THE UNIVERSITY OF ALBERTA

VOICING CONTRASTS IN FRENCH AND ENGLISH LABIAL STOPS

BY

MURRAY J. MUNRO

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 1987

ISBN 0-315-40975-4

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Murray J. Munro

TITLE OF THESIS: Voicing Contrasts in French and English Labial Stops

DEGREE: M. Sc.

YEAR THIS DEGREE GRANTED: 1987

Permission is hereby granted to the UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

M. J. Munro

(Student's signature)

#1009 - 10711 Sask. Drive
Edmonton, Alberta T6E 4S4
(Student's permanent address)

Date: October 3 1987

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 VOICING CONTRASTS IN FRENCH AND ENGLISH LABIAL STOPS, submitted by Murray Munro in partial fulfilment of the requirements for the degree of Master of Science in Speech Production and Perception.

....
Supervisor

....

....

....

Date 14 Oct 87

This thesis is dedicated to the memory of my mother, A. Lillian Munro.

Abstract

French and English word-initial labial stop consonants produced in four vowel contexts were compared in production and perception studies. In the production study, speakers of European French and Canadian English produced CVC nonsense words in a sentence frame. Duration and amplitude measurements were made of voice bars, silent intervals, bursts, and aspiration intervals. It was discovered that in French, differences between voiced and voiceless categories could be easily characterized in terms of voice onset time (VOT). Moreover, VOT in French was more dependent on the quality of the following vowel than was VOT in English. Although French /p/ was generally not as strongly aspirated as English /p/, noticeable aspiration occurred in many French /p/s produced in high-vowel contexts. A further difference between the two languages was observed in voice bar production. While French /b/ was characterized by long, uninterrupted voice bars, English voice bars tended to be shorter and to have noticeable voice breaks. Finally, a comparison of French /p/ and English /b/ revealed some small but consistent differences in VOT.

In the perceptual study, English-speaking listeners (experienced and inexperienced) were asked to categorize a set of natural tokens as either French /p/ or English /b/. Despite the small differences in VOT between these categories, most listeners were able to reliably separate the categories at levels above chance. Analysis using multiple regression revealed that the listeners may have used VOT and overall amplitude as the basis for their categorizations. These findings indicate that listeners may be sensitive to small differences between categories in their native language and analogous categories in a foreign language.

Acknowledgements

Above all, I would like to thank my supervisor, Dr. Terry Nearey, for his invaluable contribution to this thesis. Without the insightful advice he offered at all stages and the many hours he spent programming and helping me set up the experiments, this research would not have been possible.

Second, I owe much to my friends and fellow students for their criticisms and helpful input. Special thanks to Peter Assmann for making many useful suggestions on an earlier version, to Maureen Dow for reading and commenting on several drafts and always lending an ear, and especially to Tracey Derwing for giving much-needed support and encouragement as always.

Thanks are also due to my subjects (unfortunately too numerous to name), particularly those who listened without complaint (usually) to seemingly endless repetitions of /p/s and /b/s. Thanks especially to Devinder, Baljit, and Robert, and to Chloe, who fortunately chose not to participate in the experiments. Thanks also to Janet Schwegel for her help in formatting and printing the final copy.

I would also like to thank the members of my examining committee, Dr. Bernard Rochet and Dr. Kyril Holden, and my examining committee chairman, Dr. Bruce Derwing, for their useful comments and questions. I am indebted to Dr. Rochet for providing some of the recordings used in the experiments.

Finally, I would like to express my gratitude for financial support received in the form of Province of Alberta Graduate Scholarships and several teaching assistantships in the Department of Linguistics.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Cross-language studies have often constituted valuable contributions to the field of phonetics. For instance, Lisker and Abramson's (1964) important research on voice timing in a number of languages has led to a better understanding of the general mechanisms involved in the production and perception of stop consonants and has also sparked a great deal of controversy and further research. Delattre (1965, 1966), through his comparative work on English, French, and other languages, made a significant contribution to our understanding of the phonetic similarities and differences between languages with similar phonological inventories.

The particular area of concern in this thesis is that of so-called voiced and voiceless stop consonants in English and French. It has been observed that while superficially these languages have virtually the same stop inventories, there are some significant differences in stop production between the two in terms of phonological contrast. Much of the research on voiced and voiceless stop cognates in both languages has focussed on timing differences between categories within and between languages, though there is far from unanimous agreement on the precise role of voice timing in English and other languages (see Malécot, 1970; Winitz, Lariviere, and Herriman, 1975; Keating, Mikos, and Ganong, 1981). In general, English is known to contrast (in initial position) a set of aspirated voiceless stops having fairly long voice onset times with a "voiced" set characterized sometimes by short-lag voice onset times and sometimes by the presence of voicing during stop closure. French, on the other hand, contrasts a short-lag VOT category, generally described as unaspirated, with a category of fully-voiced stops. Although there has been extensive investigation, to say the least, of the voicing distinction in English, there have been relatively few studies (especially perceptual studies) of the analogous distinction in French. Furthermore, few researchers have undertaken comparative studies of stops in the two languages. There are, however, several issues which deserve further investigation. In particular, there are some inconsistencies in reports concerning voice timing and the presence of aspiration in French /p/. There is also some confusion regarding the relationship between French /p/ and English /b/. Traditional pedagogical descriptions (e.g. Delattre, 1951) seem not to point to any clear differences between the two, yet recent measurement studies suggest otherwise. A cross-language study such as the present one may prove valuable in that it may help us to understand in more detail the differences in stop production in the two languages.

In addition, cross-language perceptual experiments may improve our understanding of the perception and production of speech sounds in a second or foreign language. Flege (1984) has found evidence that a foreign accent can be detected by native speakers in short segments of speech lasting only a few milliseconds. He concludes that "human listeners are acutely sensitive to divergences from the phonetic norms of their native language" (p. 704). We can expect then that even monolingual speakers may be able to perceive small "subphonemic" differences between categories in their own language and analogous categories in a foreign language. The degree to which listeners are able to perceive the differences between phonetically similar sounds from their own language and those from another language (whether it is familiar or unfamiliar) has important theoretical (and possibly practical) ramifications. Specifically, it may help us to understand more fully how speech perception is affected by first language learning. For instance, some researchers (e.g. Pisoni, Aislin, Perey, and Hennessy, 1982) have recently attempted to determine the extent to which monolingual speakers retain their sensitivity to differences in voice onset time which are much smaller than those used in their native language. The present study also looks at this problem and focusses as well on the question of whether or not listeners can perceive differences in rise time which may reflect differences between the native and foreign language categories.

The remainder of this thesis is structured as follows. Chapter 2 is a summary of some of the recent literature on stop production and perception in French and English. Attention has been restricted to initial stops because of the vast literature available on stops in general. The results of two production studies of labials produced in word-initial position in sentence context are described in Chapters 3 and 4. Chapters 5 and 6 cover the results of two perceptual experiments. Finally, a discussion of the results is presented in Chapter 7.

Chapter 2: Production and Perception of Initial Stops: A Review of the Literature

Voice Onset Time in French and English

Lisker and Abramson (1964) showed that the difference between so-called voiced and voiceless homorganic stops in initial position in a variety of languages could be fairly simply characterized as a difference in timing between the release of the stop and the onset of quasi-periodic energy resulting from the vibration of the vocal folds. The role of this difference, commonly referred to as voice onset time (VOT), has been critically evaluated in a very large number of studies of speech perception and production in what is probably one of the most deeply researched areas of phonetics. Because of the vast number of such studies, it is not even remotely possible here to cover all of the issues relating to voice onset time. Instead, attention will be focussed on some of the major problems, and discussion will be restricted to studies of English and French stops in word-initial position.

Such studies have consistently shown that in English absolute initial /p t k/ the time of voice onset lags the time of the release of the stop by a relatively long duration while in /b d g/ a short lag or even a voicing lead is observed. As well, perceptual studies in which synthetic VOT continua were employed (e.g. Zlatin, 1974) have indicated that the time difference between a synthetic burst and the onset of periodic energy alone can serve as a powerful perceptual cue to the voicing distinction. In production studies of English word-initial stops within sentences, VOT measurements have been found to be less useful in characterizing the differences between the two categories (Lisker and Abramson, 1967). Here, some overlap is observed, even when some contextual factors known to affect VOT (e.g.

stress) are controlled for. Nevertheless, as Lisker and Abramson (1967) point out, it is consistently true that the *mean* VOTs associated with word-initial /p t k/ are longer than those associated with /b d g/.

There are some languages not included in Lisker and Abramson's (1964) study in which differences between stop cognates apparently cannot be characterized in terms of VOT. Jaeger (1983) argues for two types of languages: those in which VOT is relevant (in initial position at least) and those, sometimes referred to as "fortis-lenis" languages[1] (see discussion below) in which it is not. Among the studies apparently confirming the existence of the latter category is Caramazza, Yeni-Komshian, Zurif and Carbone's (1973) study of stops in absolute initial position before /a/ produced by ten speakers of Canadian French. Their measurements indicated some overlap in VOTs between the voiced and voiceless categories because some voiced tokens were produced in the short-lag VOT region occupied by voiceless /p t k/. Moreover, in their perceptual experiment, in which a synthetic VOT continuum was used, stimuli near 0 msec. of VOT showed ambiguous category membership. In contrast, the category boundary for English-speaking listeners was quite sharp and the number of ambiguous stimuli smaller. The Canadian French listeners' identification functions were described by the authors as "grossly non-monotonic," and it was concluded that VOT is "not phonemic" in Canadian French. However, a subsequent study by Caramazza and Yeni-Komshian (1974) indicated virtually no overlap in VOT between the two categories of stops produced (again in absolute initial position before /a/) by ten French speakers from Nantes, France. They did not, however, report a perceptual study. It was concluded that the role of VOT in French depends on dialect: VOT may be a primary perceptual cue in some dialects but not in others.

[1]Jaeger actually argues against the use of this terminology.

While Caramazza et al.'s first experiment does indicate that VOT may not be a unique perceptual cue to the voicing distinction in Canadian French, it is likely that the authors' case against VOT is overstated. In the first place, the "grossly non-monotonic" identification functions observed showed ambiguous judgments only over a relatively small range of stimuli (about -10 to +10 msec. of VOT for /p/, for instance). Although the general shape of their Canadian French subjects' identification functions differed from that of the English-speaking subjects, stimuli with large negative or large positive VOTs were always categorized by French-speaking listeners as voiced and voiceless respectively, a fact which indicates that Canadian French speakers *are sensitive* to voice timing differences in stops. Second, as in Lisker and Abramson's (1967) study of English stops in running speech, although an overlap in VOT was observed in the production study, the mean VOT values of the two Canadian French categories were quite different. No speakers ever produced French /p t k/ with prevoicing, and none produced /b d g/ with long positive VOTs. The actual overlap between the two categories was far from total. It was nothing like the extensive overlap claimed by Enstrom and Spörri-Bütler (1981) to exist in Swiss-German stops or that observed by Jaeger (1983) in Zapotec and Jawoñ, although it is perhaps comparable to the partial overlap between "fortis" and "lenis" unaspirated stops known to occur in Korean (Han and Weitzman, 1970). While no means for Canadian French voiced stops are given, it is fairly clear from Caramazza et al.'s histograms that, on the average, voicing begins far sooner in /b d g/ than it does in /p t k/. It seems premature, then, for Caramazza et al. to conclude that VOT is unimportant or "not phonemic" in Canadian French. In fact, their results are perfectly compatible with a perceptual algorithm which uses VOT as a primary cue and includes other secondary cues (see discussion below) just in case VOT is in the ambiguous range of

+/-10 msec. Another possibility is that VOT and some other cues are weighted non-hierarchically in listeners' decision criteria in such a way that VOT plays a major role in stop tokens with VOTs outside the ambiguous region, but a minor role in other cases. It is conceivable that the synthetic stimuli in this experiment lacked some cues which might have disambiguated the stops having VOT values within this range. Finally, it should be pointed out that in these two studies, all stops preceded /a/. Perhaps a wider range of VOT values would have been observed if other vowels had been included. It is known, for example, that VOT in English stops tends to be longer before high vowels than before low ones (Klatt, 1975; Smith, 1978). A similar phenomenon has been observed in European French (Fisher-Jørgensen, 1972).

