300.

Raphael, L.S. Proceding vowel duration as a cue to the perception of voicing of word final consonants in American English. *Jour. Acoust. Soc. Am.*, 1972, *51*, 1296-1303.

Shammass, S.E. An experimental investigation of segment duration and intensity in English juncture. Unpublished MSc Thesis, University of Alberta, 1980.

Stevens, K.N. & Klatt, D.H. Role of format transitions in the voiced-voiceless distinction for stops. *Jour. Acoust. Soc. Am.*, 1974, *55*, 653-659.

Stevens, K.N., Liberman, A.M., Studdert-Kennedy, M. & Ohman, S.E.G. Crosslanguage study of vowel perception. *Language and Speech*, 1958, *1*, 153-167.

Stevenson, D.C. & Stephens, R.C. A programming system for psychoacoustic experimentation. *Proceedings of the 11th DECUS (Canada) Symposium*, Ottawa, Feb. 15-17, 1978.

Stevenson, D.C. Categorical perception and selective adaptation phenomena in speech. Unpublished PhD dissertaion, University of Alberta, 1979.

Strange, W., and Jenkins, J.J. The role of linguistic experience in the perception of speech. In *Perception and Experience*, eds. H.L. Pick, Jr., and R.D. Walk, New York, Plenum, 1978.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. & Cooper, F.S., Motor theory of speech perception: A reply to

Lane's critical review. *Psych. Rev.*, 1970, *77*, 234-249.

Umeda, N., & Coker C. Subphonemic Variations in American English. In *Auditory Analysis and Perception of Speech* Eds. G. Fant and M. Tatham. London, Academic Press, 1975.

Wajskop, M. & Sweerts, J. Voicing cues in oral stop consonants. *Journal of Phonetics*, 1973, *1*, 121-130.

Winer, B. *Statistical Principles In Experimental Design.* (2nd Edition) New York, McGraw-Hill, 1971.

Zatlin, M.A. Voicing contrasts: perceptual and productive voice onset time characteristics of adults. *Jour. Acoust. Soc. Am.*, 1974, *56*, 981-994.

APPENDIX A

Stimulus list used in measurement study

till

dill

still

it's till

it's dill

it's still

it still

it till

it dill

it spill

APPENDIX B

Means of the measurements for the raw scores and the square root transformations

it's till

	RS	SQRT
I	119.46	10.91
T	53.18	7.12
B1	10.72	2.84
S	131.11	11.38
SP	83.76	9.04
B2	76.26	8.45
IL	252.55	15.85

it's dill

	RS	SQRT
I	114.46	10.67
T	52.94	7.08
B1	10.65	2.79
S	119.47	10.85
SP	98.42	9.7[illegible]
B2	23.22	4.73
IL	288.09	16.94

it's still

	RS	SQRT
I	110.23	10.46
T	59.34	7.56
B1	12.75	3.37
S	192.70	13.74
SP	66.10	8.13
B2	29.20	5.36
IL	265.85	16.25

it still

	RS	SQRT
I	116.67	10.76
T	103.63	10.02
B1	9.94	2.71
S	142.46	11.88
SP	56.07	7.38
B2	29.14	5.36
IL	264.05	16.22

APPENDIX C

List of VOT values used for the stimulus used in Expeiment I

-90 msec

-81 msec

-72 msec

-63 msec

-54 msec

-45 msec

-36 msec

-27 msec

-18 msec

-9 msec

0 msec

+6 msec

+14 msec

+23 msec

+32 msec

+41 msec

+50 msec

+59 msec

+68 msec

+77 msec

APPENDIX D

Instructions to Experiment II

In this experiment you will be presented with a series of two word items. You are asked to identify the item as one of the three sentences given as possible choices. Once identified, give a judgement on a scale of one to five. This judgement should reflect the degree to which you feel that this particular item represents a CLEAR NATURAL PRONUNCIATION of the sentence type YOU HAVE CHOSEN.

Each item will be presented twice. Draw a line through the sentence you've chosen and write down your naturalness judgement next to it.

Use the following scale as a guideline in making your judgements. You will first do a short identification task on all of the different stimuli types used so you can get a general idea as to how you will use the range of the scale.

Naturalness
(With respect to chosen category)

1	2	3	4	5
DEFINITELY BELOW AVERAGE	SLIGHTLY BELOW AVERAGE	AVERAGE	SLIGHTLY ABOVE AVERAGE	DEFINITELY ABOVE AVERAGE

APPENDIX E

Stimulus values for Experiment III

VOT value	/s/-duration	Pause duration (SP)
-54 msec	118 msec	100 msec
-45 msec	123 msec	95 msec
-36 msec	128 msec	90 msec
-27 msec	133 msec	85 msec
-18 msec	138 msec	80 msec
-9 msec	143 msec	75 msec
0 msec	148 msec	70 msec
+6 msec	153 msec	65 msec
+14 msec	158 msec	60 msec
+23 msec	163 msec	55 msec

APPENDIX F

Chi-squared test of goodness of fit

Experiment I

2-way identification	X =27.97*	df=16
3-way identification	X =27.71*	df=16

Experiment III

Listener:

DP	X =2.39	df=6
MW	X =4.10	df=6
KL	X =5.68	df=6
JK	X =1.18	df=6
GO	X =4.73	df=6

Experiment IV

Listener:

DP		
full VOT continuum	X =11.38	df=7
lead VOT continuum	X = 0.49	df=3
lag VOT continuum	X =10.89*	df=3
JK		
full VOT continuum	X =14.50*	df=7
lead VOT continuum	X = 3.62	df=3
lag VOT continuum	X =10.88*	df=3

'*' - significant to the .05 level