Voiceless Stop Production in French

An interesting question arising from Caramazza et al.'s and Caramazza and Yeni-Komshian's findings concerns their measurements of the voice onset times of French voiceless stops. In the case of /pa/, for instance, they report a mean VOT of 18 msec. in Canadian French and approximately 12-13 msec. in European French (from their graph, p. 244), and maximum VOTs of as much as 25 msec. in both French dialects, findings which appear to conflict with those of early studies such as Delattre (1966), in which it is argued that French voiceless stops are unaspirated because the vocal folds are closed prior to the stop release (p. 13). It is not clear how a VOT as long as 25 msec. could occur with no aspiration or why glottal pulsing should begin so late if the vocal folds are in fact in the appropriate configuration for voicing even before the time of release. More recently, fibrescopic techniques (Benguerel, Hirose, Sawashima, and Ushijima, 1978) have revealed that the glottis is open for most of the closure duration in European French /p/ and that glottal adduction may sometimes occur a short time *after* release. While European French voiceless stops have usually been described as unaspirated (Delattre, 1965; Kim, 1970; Keating, Linker, and Huffman, 1983), and traditionally, practical handbooks of French pronunciation seem not to have identified any articulatory differences between French /p/ and short-lag English /b/[1], Benguerel et al. (1978) and Serniclaes, D'Alimonte, and Alegria (1984) describe French /p t k/ as slightly aspirated. The latter informally report a mean VOT of 22 msec. for initial French /p/ (though they do not indicate which vowels were produced by their subjects), a value which is considerably larger than the 1 msec. reported by Lisker and Abramson for absolute initial English /b/ (excluding tokens produced with prevoicing). Two important questions, then, which deserve further investigation concern the role of aspiration in French /p/ and the similarities and differences between French /p/ and English /b/.

Voiced Stops in French and English

European French /b d g/ are usually described as "more voiced" than their English counterparts since in French the vocal folds are known to close and begin to vibrate well before the consonantal release (Delattre 1966; Benguerel et al., 1978) and to continue vibrating until the moment of release. Serniclaes et al. (1984) give a mean VOT for French initial /b/ of -145 msec. Thus, French stops, unlike English ones, illustrate a true "voiced-voiceless" distinction, one which is defined by the presence or absence of glottal activity during the closure interval. English /b d g/ show two basic patterns: voicing may begin either some time before release (prevoicing) or shortly afterward (Lisker and Abramson, 1964). Flege (1982) has

[1]See Delattre (1951, p. 53-54). He states that glottal pulsing begins at the moment of release in French /p/. Also his diagrams illustrate no difference between French /p/ and English /b/. See also Nicholson (1927) and Valdman et al. (1970).

demonstrated that it is particularly risky to make inferences about the state of the glottis on the basis of these two patterns since in many cases in which glottal adduction occurs well before the release, no prevoicing occurs. His findings, based on stops produced by ten English speakers, indicate that the presence of prevoicing appears to depend on speaker idiosyncrasies and may occur in as many as half the voiced stops produced in absolute initial position.

One interesting phenomenon which has been observed in English prevoiced stops in running speech is a "voice break": glottal pulsing may occur early during the closure interval and then "die out" several msec. before the stop release, resulting in a period of silence (Warren, 1976). Why this phenomenon occurs in English prevoiced stops is not fully understood. It is believed that voicing is possible during a closure interval only if the supraglottal cavity expands (whether passively or actively) or if some air is allowed to escape through the nasal cavity or both. (See Warren (1976) and Abramson (1977) for a discussion of the relevant literature.) It is possible that in stops with voice breaks, this expansion reaches its maximum prior to the release, and because the supraglottal and subglottal air pressures become equal, voicing cannot continue. None of the research discussed here mentions the occurrence of voice breaks in European French. Since the present study focusses on consonants within sentences, an opportunity to investigate this phenomenon further is available. If voice breaks are not found to occur in French, their absence may result from some basic difference in the production of voiced stops in the two languages.

Kim (1970) takes a functional approach in an attempt to account for the lack of congruence between English and European French voiced initial stops by arguing that the absence of aspiration in French voiceless stops makes it necessary for prevoicing to be consistently maintained in /b d g/ to keep the two categories distinct. However, such an account fails to explain the apparent partial loss of VOT contrasts in Canadian French, in which some /b/s are produced in the short-lag VOT region already occupied by /p/ (Caramazza et al., 1973), unless one accepts the possibility that in Canadian French (but perhaps not in European French) perceptual cues other than VOT are available to listeners. Measurement data, in fact, do not seem to indicate that languages necessarily (perhaps even usually) make optimal use of VOT differences in distinguishing stop categories. Keating et al. (1983, p. 280) list over twenty-five languages which distinguish two homorganic stop categories in initial position. Of these, several contrast negative versus short-lag VOT categories while others have short positive versus long positive contrasts. None distinguish categories *exclusively* by using maximal differences in VOT. That is, none exclusively contrast negative with long positive VOT.

One final point should be raised here regarding differences between English and languages such as French which have a true voiced-voiceless distinction. Some researchers (e.g. Keating et al., 1981) have pointed out that from the point of view of someone listening to natural speech tokens, a change in VOT from a negative value to a positive one involves not only a quantitative difference in timing, but a qualitative shift as well. They point out that in Polish, which contrasts stop categories in a similar manner to European French, "the initial voicing contrast can be seen as one of voicing during closure versus absence of closure voicing, and no actual VOT value need be computed by a listener. That is, our proposal is that Poles do not use VOT as English speakers do, in the sense of a temporal interval between burst and voicing onset" (p. 1268). Similarly, Abramson (1977) has stressed that "VOT is not an acoustic continuum, although it may be viewed as an articulatory or physiological continuum" (p. 296). While cues other than voice timing may be important in Polish and French, and may even be more important than timing differences, the line of reasoning used by Keating et al. in contrasting Polish and English does not seem entirely satisfactory because qualitative changes do in fact occur in English as one moves from the short-lag to long-lag region. One might just as well argue (as Winitz, Lariviere, and Herriman, 1975 essentially do) that in English no actual VOT value has to be computed either and that the distinction can be seen as one of aspirated versus non-aspirated tokens. This problem is discussed in more detail below. In any case, there is a clear need for more perceptual studies in languages other than English in order to determine the role of VOT and other proposed cues in the perception of stops.

Other Perceptual Cues

Whether or not VOT is a primary cue in French, it is probable that other cues play a role in listeners' perceptions in this and other languages. Several additional properties of stops in English and other languages are likely candidates for "cue status." Researchers have investigated the roles of a number of other differences, including differences in fundamental frequency (Haggard, Ambler, and Callow, 1970; Han and Weitzman, 1970; Ohde, 1984), the presence or absence of formant transitions (Stevens and Klatt, 1974), the degree of aspiration (Repp, 1979), closure duration (Lisker, 1957; Wajskop and Sweerts, 1973), and phonation onset type (Han and Weitzman, 1970; Darwin and Pearson, 1982; Darwin and Seaton, 1983). In general, it is believed that some of these differences may operate as secondary cues in the perception of English stops with VOT functioning as a primary cue.

Aspiration noise

One of the most important of these cues is aspiration noise. Liberman, Delattre, and Cooper (1958) observed an increase in voiceless identifications by English speakers listening to a synthetic F1-cutback continuum, when F2 and F3 were filled with noise during the cutback portion. Nevertheless, such noise did not *have to* be present in order for subjects to identify voiceless percepts. More recently, Repp (1979) observed a time-intensity tradeoff for "aspiration" noise in perceptual experiments using synthetic continua. (See also Darwin and Seton, 1983.) Stimuli were perceived as more voiceless when the aspiration phase was increased in either duration or amplitude. A 1 dB increase in the ratio of the noise amplitude to the amplitude of the following periodic portion (resulting from either an increase in noise amplitude or a decrease in synthetic vowel amplitude) caused a "leftward" movement (i.e. toward the voiced end of the continuum) of the category boundary of about 0.43 msec. It appears, then, that his listeners evaluated not simply the presence or absence of aspiration noise but perhaps the total amount of aspiration energy in making their categorizations.

In English, aspiration and voicing are largely mutually exclusive phenomena: in an aspirated stop, the aspiration phase of the stop continues as air passes relatively freely through the glottal opening until the vocal folds are approximated and glottal pulsing begins. Therefore, the end of noise excitation of the vocal tract coincides with the beginning of periodic excitation, although there may be an overlap, and the duration of the aspiration phase is highly correlated with VOT. Precisely how VOT (or duration of aspiration) is controlled is not well understood. Kim (1970) has argued that aspiration is a "function of the glottal opening at the time of release of the oral closure of a stop." (p. 115) Since the glottis requires a certain amount of time to close down, a large opening results in longer aspiration than a small opening. This view contrasts with that of Lisker and Abramson (1967, 1971) who argue that the actual time of glottal closure is controlled.

In perceptual experiments some researchers have attempted to play off VOT and aspiration in order to determine which is the stronger. Winitz et al. (1975), for example, argue that presence or absence of aspiration noise functions as the primary perceptual cue to voicing in English and that VOT "operates as a relatively unimportant secondary cue" (p. 41). They claim to have altered only the VOT in a number of naturally-produced stop + vowel syllables by making separate recordings of the aperiodic and periodic portions and then playing the two portions back with silent periods between them in the case of natural voiced tokens (to increase VOT) and with portions overlapping in the case of natural voiceless ones (to reduce VOT). They were thus able to create /p/s with VOTs supposedly identical to those of /b/s and vice versa. In general, they observed that listeners' categorizations of the manipulated segments were not much different from their categorizations of the original ones and concluded, therefore, that VOT was a less important perceptual cue than aspiration noise because radical changes in the former did not result in changes in the subjects' perceptions. In fact, they went so far as to argue that "aspiration *rather than* the tracking of time is the significant feature in the perception of the voicing contrast for initial stop consonants" (p. 52, emphasis added).

Now, strictly speaking, Winitz et al.'s claims about aspiration noise do not necessarily follow from their data. In fact, virtually any of the known differences between English voiced and voiceless stops, such as differences in fundamental frequency or in formant transitions (these were not controlled for) could have functioned as the critical perceptual cue or combination of cues in the experiment. But a more serious criticism which can be levelled at their approach concerns the nature of their stimuli. In particular, there may be serious drawbacks to their method of cutting and splicing of tape in the preparation of stimuli. It is possible that their highly unnatural modified voiceless stops with aspiration noise and vowel occurring simultaneously or their voiced stops with long silent periods (often as long as 30 - 40 msec.) inserted between burst and vowel were not even perceived by the listeners as unified CV syllables. Finally, as noted earlier, Repp's results indicate that there is a temporal aspect to the perception of aspiration noise. That is, aspiration may not be perceived in a binary "on-off" fashion. If so, then it seems incorrect to claim that timing *per se* is unimportant in the perception of stops. Rather than regarding VOT and aspiration as making entirely independent contributions to listeners' perceptions of stops, it is perhaps more appropriate to regard them as at least partially equivalent. (See Darwin and Pearson, 1982.) One may think of both VOT and amount of aspiration energy (duration and amplitude) as indicating the time difference between the release of the stop and the onset of glottal pulsing. In addition, perhaps because of their different spectral properties, both aspiration noise and periodic energy may make independent contributions in the perception of English stops.

A similar argument may be offered in response to Keating et al.'s (1981) discussion (see above) of voicing in languages such as Polish and French. That is, the amount of voicing during closure in these languages may have some importance in listeners' perceptions. In addition, other properties, such as spectral characteristics, may be relevant.

Force of articulation

Some researchers have disputed the claim that differences in glottal timing are the chief means by which stop cognates are distinguished. Malécot (1970) and Marchal (1983), for example, have preferred to describe the /p t k/ -- /b d g/ distinction in French (and in other languages) in terms of "force of articulation," which, according to Malécot, refers to "...a synesthetic interpretation of magnitudes of intrabuccal air pressure..." (p. 1588). A variety of phenomena are argued to lead to a "mistaken proprioceptive impression" of greater force in the articulation of fortis stops than in lenis ones. Jaeger (1983) summarizes some of the relevant literature, pointing out four general categories of differences: air pressure differences, resulting in differences in the intensities of bursts, frication noise and aspiration noise; articulatory differences, resulting in greater force and degree of contact between articulators; differences in closure durations; and differences in voicing. Halle, Hughes, and Radley (1957) and Lubker and Parris (1970) discuss air pressure differences in English, and the former claim that since the presence or absence of voicing during stop closures is insufficient to distinguish /p t k/ from /b d g/, these stops should be classified as "tense" and "lax" respectively. In a study of French stop production, Marchal (1983) found a greater degree of linguo-palatal contact in /t k/ than in /d g/. Malécot (1970) points out that fortis consonants in French are generally characterized by longer closure durations than lenis ones. Finally, it is well known that in many languages, the presence of voicing during closure often distinguishes /p t k/ from /b d g/.

Unfortunately, some of the research in this area suffers from the serious drawback of extremely small sample sizes. For instance, Marchal (1983) cited above used a subject pool of two. Furthermore, while there may be some reliable correlates of the fortis-lenis distinction which can be measured in natural speech, it is not at all clear what advantages there are to invoking what appears to be a rather vaguely-defined property of consonants in an attempt to characterize differences between stop categories. (See even Marchal, 1983.) In particular, as Lisker and Abramson (1964) argue, in languages such as English, a fortis-lenis description does not seem to account for any properties which cannot be explained in terms of glottal timing. While some researchers (Hardcastle, 1973) have insisted that fortis-lenis or tense-lax features are necessary in descriptions of some other languages such as Korean, others (Jaeger, 1983) have disputed fortis-lenis accounts which were previously accepted for such languages as Zapotec and Jawoñ. What is clear at the present time is that considerably more research is required, particularly in those languages having contrasts which are apparently not explicable in terms of VOT.

Overall intensity

Darwin and Pearson (1982) and Darwin and Seton (1983) have been careful to distinguish between the time of the physical onset of voicing in a stop consonant (or, in equivalent terms, the time at which aspiration ends) and the *perceived* time of voice onset. Studies in the perception of musical tones (Vos and Rasch, 1981) have established that the perceptual onset of a tone occurs after its physical onset, the degree of difference depending on the rise time of the stimulus. Vos and Rasch observed that a tone is generally perceived to onset at the time at which its intensity reaches a level 15 dB below maximum. By manipulating overall intensity levels independently of VOT in synthetic stimuli on several /pa/ - /ba/ continua, Darwin and Pearson were able to identify a similar phenomenon in speech perception. The starting point for their investigation was the observation that in natural speech the point of voice onset is marked by a number of phenomena, including the appearance of periodic energy and strong low frequency components, and an increase in energy. They presented to their listeners several synthetic continua in which the times at which these events occurred were varied independently of each other and of rise time. In general, the positions of the subjects' category boundaries were best predicted not by the time at which periodic energy first appeared in the stimuli or by most of the other correlates of voice onset, but by the time at which the overall intensity of the stimuli reached a level about 20 dB below the maximum intensity of the "vowel" phase. Specifically, Darwin and Pearson found evidence that none of the following coincided with the time at which voicing was perceived to onset: the point at which periodic energy onsets, the point at which spectral balance changes in favour of low frequency energy, and the point at which maximum intensity is reached. Also included in their stimulus set were synthetic continua, comparable to segments of whispered speech, which consisted of noise excited formants containing no periodic energy at all. Despite the absence of "voicing," some of these stimuli were perceived as /b/s. Again, the best predictor of the subjects' category boundaries was the time at which a critical overall intensity level was reached, although this level was somewhat higher than the critical level for the periodic stimuli.

While complete confirmation of Darwin and Pearson's findings using natural stop tokens as stimuli is probably not possible (because some of their tokens may represent impossible configurations of the human vocal tract), it is possible to use naturally-produced stops differing in rise time but not in VOT to test their claims. Some further experiments using natural speech tokens and a minimum of manipulation are included in the present study. (See Chapters 5 and 6.)

Darwin and Pearson's results may prove useful in accounting for the apparent failure of VOT to completely separate stop categories in some languages. It is conceivable that in such languages the type of phonation onset plays a significant role in listeners' perceptions. In particular, although there may be a consistent overlap in measured VOTs between stop cognates, the *perceived* times of voice onset may not overlap or may overlap less than measured voice onset time because of differences in phonation onset. In Korean, for instance, which has three voiceless stop categories contrasting in initial position, it is apparently not possible to characterize the differences between categories in terms of VOT alone because there is some VOT overlap (Lisker and Abramson, 1964; Han and Weitzman, 1970; but cf. Hardcastle, 1973, who found no such overlap). Nevertheless, on the average there are consistent differences in mean VOTs for these categories. Han and Weitzman, as well as Hardcastle, have argued that at least part of the difference between the so-called fortis and lenis categories may be described as a difference in intensity characteristics at the onset of voicing, although fundamental frequency differences may be important as well. According to Han and Weitzman's production study, periodic energy onsets more gradually in the lenis stops than in the fortis ones, in which voicing abruptly reaches its maximum level. They were also able to find evidence of the perceptual relevance of rise time by using a tape cutting and splicing method similar to that used by Winitz et al. (1975). They concluded that "the difference between weak and strong stops is a function of the slope of the leading edge of the intensity contour during the first few centiseconds of voicing following stop release ... If the slope rises abruptly, then the stop will be heard as strong, and if the slope rises gradually with the same time interval, then the stop will be heard as weak" (p. 127).

The problem of the lack of a simple correspondence between physical measurements, such as VOT, and their perceptual correlates is not a new one. Lisker and Abramson (1964) discuss at length their difficulties in evaluating the perceptual relevance of "edge vibrations" in making VOT measurements of non-initial stops from spectrograms. These vibrations are seen as "faint vertical striations at glottal rates near the baseline of the spectrogram below the clearly aperiodic high-frequency noise of aspiration" (p. 416). While an investigator might be tempted to analyze these as marking the moment of voice onset, they are normally not audible and therefore apparently have no relevance to stop perception.

Psychophysical Basis for the Perception of VOT

A general psychophysical account of how listeners may make use of VOT in the perception of stops is offered by Pisoni (1977), who argues that languages may take advantage of "naturally defined regions of high discriminability" (p. 1353) in the placement of their VOT boundaries. In particular, his experiments reveal that in order for English-speaking listeners to correctly identify the temporal order of two pure tones differing in frequency and onset time, the onset of one tone must lag that of the other by at least 20 msec. His naive subjects were easily trained to identify three categories of tone stimuli on a continuum: one with the low tone leading by at least 20 msec.; one with the low tone lagging by the same amount, and one in which the difference in onset times varied between +20 and -20 msec. Apparently the tones in the latter category were perceived as having simultaneous onsets while in the other categories their temporal order was perceptually resolvable since, as in the perception of stop consonants, within-category discrimination of the stimuli was found to be much poorer than between-category discrimination in an ABX discrimination task. Pisoni hypothesized that the underlying basis for the discrimination of consonants in terms of VOT is the listener's threshold for the temporal resolution of two events.

Such an account is seductive because of its simplicity; also, it may explain how animals such as chinchillas (Kuhl and Miller, 1978) and rhesus monkeys (Waters and Wilson, 1976) can be trained to distinguish voiced from voiceless consonants and have been found to have very similar category boundaries to those of humans. On the basis of this hypothesis, one can predict that languages in which VOT functions as a perceptual cue will locate their stop categories on the VOT continuum in such a way as to ensure that perceptual discontinuities are taken advantage of. Presumably no language will contrast two stops which both have VOTs between -20 and +20 msec. Instead, a language with two stop categories should contrast stops from any two of the three possible categories defined by the +20 and -20 msec. boundaries. In fact, such a prediction is fairly consistent with the available data. An examination of Lisker and Abramson's (1971) chart summarizing mean VOTs in several languages (p. 172) reveals that the three VOT categories observed to occur cross-linguistically - voicing lead, short lag, and long lag - correspond more or less to the three categories of tone stimuli.

However, this explanation fails to explain differences between languages in mean VOT values within the three "universal" categories. For example, the mean VOT value of English long-lag stops is considerably shorter than that of Cantonese or Hindi long-lag stops (Lisker and Abramson, 1971). It appears, then, that the exact positions of category boundaries within languages must be to some extent learned. Another limitation of Pisoni's account is the fact that the positions of subjects' perceptual boundaries in experiments involving synthetic continua have been shown to depend on the range of VOTs in the stimulus set (Brady and Darwin, 1978; Keating et al., 1981). If VOT boundaries are determined by basic psychophysical constraints on temporal processing, it is hard to explain why these boundaries should move around at all.

Overview of the Present Study

As can be seen from the above discussion, the issues relating to the perception and production of stop cognates are very complex. Nonetheless, there are several important questions which can be addressed in a study such as the present one. In general, the purpose of this study is to compare and contrast French and English stop consonants. In Chapters 3 and 4, two experiments are described in which attention was focussed on the production of word-initial labial stops in several different vowel contexts. The experiments described in these chapters were aimed at clarifying some of the questions discussed above concerning the production of French stops. In general, an attempt will be made to identify the temporal and amplitude characteristics which distinguish French initial labial stops from English ones. Should French /p/ be regarded as aspirated, and under what conditions does aspiration occur? Do French /b/s exhibit voice breaks? Are there significant differences between languages in voice bar and aspiration amplitudes? Special consideration will also be given here to the similarities and differences between the French /p/ and English /b/ categories. Are these two categories virtually the same? Contrastive studies of this type are particularly important in improving our understanding of how languages may differ in the ways in which they distinguish superficially similar consonant categories.

Chapters 5 and 6 describe two perceptual experiments in which English listeners were tested in order to determine whether they could actually perceive differences between French /p/ and English /b/. The findings have some relevance to our understanding of the perception of the speech sounds of a foreign language. Can listeners attend to small differences in timing which are not distinctive in their native language? As will be seen in Chapter 5, the nature of the two categories also offers a good opportunity to use naturally produced stimuli to test Darwin and Pearson's (1982) findings about the effects of intensity on the perceived point of voice onset in stops.

Chapter 3: An Exploratory Study of Initial Stop Production in French and English (Experiment 1)

Introduction

In this experiment some properties of English and French initial bilabial stop consonants were analyzed and compared. Attention was focussed on the durations of several segments of these consonants: voice bars, silent periods, closure intervals, release bursts, and aspiration noise intervals.

Experimental Method

Speakers

Speakers in this experiment were ten native speakers of European French and seven native speakers of Canadian English. Of the French speakers, five were female and five were male. All were between 25 and 65 years of age. The dialect regions represented by these speakers included Switzerland, Corsica, Lyon, and Northern France. Four of the English speakers were female, and three were male. All were between 20 and 35 years of age. None of the participants had any known speech abnormalities.

Speech materials

The French speakers read aloud the sentence frame "Répétez _____ une fois," filling in the blank with each of 16 nonsense words. The words were monosyllables representing all possible CVC combinations of the French consonants /p/ and /b/ and the vowels /a/, /ɛ/, /i/, and /u/ (e.g., /pip/, /pib/, /bup/, /bub/, etc.). The English speakers read the frame "Say _____ again, please," filling in the blank with similar words consisting of all possible CVC combinations of /p/, /b/, /æ/, /ɛ/, /i/, and /u/. Both frames were designed so that the nonsense word was preceded by a word ending in a vowel and followed by a word beginning with a vowel. Only the initial consonant in each word was analyzed.

Equipment

A set of recordings of the French speakers had been made by Dr. B. Rochet of the Department of Romance Languages at the University of Alberta prior to the present study. The speech samples had been recorded in a quiet room on a Sony TC 182 monaural cassette recorder with accompanying microphone. Recordings of the English speakers were made in a sound-treated room at 15 ips. on the left channel of a Teac A-7030 stereo reel-to-reel tape recorder. A Sennheiser MD 421N cardioid directional microphone set to "speech" was used. It is recognized that differences in equipment and recording conditions may have effects on the outcome of a comparative study such as this one. However, since it was very difficult to locate and obtain speech samples from a large number of native speakers of French, it was decided that the recordings would be acceptable for an exploratory study of stops in the two languages. The potential difficulties in drawing conclusions about the differences between English and French are dealt with in the confirmatory study reported in Chapter 4.

The signals were bandpass filtered with a Rockland Programmable Dual Hi/Lo Filter, Series 1520. During digitization, a Sony TC K61 Stereo cassette deck and a Teac A-7030 tape recorder were used. These were connected to a Digital

PDP-12A computer through analog-to-digital converters. The software used was written in OS/8 and the Alligator system (Stevenson and Stephens, 1979) which is available at the Department of Linguistics at the University of Alberta.

Procedure

The speakers were asked to read the sentence frame sixteen times at a normal conversational volume and rate, filling in the blank with a different nonsense word each time. Three replications were made by each speaker. From these, a set exhibiting no instances of speaker hesitation and having little variation in volume was chosen for measurement. The signals were band-pass filtered (68-6800 Hz) in preparation for digitizing in order to eliminate line-frequency hum and any high-frequency speech components which might result in aliasing. The signals were then digitized at 16 kHz and edited with a visual editing program available on the Alligator system. This program allowed the experimenter to display the waveforms of the signals on the CRT screen of the PDP-12, edit them, and save the desired parts by positioning a pair of cursors (using potentiometers) on the signals. The portion of the signals between the cursors could be played through headphones so that the experimenter could place the cursors exactly where desired. In each case the CVC portion of the signal was saved along with the last few milliseconds of the vowel preceding it and the first few milliseconds of the vowel following it. The saved portions were stored digitally on hard disks and then transferred to digital tape for later use.

For the measurements, an interactive program written in Fortran IV was used.[1] The program displayed one signal at a time at the top of the CRT screen, and the experimenter made measurements by placing a set of cursors on the signal at several points: the beginning of the voice bar and/or silent period associated with the closure interval of the stop, the beginning of the release burst, the beginning of the aspiration phase (if any), the beginning of the periodic (vowel) phase, and the end of the periodic phase. The program automatically calculated (to the nearest 0.1 msec.) the durations of each of the segments determined by the positions of the cursors. RMS amplitudes of the segments were also calculated to the nearest 0.1 unit of a sampling range of 1024. For each initial stop, up to five segments were measured, depending on which were present: voice bar, silent period, release burst (including frication noise, if any), aspiration noise, and vowel.

There were a number of ways to determine where exactly to place the cursors on the signal. For precise adjustments a small portion of a signal could be "magnified" and placed in a window at the bottom of the screen. Also, the experimenter could listen to any portion of the signal through headphones by positioning the cursors on either side of the segment to be listened to and pressing a "play" switch. Finally, the program allowed a spectral section of any portion of the signal to be viewed on a Tectronics Type 611 display.

Segmentation Criteria

For the sake of consistency in making the measurements, several criteria were adopted for the positioning of the cursors. In the majority of cases, the

[1]The program had been developed by Dr. T. Nearey of the Department of Linguistics, University of Alberta.

decisions were quite straightforward. The beginning of a voice bar was defined as the point at which a sudden change in the waveform was observed following the final /e/ of "répétez" or "say" (cf. Keating et al., 1983). All voice bars had a distinctive low-frequency, relatively low-amplitude sinusoidal pattern which could be easily distinguished from the complex pattern of the preceding vowel. The beginning of a silent period was defined as the point at which a signal reached almost zero. Peaks in the signal which were clearly tape noise or ambient noise were ignored. A release burst onset was judged to occur at the first point at which transient energy was observed in a signal. In some instances (mainly in French /b/s) the burst was superimposed on a periodic waveform apparently because voicing continued through the burst phase. Also, in some French /p/s and English /b/s it was difficult to determine whether or not there was an aspiration phase following the burst and, if so, where the aspiration began. Very short noisy segments (usually of low amplitude) which did not sound like aspiration were judged to be frication noise and were treated as part of the burst. Those segments which were judged to be aspiration noise often sounded like brief whispered vowels and were usually of higher amplitude. The aspiration phase was judged to begin at the point at which the amplitude of the signal following the transient burst (and any frication) reached a minimum and then either levelled off or began to increase. Aspiration was judged to end at the first sign of periodicity following the burst. Such periodicity was usually signalled by the presence of a periodic wave pattern of low amplitude which, in some cases, had aspiration noise lasting a few milliseconds superimposed on it (breathy voice). This portion of the aspiration was ignored since the measurement program did not allow the cursors to be placed in such a way that the measured segments overlapped. Finally, the end of the vowel was defined as the beginning of the voice bar or silent period following it.

Since there were two tokens of each initial CV in each language, a total of 160 French and 112 English stops were measured. All measurements were stored on hard disks and then transferred to the University of Alberta's Amdahl 5860 computer system for analysis.

Results

Table 3-1 gives the mean durations and Table 3-2 the mean relative amplitudes of the measured segments (voice bar, silent period, burst, aspiration noise) of each consonant. Also included are closure durations which were obtained by adding the voice bar and silent period durations for each segment. Because of possible variations from speaker to speaker in speaking effort and distance from the microphone, the amplitude measurements of the aperiodic portions and voice bars of each stop were normalized — they were divided by the RMS amplitudes of the vowels following them. Therefore, "relative amplitude" in this study always refers to the RMS amplitude of a segment divided by the RMS amplitude of the vowel following it. Although the relative amplitudes of the segments are presented here, they were not analyzed statistically for reasons discussed below. Instead Experiment 1 focusses on the segment durations only. A statistical analysis of amplitudes was performed in Experiment 2 (See Chapter 4).

The mean durations and amplitudes for all segments are also represented in the schematic diagrams in Fig. 3-1 and 3-2. Each consonant is drawn to scale with respect to both duration and amplitude to give a representation of the "average case" for each type of consonant.

French

The 80 French /p/s fell into two categories: 44 of them showed a measurable amount of aspiration noise ranging from 5.8 to 62.8 msec. (mean 26.5 msec.), and 36 tokens showed no measurable aspiration phase. The relatively large mean duration of aspiration noise (when present) is somewhat surprising, given that French /p/ has often been described as "unaspirated." In fact, some of the French /p/s had longer aspiration phases than some of the English /p/s. Table 3-3 shows the mean aspiration duration (when aspiration was present) associated with each of the four French vowels. It is interesting to note that aspiration occurred three times as often before the two high vowels (33 of 40 cases) as it did before the non-high ones (11 of 40 cases). Moreover, the mean duration of aspiration (when present) preceding the high vowels was more than 2.5 times as long as it was before the other vowels. Also, there were differences in the degree to which individual speakers aspirated. Individual patterns are shown in Table 3-4. While all speakers aspirated at least some tokens, some aspirated as few as two /p/s (of a total of eight), and one subject aspirated seven. The range of the subjects' mean durations was from 11.6 to 39.0 msec.

Voice onset times for the /p/s were determined by adding the burst duration of each stop to its aspiration duration (if any). Mean VOTs associated with each vowel are given in Table 3-5. The mean VOT for /pa/ is 14.1 msec., which appears very similar to Caramazza and Yeni-Komshian's finding of 12-13 msec. (See Chapter 2) for absolute initial /pa/. For /pi/, /pε/, and /pu/ the means were 32.6, 14.5, and 38.6 msec. respectively.

The French /b/s all followed one pattern. In no case was a silent period observed during the stop closure. Instead, the closure interval was always marked by a strong voice bar beginning at the moment of closure (actually a continuation of the voicing of the final /e/ in ***répétez***) and lasting up to the point of release. The release burst was a phase averaging 7.6 msec. between the voice bar and the vowel. No aspiration noise was observed in any of the French /b/s. Since no silent periods occurred, all /b/s were assigned negative voice onset times equivalent to their closure durations.[1] The mean VOTs are given in Table 3-1 and again in Fig. 3-3 in a histogram comparable to the ones used by Lisker and Abramson (1964) and Caramazza et al. (1973). There is no overlap in VOT between the two French categories.

English

Unlike their French counterparts, all English /p/s showed large amounts of aspiration as expected. (Range: 19.2 to 103.4 msec., mean 56.2 msec.) The voice onset times for the English /p/s were determined as they were for the French /p/s. The overall mean VOT was 63.1 msec.

In 54 of the 56 English /b/s, measurable voice bars were observed. These apparently resulted from a continuation into the closure interval of the voicing of the vowel in *say*. However, in 42 cases there was also a noticeable silent period or voice break between this voice bar and the release burst of the initial /b/. In most instances the voice bar appeared to decrease in amplitude gradually to nearly zero at some point before the burst, a pattern quite different from that observed in

[1]It should be noted that this procedure differs from that of Lisker and Abramson (1967) who chose not to assign any VOT values to word-initial voiced consonants in running speech in which voicing continued from a vowel into a closure interval.

French, in which the voice bar amplitude remained more or less constant up to the point of release. Twelve of the fourteen English cases without a voice break can be attributed to two speakers. In other words, these two speakers frequently produced "French-like" voice bars. To summarize, there were three types of English /b/s in these data: those with no voicing whatsoever during the closure interval (2 cases), those with voice bars followed by a voice break (40 cases), and those with voice bars lasting throughout the entire closure interval (14 cases). See Tables 3-1 and 3-2 for the mean durations and amplitudes of the English segments and Fig. 3-2 for the schematic representations of these values.

Assigning VOT values to the English /b/s posed some difficulties. Of course the two cases in which no voicing occurred during closure were assigned positive VOTs equal to the sum of burst and aspiration durations. Those which resembled the French /b/s, in which voicing continued throughout the entire closure duration, were dealt with in the same manner as the French /b/s: they were assigned negative VOTs equal to their closure durations. For these /b/s an overall mean VOT of -111.5 msec. was calculated. For those with both a period of voicing and a voice break during the closure interval, it was not clear how to proceed. Lisker and Abramson (1967), who examined word initial stops in context, mention stops in which voicing continues throughout the entire closure (p. 9), but they apparently do not discuss those with interrupted voicing. One might infer from their discussion that such stops were assigned positive VOTs. Also, other researchers (e. g. Keating et al., 1983) appear to have ignored these voiced intervals during closures in medial stops and have measured VOT from the point of release to the onset of the following vowel. Two problems come to mind here. First, it is conceivable that voice bars of short duration and relatively low amplitude which are followed by long periods of silence (and some such cases were observed here) may not actually function as perceptual cues to the voicing distinction. In fact, they may not even be noticed at all by a listener. From the point of view of a listener, then, it would be misleading to evaluate tokens exhibiting such properties as having long negative VOTs because the voicing preceding the burst may be irrelevant as a perceptual cue. This problem is similar to the problem of "edge vibrations" discussed by Lisker and Abramson (1964, 1967), but these short voice bars differ from edge vibrations. (See Chapter 2.) In fact, the voice bars are always audible to some degree if they are isolated from the rest of the signal. Second (and conversely), voice bars of relatively long duration may well play an important role as perceptual cues even though in tokens of this type, voicing is interrupted. In such cases it would be misleading to assign positive VOT values to the stops.

Nonetheless, some criterion for assigning a VOT value had to be adopted here. The negative VOT values given for the English /b/s in Table 3-1 and represented in the histogram in Fig. 3-4 represent only those cases in which uninterrupted voicing occurred during the closure interval. In cases in which there was a voice break between the voice bar and the release of a stop, the VOT value assigned was positive. For stops of this type, the overall mean VOT was 8.5 msec. In order to present a more accurate picture of the nature of the English voice bars, however, additional histograms showing the distribution of voice bar durations (Fig. 3-5) and of the percentages of closure durations during which voicing was observed (Fig. 3-6) are also given. The latter appear to be distributed with two peaks: one in the 60-70% and the other in the 90-100% category. It is interesting to note the gap between the 0-10% and 20-30% categories, apparently indicating some lower limit on voice bar durations in this study. A similar finding has been reported by Lisker and Abramson (1964, 1967) and by Flege (1982).

The correlation (Pearson *r*) between the voice bar durations and their corresponding burst + noise durations was also calculated. If voice bar duration and burst + noise durations are both determined directly by the difference in glottal timing believed to distinguish stops in English, one might expect to find a significant correlation here; however, no significant relationship was found ($r = 0.09$). This finding may indicate that speakers have independent control over the two stop parameters, but the evidence here is not sufficient for a firm conclusion to be drawn.

Statistical analyses of durations

In order to explore further the differences in segment durations within and between languages, several analyses of variance with repeated measures were performed using BMDP statistical software. The measures of interest were the durations of closure intervals, bursts, voice bars and aspiration noise. For the first two of these, three-way analyses were performed with Language (L) as a grouping factor and Vowel (V) and Consonant (C) as within subject factors. There were two groups (English and French), four levels of V (English /æ/ and French /a/ were treated as a single vowel category and will be referred to here as /a/.) and two levels of C. For the other two measures, two-way analyses were performed. Because no voice bars occurred with /p/s, C was not a factor in the voice bar analysis. Also, since no aspiration was ever observed with French /b/, C was not a factor in the analysis of aspiration noise. Because there were two tokens of each CV combination from each speaker, the mean of the two replications was used in the analyses. A summary of the results of the analyses is given below.

1. Burst durations

This variable showed the most straightforward results. There was a significant effect of L ($F(1,15) = 10.58$, $p < .01$) which was due to longer bursts in French (mean = 8.98 msec.) than in English (6.40 msec.). There was also a significant effect of C ($F(1,15) = 6.81$, $p < .05$) resulting from longer /p/ bursts (mean = 8.93 msec.) than /b/ bursts (6.90 msec.). There was no significant effect of V, and there were no significant interaction effects.

2. Voice bars

Here, a significant effect of L was observed ($F(1,15) = 23.95$, $p < .01$). In general, the French voice bars were much longer than the English ones. The mean voice bar duration for French was 107.78 msec.; for English it was 76.90 msec. There was also a significant effect of V ($F(3,45) = 6.76$, $p < .01$). A Tukey (a) test (Winer, 1971) on the means for the vowels revealed that this effect was due to the voice bars associated with /u/ being longer than those associated with /a/ ($Q(4,45) = 5.63$, $p < .01$), which confirms Smith's (1978) findings for English. He generally found longer voice bars with /i/ and /u/ than with /æ/ and /a/, but no other significant differences were observed in the present study. There did however, appear to be a general trend toward $Vbar_u > Vbar_\varepsilon > Vbar_i > Vbar_a$. No significant VL interaction was observed.

3. Aspiration noise

For this variable there was a significant effect of L ($F(1,15) = 32.39$, $p < .01$). The mean duration of aspiration for *all* /p/ tokens was almost four times as long in

English (56.16 msec.) as it was in French (14.57 msec.). There was also a significant effect of V ($F(3,45) = 7.23$, $p<.01$). Because of the significant VL interaction ($F(3,45) = 7.73$, $p<.01$), a Tukey test on the mean duration of aspiration for each vowel in each language was performed. It revealed that in French, $Asp_u > Asp_a$ ($Q(8,45) = 6.37$, $p<.01$), $Asp_u > Asp_\varepsilon$ ($Q(8,45 = 5.66$, $p<.01$), $Asp_i > Asp_a$ ($Q(8,45) = 5.37$, $p<.01$), and $Asp_i > Asp_\varepsilon$ ($Q(8,45) = 4.66$, $p<.05$). In summary, the high vowels were accompanied by more aspiration than the non-high ones. In English, however, there were no significant differences among vowels. This interaction effect is plotted in Fig. 3-7. It should be noted here that the low mean durations for French /a/ and /ɛ/ are partly due to the occurrence of a larger number of unaspirated /p/ tokens before these vowels than before /i/ and /u/ (i.e. several instances in which aspiration duration = 0 msec.). However, even when only the aspirated cases were used in the analysis, there were significant effects of L, V, and VL ($F(1,8) = 8.97$, $p<.02$; $F(3,24) = 6.52$, $p<.01$; $F(3,24) = 7.00$, $p<.01$ respectively).

4. Closure durations

For the closure durations, there were significant effects of C ($F(1,15) = 10.72$, $p<.01$) and V ($F(3,45) = 4.76$), $p<.01$) but not of L, and significant VL ($F(3,45) = 5.48$, $p<.01$), VC ($F(3,45) = 4.13$, $p<.05$), and VCL ($F(3,45) = 3.36$, $p<.05$) interactions. The VCL interaction is plotted in Fig. 3-8. In general, it appears that the effects of V and C are stronger in English than in French, but there are no obvious patterns in the data in terms of vowel height or advancement. A Tukey procedure performed on the means of each vowel separately revealed that closure durations associated with /b/ were longer than those associated with /p/ in the environment of /u/ in both French and English ($Q(16,45) = 6.73$, $p<.01$ and $Q(16,45) = 6.27$, $p\leq.01$, respectively) and before /ɛ/ in English ($Q(16,45) = 5.46$,

$p<.05$). This finding contrasts with an earlier finding of longer closures for /p/ in word medial position in English (Lisker, 1957).

Chapter 4: Amplitude Measurements (Experiment 2)

Introduction

As mentioned in the previous chapter, one confounding variable in Experiment 1 was the difference in recording conditions for the two languages. In particular, the equipment used for recording the English samples was of higher quality than that used for recording the French ones. While initially it was predicted that the use of two sets of equipment would not have important effects on the outcome of the experiment, some differences in frequency response (particularly in the microphones used) may have caused low frequency components to be amplified and higher frequency components to be attenuated in the French recordings. While such differences are very unlikely to have affected the durations of the segments measured in the experiment (see comparisons below), some of the amplitude measurements for French and English may not be comparable. Therefore, it was decided that the Experiment 1 production study should be partially replicated. Speech samples from three French-speaking subjects were obtained using the same equipment that was used for the English-speaking subjects in Experiment 1. The purpose of this second experiment was to determine whether the different conditions affected the outcome of the original experiment.

Experimental Method

Subjects

The participants in this experiment were three native speakers of European French ranging in age from 20 years to 45 years, one from Bordeaux (S1), one from

Lyon (S2), and one from rural Brittany (S3). S2 had participated in the first experiment. Two of the speakers were female, and one was male. None had any known speech or hearing abnormalities.

Equipment and Procedure

The equipment and procedure in this study were identical to those used for the English speakers in Experiment 1, except, of course, that the subjects read the French sentence frame, "Répétez _____ une fois." Recordings were made of the 16 CVC combinations of /b/, /p/, /a/, /i/, /ɛ/, and /u/. These were digitized and measured using the same equipment and procedure as in Experiment 1. The new set of French recordings is therefore comparable with the original set of English recordings.

Results

General findings

In general, the tokens analyzed in this experiment resembled the French stops in the first experiment quite closely. For example, as can be seen from Table 4-1, aspiration in initial /p/ occurred more frequently (100% of the time) and for apparently longer durations with the high vowels than with the others. This time, however, a greater proportion of the tokens (17 of 24) showed significant aspiration. All of the /i/ and /u/ tokens were preceded by aspiration which lasted, on the average, about 3 to 4 times as long as the aspiration which occurred in the 5 aspirated (as compared with 7 unaspirated) /a/ and /ɛ/ tokens. As well, significant voice bars were always found with /b/s in the new sample. Nevertheless, a few differences emerged in the case of one subject (S3), who seemed to produce more heavily aspirated stops than the other two subjects here and than the French-speaking subjects in Experiment 1. For this reason, mean aspiration amplitudes for each subject are given in Table 4-2. The mean relative aspiration noise amplitude of the 11 aspirated tokens produced by S1 and S2 was 0.21 (cf. overall mean of 0.13 in Experiment 1). For S3 it was 0.59. The large mean amplitude for S3 is mainly due to two /pi/ tokens in which very strong aspiration accompanied vowels of unusually low amplitude. Also, significant aspiration (on the order of 12-14 msec.) was noted even in the two /bi/ tokens produced by this subject. Finally, five of the eight /b/ tokens she produced had voice bar patterns similar to those found in English: that is, she sometimes produced voice bars which "died out" several msec. before the burst. She was the only French subject in either experiment to exhibit such a pattern, and it is possible that the differences observed in her stops were due to dialect. The subject was originally from rural Brittany, a dialect region not represented in the original sample.

The mean amplitudes for all segments are given in Table 4-3, and voice onset times for the new sample are shown in a histogram in Fig. 4-1. In contrast with the results of Experiment 1, there is a slight overlap in VOT here because the five /b/ tokens produced with voice breaks by S3 fell into the short lag category. However, all of these /b/s were preceded by noticeable voice bars ranging from 68 msec. to 124 msec. in duration. Therefore, it is still the case that the initial French stops in this study are always distinguishable by the presence or absence of voicing during the closure interval.

One of the participants in Experiment 2 had also participated in Experiment 1. Her results in both experiments are given in Tables 4–4 (a) and 4-4 (b).

Statistical Analyses

In order to determine whether the recording conditions of Experiment 1 might have affected the measures made, several analyses of variance were performed, the purpose of which was to compare the measurements in the two French samples. The approach used was very similar to that followed in Experiment 1. A partially repeated measures design was used, this time with Experimental Conditions (E) as the grouping factor, and with V and C as within subject factors. There were two groups (Experiment 1 and Experiment 2), two levels of C (/p/ and /b/, and four levels of V (/a/, /i/, /ɛ/, and /u/). Two sets of analyses were run, one for the duration measures and one for the amplitude measures.

As expected, there was no significant effect of E on burst durations, aspiration noise durations, or voice bar durations. Nor were there significant CE, VE, or VCE interactions. However, there was an effect of E on closure durations which was significant at the .05 level but not at the .01 level ($F(1,11) = 6.47$). This was due to longer closure durations in the new set of recordings. Nevertheless, it is highly unlikely that the recording conditions alone could have affected the closure duration measurements but not the other duration measurements. It may simply be that the three subjects from the new, smaller sample collectively spoke more slowly than those in the first sample and that the only noteworthy effects of this difference in speaking rate were on closure durations. In fact, there was a weak, non-significant trend toward longer voice bars and longer aspiration in the second experiment as well.

Analyses of variance were also carried out on the amplitude measurements. It was found that there was a significant effect of E on burst and aspiration amplitudes ($F(1,11) = 5.04$, $p<.05$ and $F(1,11) = 9.67$, $p<.01$ respectively) and on voice bar amplitudes ($F(1,11) = 8.02$, $p<.05$). The mean relative amplitudes for each segment for the new French-speaking subjects are given in Table 4-5 along with the corresponding mean amplitudes from Experiment 1. In general, these figures indicate lower voice bar amplitudes, but higher burst and aspiration amplitudes in the second experiment. However, when S3 was omitted from the second sample, the difference in aspiration amplitudes was no longer significant, even at the .05 level.

Because there was no evidence of strong differences in burst, aspiration, or voice bar durations between the two groups of French-speaking subjects, it was concluded that the results of the analyses on these durations in Experiment 1 should be accepted without further investigation. However, since there were differences in the amplitude measurements in Experiments 1 and 2, it was decided that the amplitude measurements of the French subjects in Experiment 2 should be compared with those of the English subjects from Experiment 1. Several repeated measures analyses of variance were carried out. The design used was the same as that used for the duration measurements in the previous experiment. It should be pointed out that because of the small sample size, the results reported below must be interpreted with some caution.

1. Voice bars

The analyses revealed no significant effect of L on voice bar amplitudes. There was, however, a significant effect of V ($F(3,24) = 35.44$, $p<.01$). Also, the VL interaction was found to be significant ($F(3,24) = 3.43$, $p<.05$).[1] When the VL interaction was explored further using a Tukey (a) procedure, a tendency for the differences in voice bar amplitudes among vowels to be greater in French than in English was observed. It was discovered that in French $Vbar_i > Vbar_a$ ($Q(8,24) = 10.50$, $p<.01$), $Vbar_i > Vbar_\varepsilon$ ($Q(8,24) = 10.42$, $p<.01$), $Vbar_u > Vbar_a$ ($Q(8,24) = 8.64$, $p<.01$, and $Vbar_u > Vbar_\varepsilon$ ($Q(8,24) = 8.57$, $p<.01$). While the voice bars associated with both high vowels in French were stronger than those associated with the non-high vowels, English /bu/ and /bi/ had stronger voice bars than /ba/ ($Q(8,24) = 6.00$, $p<.01$ and $Q(8,24) = 5.50$, $p<.05$, respectively), but only /bu/ had stronger voice bars than /bɛ/ ($Q(8,24) = 5.14$, $p<.05$). There was, however, a trend toward stronger voice bars before both high vowels in English.

2. Bursts

There were no significant effects of L or C on burst amplitudes, nor were there significant CL or VCL interactions. There was, however, a significant effect of V ($F(3,24) = 11.24$, $p<.01$), and the VL interaction was significant when calculated by the usual method ($F(3,24) = 3.55$, $p<.05$) but not using the Greenhouse-Geisser criteria. A Tukey analysis showed that /i/ was preceded by stronger bursts than /ɛ/ in French ($Q(8,24) = 6.15$, $p<.01$) but that the difference was not significant in

[1]The VL interaction was not found to be significant when the more conservative Greenhouse-Geisser method (see Myers, 1979) was employed. Unless otherwise indicated, all other effects described here as significant were also significant by the Greenhouse-Geisser method.

English. Nevertheless, a tendency for the high vowels to be preceded by stronger bursts was evident in both languages.

3. Aspiration

Interestingly, there was no effect of L on aspiration amplitude. Initially it was felt that this finding might be somewhat misleading. Because S3 produced heavy aspiration in /pi/ syllables, the overall mean aspiration amplitude for French /p/ was fairly high, as was its standard deviation. There was a trend in the data from the other subjects (and from S3's data for the other vowels) for French consonants to have lower aspiration amplitudes than English consonants, but when an analysis of variance was performed with S3 excluded, the effect of L was still not significant, even at the .05 level.

There was a significant effect of V ($F(3,24)$ = 5.6[illegible] p< .01) and a significant VL interaction ($F(3,24) = 4.47, p<.05$). When this i[illegible]n was explored further with a Tukey (a) procedure, it was found that French /pi/ had significantly stronger aspiration than did French /pa/ ($Q(8,24) = 6.82, p<.01$), /pɛ/ ($Q(8,24) = 6.31, p<.01$), or /pu/ ($Q(8,24) = 5.49, p<.05$). This finding must be viewed with some caution since the differences may have been exaggerated by S3's heavy aspiration. No significant differences were found within English.

A Brief Summary of the Production Studies

From the results of these two studies, the differences between the stop categories in the two languages may be briefly characterized as follows. While initial English /p/ is always significantly aspirated, the presence and degree of aspiration in French /p/ seems to be conditioned by the height of the following vowel. French /p/ is frequently (but not always) aspirated before /i/ and /u/ although the duration of the aspiration phase is likely to be shorter than it is in English. French /p/ therefore has a shorter VOT than English /p/ but a longer VOT than English /b/. There is no evidence, however, that aspiration in French is of lower amplitude than aspiration in English. Initial French /b/ shows longer voice bars than English /b/, and while in French these almost always continue through the entire closure period, in English they frequently end several msec. before the stop is released. On the average, French voice bars do not appear to be of higher amplitude than English ones. Stop bursts in French appear to be longer than those in English, but do not differ in amplitude. In neither language do the two stop categories show an overlap in terms of voice onset time.

Chapter 5: An Exploratory Cross-Language Perception Study (Experiment 3)

Introduction

One interesting question arising from the results of Experiments 1 and 2 concerns the relationship between English /b/ and French /p/. As mentioned earlier, some previous comparative descriptions of French and English stops make no clear distinction between these two categories. One might infer that an English speaker attempting to produce French /p/ need only produce an English /b/ without prevoicing. The results of the production experiments, however, indicate that while there is some overlap between the two, there may also be significant differences, particularly in terms of voice onset time. Even if we ignore differences resulting from the presence of voice bars in some English /b/s, the sums of burst and aspiration durations appear to be smaller on the average for English /b/. It may well be incorrect then to characterize a word-initial French /p/ as simply an English /b/ with no voice bar, and anglophones learning French may have to make some fairly subtle adjustments to produce stops from the French category correctly.

For these reasons, it was decided that a perceptual experiment in which subjects would be asked to categorize a set of natural French /p/ and English /b/ tokens should be performed. The aim of Experiment 3, then, was to determine whether a number of measurable differences between the two categories could be perceived by native speakers of English. Not only did such an experiment afford an opportunity to test listeners' abilities to recognize small differences between categories which are superficially similar, but in this case it was also possible to test Darwin and Pearson's (1982) claims about the role of differences in overall intensity in the perception of stops by English speakers. In particular, a number of exploratory intensity measurements were made of the stimuli. The relationships between these and the subjects' identification scores were then considered.

Experimental Method

Subjects

Listeners in this experiment were eight native speakers of English between 22 and 40 years of age, all of whom had studied some phonetics. Two were male and six were female. All had at least a basic knowledge of French, and seven had completed at least one university-level French course.

Stimuli

The subjects listened to a modified subset of the words recorded in Experiment 1. Recordings from eight speakers, four French-speaking and four English-speaking were used. For each language condition, two speakers were female and two were male. Except with respect to sex and language category, the selection of speakers was arbitrary: no reference was made to the measurement data from Experiment 1. From the French recordings, only the six /pa/, /pɛ/, and /pi/ "words" were selected. The six English /b/ words containing the vowels /æ/, /ɛ/, and /i/ were also used. The words containing /u/ were excluded because of the difference in the quality of this vowel in the two languages. It is conceivable that listeners might have been able to distinguish French syllables from English ones on this basis alone without attending to the properties of the consonants.

The stimuli were windowed on a Digital PDP-12 computer by a Fortran 4 program. All speech components preceding the burst in each signal were removed, and the signal was truncated at 68 msec. The final eight msec. were windowed with a ramp function so that the signal decreased in amplitude to 0 at 68 msec. As a result, each of the 48 stimuli consisted of an initial French /p/ or English /b/ burst followed by the initial portion of the following vowel. No part of the final consonant was included, and the stimuli were deliberately made short so that the vowel quality would be unclear and thus not give any indication of the language of the speaker.

Following the windowing procedure, each stimulus was inspected both visually and aurally with the Alligator system. It was found that the vowel quality in most cases was highly ambiguous; there was not enough information in the vowel portions of the signals to indicate whether they were French or English. In particular, in the English stimuli there was no evidence of the diphthongization whi[illegible]aracteristic of English vowels but not of French ones. It is unlikely, t[illegible]vowel quality could have been used by listeners as a means of identifying [illegible]mulus items. In the case of the French /i/ syllables, however, it was noted that many of the consonants were easily identifiable as /p/s because they were strongly aspirated.

Each stimulus set consisted of four replications of each of the 48 tokens in random order for a total of 192 items. A different randomization was used for each replication of the experiment.

Equipment

A Digital PDP-12 computer and the Alligator system were used to play the stimuli to the listeners and to record their responses. The stimuli were played through digital-to-analog converters, band-pass filtered (68 - 6800 Hz) with a Rockland Programmable Dual Hi/Lo Filter (Series 1520) and amplified through the left channel of a Sony Integrated Stereo Amplifier (TAF35) set to a comfortable listening level. The subjects listened on Telephonics TDH-49 headphones and registered their responses on a switchbox with two switches labelled *French p* and *English b*. Responses were recorded automatically by an Alligator program and stored in Alligator files for later sorting and analysis.

Procedure

The stimuli were presented to the subjects' right ears through the headphones in a sound-treated room. They were advised that they would hear a stimulus set consisting of French /p/s and English /b/s and that they were to identify which category each item belonged to by pressing the appropriate button on the switchbox. They were not to assume that the two sounds were equally represented in the set. Each stimulus was continually replayed with an interstimulus interval of 1 second until one of the buttons on the switchbox was pressed. The subjects were told that they could listen to a stimulus as many times as necessary but that they should make their choices as quickly as possible rather than listen to a stimulus many times over. They were also instructed to ignore the vowel quality of the stimulus items and to make their judgments strictly on the basis of the consonants they heard. Finally, they were told that if they heard a stimulus which did not seem to fit either of the two categories, they were to choose the category which seemed closest to the stimulus item in question.

Each subject was asked to listen to a total of five different randomizations of the stimulus set although one subject dropped out after listening to only four. Therefore, a total of 39 listening sessions were held, and 7488 responses were collected for analysis. No more than two sessions were held on a single day for any listener, and a short break of 15 to 20 minutes between sessions was given. Individual sessions lasted from 15 to 25 minutes, depending on how quickly the listeners entered their responses. Following each replication, a short debriefing session was held during which subjects were asked for their impressions of the experiment. In general, the subjects reported that their judgments were not affected by vowel quality because they could not identify the language of the speakers on the basis of the vowels. They also reported that the heavily aspirated stimuli were particularly easy to categorize.

Results

A preliminary examination of the results indicated that the subjects were generally successful at the task. The highest identification rates were obtained for the /pi/ syllables (seven of the eight tokens were correctly identified over 99% of the time), which, as mentioned earlier, sounded heavily aspirated in comparison with the other sounds. Because these stimuli were much more aspirated than the others, it was felt that the subjects may have used different criteria in making their judgments on them than on the other items. That is, they may have labelled virtually all of these as /p/s simply because the aspiration was clearly present. Inclusion of the /pi/ correct identification (CID) scores in the analysis might have led to some distortion of the results because some other stimuli with far less aspiration and with shorter VOT values were also well identified. For these reasons, the identification scores on the 16 /pi/ and /bi/ syllables were not used in the subsequent analyses. The results reported are therefore based on the subjects' categorizations of the remaining 32 stimulus items.

A *t*-test of $\mu = 50\%$ revealed that the subjects were able to identify the remaining stimuli correctly at levels well above chance ($t(7) = 6.02$, $p<.001$).The CID scores for each subject on the non-/i/ stimuli are given in Table 5-1. The highest score was 80.9%; the lowest was 54.1%. The overall rate was 69.1%. The six highest-scoring subjects, all of whom had CID scores above 65%, showed a slight response bias toward French /p/: of the 4992 responses collected, 54% were /p/ categorizations. The total CID score for French /p/ was 73.3%; for English /b/ it was 65.0%. It appears, then, that most of the listeners were able to use one or more cues to distinguish at least some members of one category from those of the other category.

Analysis of the stimuli

From the preliminary examination of the data, it appears that in terms of perception, French /p/ and English /b/ are only partially overlapping categories which can be distinguished at least part of the time. The next step in this study was to determine what characteristics of the stimuli the listeners may have attended to when making their judgments. The measurements made in Experiment 1, along with some additional ones were considered in an attempt to identify possible perceptual cues.

1. VOT

VOT values were determined for each of the stimulus items on the basis of the results of Experiment 1. For the French /p/s and for the English /b/s without voice bars, the VOT values were the same ones calculated previously. However, the VOT values for five of the /b/ tokens had been changed by the windowing procedure because their voice bars had been removed. The VOT values for these items were determined by adding their burst durations to their aspiration durations, except in the case of two stimuli which clearly showed high-amplitude periodic energy right at the moment of release. Both of these tokens had had strong voice bars prior to windowing, and voicing apparently continued right through the burst phase. They were both assigned a VOT value of 0 msec.

A description of the stimuli in terms of VOT is given in Table 5-2. The overall range of VOTs was from 0 to 25.5 msec. (0 - 17.5 msec. for English /b/ and 7.1 - 25.5 msec. for French /p/). The mean VOT for French /p/ was 12.0 msec.; for English /b/ it was 6.9 msec. An examination of the VOT values of the individual stimuli revealed a trend for tokens with long VOTs to be labelled as /p/s and for those with short VOTs to be labelled as /b/s: of the 7 "best identified" /b/ stimuli (CID over 75%), 6 had VOTs of less than 10 msec. The worst /b/ (CID 15%) had a VOT of 17.5 msec. The 5 best French /p/s (CID over 95%) had VOTs ranging from 13.8 to 25.5 msec. Of the four worst /p/s (CID 44-47%), three had VOTs of less than 8.5 msec.

In order to explore further the relationship between VOT and the subjects' categorizations, some additional analysis was carried out. The total number of times each item was identified as a /p/ (hereafter labelled the /p/-id score) by all subjects was transformed using an arc sine transformation (see Ferguson, 1981). A Pearson *r* of 0.767 ($p<.01$), indicating a fairly high correlation, was calculated between the transformed scores and the VOT values. It is possible, then, that the subjects relied at least partly on VOT (or on some property or properties of the stimuli highly correlated with VOT) as a means of categorizing the stimuli. Nevertheless, two important observations should be made here. First, VOT could not have been the only factor influencing the subjects' judgments. For instance, one of the most poorly identified /p/s had a fairly long VOT of 13.0 msec. and one of the worst /b/s had a short VOT of only 7.5 msec. Furthermore, some pairs of stimuli with very similar or identical VOTs were found to have markedly different identification scores associated with them (see Table 5-4). It was decided that the roles of other potential predictors of the subjects' scores should be assessed in an attempt to account for these discrepancies. Second, the range of VOT values in the stimulus set was much smaller than the range usually identified for either English or French bilabial stops. Lisker and Abramson (1964), for instance, report a total VOT range of from -130 to +120 msec. for the English bilabial categories. Caramazza and Yeni-Komshian (1974) give an overall range of from less than -150 msec. to almost +25 msec. for European French. It appears that the subjects in the present study were able to make distinctions finer than those normally expected of subjects in similar perceptual studies involving synthetic stimuli. It is possible that the rather fine distinctions made were facilitated by perceptual cues in addition to or other than VOT.

2. Other Predictors

In order to explore further the perceptual cues which might have influenced the subjects' identification scores, a multiple regression analysis was performed using Minitab statistical software. A number of possible predictors were entered. Among these were some measures of "attack" or "rise time" based very roughly on Darwin and Pearson's (1982) study. They argued that voicing onset was *perceived* to occur not necessarily at the time at which the periodic portion of a synthetic CV syllable began, but at the time at which the overall intensity of the stimulus reached a critical level of 20 dB below the vowel steady-state. One might expect that in natural speech the time at which such a critical intensity level is reached is positively correlated with VOT since the onset of voicing is accompanied by an increase in overall intensity. The critical intensity level should be reached some time after the onset of periodicity, assuming that periodic energy first appears in the signal at a relatively low amplitude and thereafter increases to a maximum level during the steady-state portion of the vowel. However, this correlation is unlikely to be perfect. It is possible that intensities in two stops with identical VOTs may reach the critical level at different times because of different rise slopes. One exhibiting a steep rise portion should be perceived as more voiced because the critical intensity level is reached earlier.

To obtain the necessary information for the present study, the amplitudes of the stimuli at fixed times were measured using a Fortran 4 program on a Digital PDP-12A computer. A 16 msec. rectangular integrating window was moved along each signal beginning at 0 msec., and RMS amplitudes were calculated at 2 msec. intervals. Thus, for each 68 msec. signal, 35 measurements were made. These values were entered into another program which calculated a "mean peak" amplitude by averaging the 15 measurements from 20 to 48 msec. inclusive. This value represents a mean amplitude for the more or less steady-state portion of the vowel. The values discussed above were used to calculate three different attack measures labelled *d*, *t*, and *na*. First a program was used to determine, by linear interpolation, the approximate time (relative to the onset of the stimulus) at which the RMS amplitude of each signal first reached a critical percentage of the mean peak. The critical percentages selected for exploration were 10, 25, 50, 75, and 90% (-20, -12, -6, -2.5 and -.9 dB) and were labelled d10, d25, d50, d75, and d90 respectively. For example, the value of d75 for a particular stimulus item is the approximate time (in msec.) that the RMS amplitude of that item first reached 75% of the mean peak. A small value for d75 indicates that the amplitude of the stimulus item reached the critical level early and therefore exhibited a relatively "hard onset" (see Flege, 1982) or, in equivalent terms, a steep rise envelope. A preliminary examination of the *d* values for the stimulus items revealed differences between the two stop categories. (See Table 5-2.) For example, the range of d75 values for English /b/ was 9.95 to 17.65 msec., with a mean of 12.94 msec. For French /p/ the range was 16.36 to 34.61 msec., and the mean was 24.14 msec. In general, then, the English /b/ amplitudes reached 75% of peak earlier than the French /p/ amplitudes. Furthermore, there appeared to be relatively little overlap between the d75 values for the two categories.

Obviously, in a study such as this one, there is no way to predict *a priori* all of the characteristics of a set of stimuli which may figure in subjects' judgments. The measure of attack discussed above does not conform exactly to any theoretical model; it is merely an educated guess. In fact, many other measures could have been used in the regression analysis, and many other plausible methods of calculation could have been used, but in an exploratory study there are obvious practical constraints on the number of potential cues which can be investigated. One potential difficulty with the *d* measure was that it did not take into consideration the differing onset amplitudes of the stimuli. It is conceivable that the listeners used as a reference point not only the mean amplitude of the vowel steady-state, but the amplitude of the signal at its onset. In synthetic stimuli, onset amplitudes can be controlled; however, in these natural utterances a wide range of onset amplitudes was found. For this reason, a more complex set of measurements was also used in the regression analysis. In this case, the mean peak amplitude was calculated as before, but this time the critical amplitude was defined as the amplitude x% of the range between the amplitude measured at 0 msec. and the mean peak, where x = 10, 25, 50, 75, and 90. Each of the *t* values, then, was the approximate time (calculated by linear interpolation) at which one of these critical amplitudes was reached and was labelled t10, t25, etc. A small *t* measurement has the same significance as a small *d* measurement; that is, it indicates a steep onset amplitude slope. Also, like the *d* values, the *t* values for the English stimuli appeared to be smaller than the French ones although the overlap in the two categories appeared to be greater. For example, the range of t25 values for English /b/ was from 2.56 to 10.52 msec.; the mean was 4.89 msec. For French /p/ the range was from 4.13 to 19.30 msec., and the mean was 11.93 msec.

A third, simpler measure was also used: the first ten amplitude measures for each stimulus item were divided by the mean peak value. The result was a set of 10 "normalized amplitude" measurements for each stimulus item taken at 2 msec. intervals from 0 msec. to 18 msec. They were labelled na0, na2, etc. and may be interpreted as the proportion of maximum amplitude reached at fixed times during the stimulus. A large *na* value at an early point in the signal indicates a steep onset slope. Overall, there was a trend for the English stops to have larger *na* values than the French ones. For na8, for example, the range in English was from 0.0942 to 0.7291 with a mean of 0.5338. The range for the French stimuli was from 0.0577 to 0.5357 with a mean of 0.2418.

Finally, it was noted that the only stimulus item with a /p/-id score of 100% had the longest aspiration noise duration (14.9 msec.) of all the stimulus items according to the measurements made in Experiment 1. Furthermore, the "worst" /b/ (which was identified as a /p/ 85% of the time) had the second longest aspiration noise duration (12 msec.). While these tokens had fairly long VOT values associated with them, not all of the stimulus items with relatively long VOTs were aspirated (many simply had long bursts, accompanied in some cases by low intensity frication noise, which was not treated as aspiration), nor did all of the items with high /p/-id scores. It is possible that aspiration noise made a contribution to the subjects' categorizations independently of VOT. For this reason, the aspiration noise durations and aspiration noise amplitudes of the stimulus items (measured in Experiment 1) were also entered into the analysis. The durations ranged from 0 msec. in 23 of the 32 cases to a maximum of 14.9 msec. It should be noted that entering the aspiration durations into the analysis together with VOT was effectively the same as entering the burst durations because VOT is the sum of the burst and aspiration durations.

Statistical analysis

A correlation matrix for the subjects' transformed /p/-id scores, VOT, aspiration noise duration, and other selected predictors is given in Table 5-3. VOT is the predictor which is the most highly correlated ($r = 0.767$) with the transformed identification scores. The *d*, *t*, and *na* measurements most highly correlated with the transformed scores were d75 ($r = 0.649$), t25 ($r = 0.634$), and na8 ($r = -0.713$) ($p<.01$ in all cases). Furthermore, these three highly-correlated predictors showed moderate to high correlations with VOT of 0.443, 0.675, and -0.761 respectively. The correlations between aspiration noise duration and amplitude and the transformed identification scores were relatively low ($r = 0.320$ and 0.284 respectively).

An examination of the predictors discussed above in relation to the subjects' categorizations revealed some interesting trends. Table 5-4 lists five pairs of stimulus items in which the two items had identical or almost identical VOTs yet very different /p/-id scores. The subjects identified one member of each pair fairly consistently but perceived the other as relatively ambiguous. In every case, the d75 value for the item with the lower /p/-id score was smaller, indicating a steeper rise slope. For instance, stimulus items 23 and 30, both /p/s, had identical VOTs of 8.4 msec. Yet item 30 was perceived as a /p/ 84% of the time while item 23 was much more ambiguous and was identified as a /p/ only 46% of the time. One dimension on which the two items seemed clearly to differ was the d75 measure: the amplitude of the more /p/-like stimulus did not reach 75% of peak until nearly 25 msec. In the case of the other token, this level was reached 7.5 msec. earlier. A similar pattern is seen in the other four cases. It should be noted, however, that although the d75 measure apparently can account for some of the differences in the pairs of stimuli listed in Table 5-4, it is less successful in accounting for other differences. For example, items 32 and 30 have very similar d75 values but quite different /p/-id scores. Sample waveforms of some of the stimulus items from Table 5-4 are given in Fig. 5-2.

The arc sine-transformed p-identification scores were regressed on the three sets of attack measures discussed above, along with the voice onset times and the aspiration noise durations and amplitudes measured in Experiment 1. Since there were five sets of d measurements, five sets of t measurements, and ten sets of na measurements, as well as the aspiration noise durations and amplitudes, the total number of predictors was 22. A stepwise procedure (Myers, 1979) was used to calculate regression equations and their corresponding multiple correlation coefficients (R) for the best combinations (those yielding the highest R values) of predictors with F to enter set at 4.00. The highest R was obtained for an equation which included VOT, aspiration noise duration, and d75 ($F(1,28)$ = 81.27,18.07, and 10.90, respectively; $p<.01$ in all cases). With this equation it was possible to account for almost 78% (R^2) of the variance in the transformed scores, as compared with 57% with the regression equation including VOT alone. The five "next best" equations were also determined and were found to have R^2 values ranging from 61% to 70%. All of these included VOT, and three of them included one set of d measurements (d75, d90 and d50). It appears, then, that in combination with VOT the d measure can be a particularly useful predictor. Figure 5-1 illustrates the differences in all 5 mean d values for the two categories. In general, the categories appear to be farther apart for the d75, d50, and d90 values. In order to determine how well the subjects' scores could be predicted without VOT, an additional analysis was performed with VOT excluded as a predictor. In this case the highest R^2 (55.4%) was obtained for an equation which included na8 and d75, both of which were correlated with VOT anyway. It appears, then, that although some predictors in combination with VOT are better than VOT alone, no combination of the examined predictors which excludes VOT gives a better prediction than VOT alone. The eight equations discussed above are given in Table 5-5 along with their corresponding R^2 values and the standard deviations about the regression lines.

This preliminary investigation of listeners' perceptions of French and English stops indicates that VOT is a fairly powerful predictor of the subjects' categorizations. In addition, some measure of attack, such as a d value may be used in combination with VOT to increase the accuracy of the prediction. Nevertheless, some caution is required in drawing conclusions based on these results. In particular, the use of a large number of predictors in a stepwise multiple regression procedure increases the chances that significant results will be found. In order to confirm the findings, a perceptual study similar to the present one was performed using different stimuli and a new group of listeners.

Chapter 6: Cross-validation of Perceptual Study (Experiment 4)

Introduction

Because Experiment 3 was intended as an exploratory study involving a preliminary assessment of a large number of variables and because the two sets (French and English) of stop tokens used in the experiment had been recorded on different equipment, it was felt that the results of Experiment 3 should be cross-validated. In Experiment 4, two groups of listeners, one naive and one experienced, were asked to categorize a new set of French /p/ and English /b/ stimuli. The new results were compared with those of the previous experiment.

Experimental Method

Subjects

The participants were ten native speakers of English, eight female and two male. All were between the ages of 18 and 35. Five of the speakers were classified as "experienced": all had studied some French at the university level or equivalent, and all had studied some phonetics. The other five subjects were naive listeners who had essentially no knowledge of either French or phonetics. Two of the experienced subjects had participated in the previous experiment.

Stimuli

In this study, the stimulus items were modified tokens taken from the two production studies discussed in Chapters 3 and 4. The stimuli were similar to those used in Experiment 3, except that in this case the French tokens all came from the three speakers who had participated in the Experiment 2 production study. Therefore, all recordings from which tokens were taken had been made on the same tape recorder in the same sound-treated room. Also, a slightly smaller stimulus set was used. All stimuli were modified in exactly the same way as in Experiment 3. The recordings from the three English speakers in Experiment 1 whose recordings had not been used in Experiment 3 were used.

The six French /pa/, /pɛ/, and /pi/ syllables from each subject and the English /ba/, /bɛ/, and /bi/ tokens were modified as in Experiment 3. Although the subjects' identification scores on the /pi/ and /bi/ stimuli were not used in the final analysis (as in Experiment 3), these items were included to ensure that the conditions of Experiment 4 were as much like those of Experiment 3 as possible. In particular, since some researchers (Brady and Darwin, 1978; Keating et al., 1981) have found evidence of range effects on identifications of stimuli from synthetic VOT continua, it was necessary to use roughly the same range of tokens (in terms of VOT) in this study as in the previous one. The stimulus set consisted of 36 items, each one repeated four times, for a total of 144 stimuli.

Procedure

The subjects listened to the stimuli through headphones in a quiet room and registered their judgments on the same switchbox that was used in Experiment 3. The experienced listeners were told that they would hear a stimulus set consisting of French /p/s and English /b/s, and the switches on the box were labelled accordingly; for the naive condition the labels were simply *p* and *b* since information about the languages of the speakers would be of no use to them. As before, each stimulus was played repeatedly until a response was entered, but the subjects were encouraged to make their decisions as quickly as possible. Each participant listened to the stimulus set four times. Eight subjects listened to two replications per day on separate days while the other two listened to all four replications on the same day. A break of about fifteen minutes was given between replications.

Results

General Findings

A preliminary examination of the subjects' responses revealed similar results to those found in the previous experiment. As before, a *t*-test of $\mu = 50\%$ on the CID scores indicated that the subjects generally performed at levels well above chance ($t(9) = 5.19$, $p<.001$). Table 6-1 gives the individual subjects' CID scores in percentages. For the experienced listeners, these ranged from 66.1% to 72.7% with an overall mean of 69.4% - virtually identical to the overall mean identification score obtained for the experienced subjects in Experiment 3. Among the naive listeners, the CID range was from 47.4% to 75.5% with a mean of 61.4%. The lower mean for the naive group was due to poor performance by two listeners (S4 and S5), who were apparently unable to distinguish the two categories: they obtained CID scores very near 50% while the other three naive subjects scored above 65%. With the two lowest-scoring subjects excluded, the mean for the naive group was 69.9%. A *t*-test on the means of the two groups (naive and experienced) revealed no significant difference. The overall mean for all ten subjects was 65.4%. Both the highest and the lowest CID scores were observed in the naive group.

A response bias in favour of French /p/ (which was also observed in Experiment 3) was evident in the categorizations of four of the five experienced listeners in Experiment 4 (on the average they classified 58% of the tokens as /p/) and in those of the two "unsuccessful" naive subjects (70% /p/ identifications). A bias toward English /b/ was seen in the responses of the successful naive listeners (64% /b/ identifications).

Analysis of the Stimuli

The main purpose of the statistical analyses performed here was to determine whether the subjects' categorizations in Experiment 4 were comparable to the categorizations in Experiment 3. Therefore, the amplitudes of the stimuli were measured, and *d* and *t* values were calculated as in Experiment 3. As well, the VOTs and aspiration durations measured in the production studies were included in the analyses.

Statistical Analysis

These values were entered into the regression equations calculated in Experiment 3. (See Table 5-5.) In each case a RMS error value for the equation was calculated using the following:

$$\text{RMS error} = (\Sigma\,(y - y')^2/n)^{1/2},$$

where y = the arc sine-transformed value of the total score for all subjects on a particular stimulus item, y' = the predicted value of y according to the regression equation in question, and n = the total number of stimulus items for which identification scores were predicted (24).

Table 6-2 gives this error value along with the standard deviation about the regression line for each equation determined in Experiment 3. In general, these values are quite comparable, indicating that the equations from Experiment 3 can predict fairly well the subjects' performance in Experiment 4. The three lowest RMS error values were those associated with the regression equations including VOT and d50, VOT and t90, and VOT and d75.

In addition, the correlations between some of the more "successful" predictors found in Experiment 3 and the transformed /p/-identification scores were calculated, and completely independent regression analyses were performed on the data from Experiment 4 using the five "best" combinations of predictors found in Experiment 3 (VOT, aspiration duration and d75; VOT and aspiration duration; VOT and d75; VOT and d90; and VOT and d50) and using VOT alone. These combinations of predictors were forced into the analysis in order to determine which ones accounted for significant proportions of the variance in the transformed scores.

The results of these analyses are given in Table 6-3. As in Experiment 3, VOT (r = 0.793) was more highly correlated with the transformed scores than were any of sets of d or t values. In the regression analysis, VOT accounted for a significant proportion of the variance in the transformed scores ($F(1,22) = 37.19$, $p<.01$). This time the most highly-correlated d value was d50. However, according to the new regression analysis, aspiration duration, d50, and d90 were not found to account for a significant proportion of the variance in the transformed scores after

VOT was entered, but d75 did make a significant contribution ($F(1,21) = 6.19$, $p<.05$) when used in combination with VOT.

To summarize, in both perceptual experiments VOT appeared to be a good predictor of the subjects' categorizations, and when VOT was used in combination with d75, even more accurate predictions were possible. While in Experiment 3, other predictors such as aspiration noise duration, d50, and t90 were found to be useful predictors when included with VOT, such was not the case in Experiment 4.

Chapter 7: Discussion and Conclusions

French and English Labial Stops: Comparison and Contrast

Release bursts and aspiration noise

In contrast with English labial stops, the word-initial European French labial stops examined here are characterized by significantly longer release bursts and, in the case of /p/, significantly shorter or non-existent aspiration phases, the durations of which are greater before high vowels than before non-high ones. Differences in burst durations may reflect differences in the amount of supraglottal air in the vocal tract at the time of release. It is also interesting to note that aspirated segments in both languages tend to have shorter bursts than unaspirated ones. This finding may reflect a greater overlap between the burst and aspiration phases than between burst and vowel when no aspiration is present. In other words the end of a burst may be less well defined when it is immediately followed by aspiration. As a result, because of the greater frequency of occurrence of aspiration in English (in both /p/ and /b/), English bursts may be measured as shorter than French ones.

While French burst amplitudes seem to be affected by vowel context more than English burst amplitudes (in French, bursts preceding /i/ are stronger than those preceding /ɛ/), there is no significant difference overall in burst amplitudes between the two languages, nor are there differences between /p/ and /b/ burst amplitudes within languages. Thus, no support is found here for the claim (Halle et al., 1957; Malécot, 1970) that differences in force of articulation result in stronger bursts for /p/ than for /b/. However, their claim is not refuted. It is conceivable that if the peak rather than the RMS amplitudes of the bursts had been measured, significant differences might have been found between /p/ and /b/. In fact, in both languages /p/ bursts are significantly *longer* than /b/ bursts (and the possibility that this difference has some importance as a perceptual cue cannot be ruled out). When the burst measurements were being made, it was noted that /p/ bursts in both languages often included a short period of fairly high amplitude followed by a significant period of low-amplitude energy frequently attributable to frication noise. This was less often true in the case of /b/. Consequently, it is possible that the RMS amplitudes of /p/ bursts (but not of /b/ bursts) might be considerably smaller than their peak amplitudes, and a difference in burst amplitudes may still exist in the /p/ - /b/ opposition. On the other hand, it is also possible that the burst and aspiration phases of stop consonants are not perceived as separate. Because of their short duration and their spectral properties, bursts may not be very salient when followed by audible aspiration noise.

The findings of the present study do refute the traditional claim that French /p/ is unaspirated (Delattre, 1965; Kim, 1970 and others). They also clarify the more recent findings of Benguerel et al. (1978) and Serniclaes et al. (1984) who have stated that French /p/ is slightly aspirated. It appears that the truth lies somewhere between these two claims. When a variety of dialects and speakers are investigated, word-initial French /p/ is *sometimes* (about 55% to 70% of the time) found to be accompanied by measurable aspiration, which may in some rare instances last as long as the aspiration associated with English /p/. In French, aspiration noise occurs more frequently and lasts for greater durations before high vowels than before non-high ones. In contrast, aspiration duration in the English /p/ tokens examined here is not vowel-contingent. A somewhat similar pattern is observed when aspiration amplitude is considered. While there is no overall difference between French and English in aspiration amplitude, like French burst amplitudes, French aspiration amplitudes seem to be influenced more by vowel context than do English ones.

Although the findings of the present study confirm to some degree the widely-held belief that glottal adduction is effected relatively early in initial French voiceless stops, the presence of significant aspiration also indicates that adduction is not necessarily completed prior to the release of the stop. The reasons for discrepancy between these and earlier findings are not entirely clear, but several possibilities may be suggested. One has to do with the widespread tendency of researchers to draw conclusions based on measurements of absolute initial stop consonants only. (Yet absolute-initial tokens probably constitute only a tiny minority of the total number of stops produced in natural speech.) It is possible that in running speech, precise control over the relative timing of glottal approximation and stop release may become more difficult. As a result, glottal adduction in French may be less likely to coincide exactly with or immediately precede the release of the stop than it is in the case of absolute initial stops. A second possibility is that the presence of aspiration in French may depend to some degree on the dialect of the speaker.[1] Thus, results of studies on very small subject pools representing a single dialect may be misleading. In the present research, a variety of regional dialects are represented, and these together with speaker idiosyncracies, including level of education, may have affected the results. Also, it may be the case that some registers in French are characterized by the presence of less aspiration than others. Another possibility is that a change in the way in which French stops are produced is in progress. Therefore, claims made some time ago (Nicholson, 1927) may be less true today than they were earlier. It is possible,

[1]There is some anecdotal evidence of such influences.

too, that some of the discussion of French voiceless stop production in pedagogical writings (Delattre, 1951; Valdman et al. 1970) is a hypercharacterization of the difference between the English and French categories. In other words, some writers may have wished to stress to English-speaking learners of French that there is generally less aspiration in French stops. Finally, one confounding variable in the present study is the fact that some of the French speakers have been long-time residents of western Canada. The extensive exposure which they have had to Canadian English may have had some effects on their stop production. In any case, there are large differences in the degree to which the individual French speakers aspirated (between, as well as within subjects), both in terms of frequency of aspirated tokens and the duration of aspiration when present.

Voicing during closure

When word-initial French /b/ occurs in running speech after a word ending in a vowel, it appears that voicing always continues into the closure interval of the stop, resulting in a voice bar. The same is usually true in English, but in a very small minority of cases (fewer than 4%), no measurable voicing occurs during closure. These findings are consistent with the findings of others who have investigated European French stop consonants in absolute initial position - all occur with voice bars (Caramazza and Yeni-Komshian, 1974) - and more or less consistent with the results of similar studies of English (Lisker and Abramson, 1967), except that a strong focus on absolute initial English stops may lead one to underestimate the extent to which some degree of voicing during closure actually occurs in English stops within sentences (cf. Smith, 1978; Flege, 1982). Flege also found that almost all absolute-initial prevoiced /b/s were produced without voice breaks. However, the present study indicates that this is not the case for initial /b/s produced in context. Caution must therefore be exercised when drawing conclusions about the production and perception of natural speech based on evidence from studies in which words are produced completely out of context.

One problem which appears not to have been clearly addressed in the literature is the question of how long an English voice bar must be before it can be deemed to have "cue status." The results of the present study reveal that audible voicing during a closure following a vowel may last for as little as 27 msec. and may be followed by a significant silent period. Such voice bars may have little perceptual relevance, especially if they are of low amplitude. However, longer ones, even though they may be accompanied by voice breaks, may play a significant role as perceptual cues. Lisker and Abramson (1964, 1967) were careful to point out that when making their VOT measurements they ignored inaudible "edge vibrations" which they felt were perceptually unimportant. Their discussion makes clear the importance of distinguishing measurable but perceptually irrelevant properties of speech from characteristics which are important from the listener's point of view. But all of the voice bars measured in Experiments 1 and 2 were audible to some extent; hence the researcher who wishes to evaluate the role of closure voicing in English does not have available any simple basis for rejecting the possibility that this phenomenon is perceptually important. The fact remains that the only way to answer the question of what constitutes a perceptual cue is to perform perceptual experiments in which the efficacy of the cue in question is evaluated. A need for further exploration of this problem is clearly indicated.

Despite the high frequency of occurrence of glottal activity during stop closure in both languages, the presence of voice breaks in English but not in French suggests some significant difference in the production mechanisms associated with

/b/ in the two languages. While it is not possible here to draw firm conclusions about the precise mechanisms involved, it is conceivable that in English /b/ (following Warren, 1976), but not in French /b/, some limit on the expansion of the oral cavity is reached before release. This may occur because the cavity is not allowed to expand as much in English as it is in French or because the size of the cavity is larger in French. Another possibility is that the velo-pharyngeal seal is not as tight in French, and as a result, some air may escape through the nose. It appears that the difference described here may involve factors other than simple differences in timing. That is, English voice bars do not simply "start sooner" and therefore "die out" sooner than French ones; in fact, the English voice bars are significantly shorter than the French ones.

A further difference between /b/s in the two languages is the occasional presence of short, low-amplitude aspiration after English /b/, which contrasts with a nearly total absence of aspiration in French /b/. Again, while it is not possible to use measurements from waveforms to draw precise conclusions about glottal events, it may be that some degree of voicing continues from the closure right through the stop release in virtually every French /b/, thus prohibiting aspiration. In English, since voice breaks regularly occur, some slight aspiration may be observed just before glottal pulsing resumes. Interestingly, there is no significant correlation between the amount of closure voicing in English /b/ and the amount of aspiration following release. If both closure voicing and aspiration are effects of a single underlying property of stop consonants (i.e. voice timing) as Lisker and Abramson (1964, 1967, 1971) argue, one might expect some degree of negative correlation here. However, it is true that in those English /b/s in which voicing continues unbroken throughout the closure interval, aspiration is unlikely to occur.

In fact, only three of the fourteen /b/s which were "fully voiced" showed any measurable aspiration.

Some of the findings of the production experiments described here agree with those of Smith (1978), who found that voice bar durations in English were significantly longer in the context of high vowels than they were before low ones. He speculates that the differences may be attributable to larger supraglottal cavity volumes for high vowels. In the case of voice bars produced with larger cavity volumes, it may take longer for supra- and subglottal air pressures to become equal, and glottal pulsing may last longer. In Experiment 1, English /bi/ and /bu/ were found to have significantly longer voice bars than /bæ/, and French /bi/ and /bu/ had longer voice bars than /ba/; also, in both languages, /bi/ and /bu/ had voice bars of higher amplitude than /ba/. These findings give some support to Smith's claims because one might expect such differences in durations to occur across languages. However, this question still requires further exploration.

Closure durations

The finding of longer closure durations in some instances of initial French and English /b/ than in /p/ in Experiment 1 does not agree with earlier findings for medial stops in English (Lisker, 1957) and for post-vocalic stops in French (Wajskop and Sweerts, 1973), although it has been pointed out (Lisker, 1978) that closure durations probably have little perceptual importance for English stops in (presumably absolute) initial position. The difference found between categories in Experiment 1 is quite small (about 12 msec.) in comparison with the mean overall closure duration (over 100 msec.) It should be noted that there is considerable variation in closure durations, both among and within subjects, in the present study an[illegible]rs. In Experiment 1, several subjects from both language groups often produced [illegible]/s with longer closure durations than /b/s. Flege (1982), also reports that four of ten English-speaking subjects produced /b/s with average closure durations longer than those of /p/s in absolute initial position. Given the small degree of difference between categories found here, the considerable variation among subjects, and the probable susceptibility of closure durations to strong effects of speaking rate, it seems likely that Lisker's claim is valid for both English and French initial stops produced within sentences. Moreover, the frequent mention of differences in closure duration as an argument for a "fortis-lenis" account of stop production or perception (e.g. Malécot, 1970) without due attention to the context or contexts in which such differences are found to have perceptual relevance (see Lisker, 1978) appears to be misleading.

Voice onset time

Despite the problems in measuring VOT resulting from voice breaks in English (see Chapter 3), no matter how VOT is measured (from the onset of a voice bar, even if there is a voice break, or from the beginning of the release burst to the end of aspiration), it is virtually always possible to distinguish /p/ from /b/ in the environments under consideration here on the basis of VOT alone in both English and European French. These findings therefore agree with those of Caramazza and Yeni-Komshian (1974) in that it appears that in European French initial stops, voice timing clearly separates the categories and is thus a plausible perceptual cue. Nevertheless, perceptual experiments of the type performed by Caramazza et al. (1973) on Canadian French subjects are still needed to demonstrate that European French listeners are indeed sensitive to timing differences in the same way that English listeners are.

Differences in VOT are clearly useful in describing differences in stops *between* the two languages as well. Examining the VOT "axis," one finds that the negative side is occupied by French /b/ which is confined exclusively to that region. This category overlaps with English /b/, but the degree of overlap depends on how VOT is defined. In the positive short- to moderate-lag region of the continuum, overlapping with short-lag English /b/, is French /p/. Finally, on the extreme positive side is English /p/, which overlaps to some extent with French /p/. These findings generally confirm Serniclaes et al.'s (1984) informal reporting of mean VOTs for English and French.

French /p/ and English /b/: Implications of the Experiments

The production experiments described in Chapters 3 and 4 indicate that there are some important differences between French /p/ and English /b/ in running speech. The most obvious of these is of course the presence of some voicing during the closure interval of English /b/. But even when a VOT interval is defined as the sum of release burst and aspiration noise durations (except when voicing is present throughout an entire closure interval), French /p/ is produced with a mean VOT which is longer than that of English /b/. This is true because the former has longer burst durations *and* is frequently accompanied by longer aspiration noise.

The perception experiments described in Chapters 5 and 6 indicate that English-speaking listeners are able to distinguish French /p/ from English /b/ reliably between 65% and 75% of the time (at least under the nearly ideal listening conditions of the experiments). Naive listeners who have no experience with French may sometimes be less successful at this task than phonetically-trained listeners who have studied French; there is clear evidence here, however, that listeners need not have experience with French in order to make correct identifications at above-chance levels. In fact, the subject with the highest correct identification score in Experiment 4 was a naive listener. Analysis of the subjects' identification scores in both experiments reveals that voice onset time alone is a fairly good predictor of how subjects will categorize individual tokens. In general, tokens with long VOTs are categorized as French /p/s. Of course, this finding does not prove that listeners actually use VOT in making their judgments, but at the very least it suggests that English-speaking listeners, regardless of their linguistic experience, are fairly sensitive to the acoustical consequences of slight differences in laryngeal timing which are smaller than those normally found between initial stop categories in English. Moreover, VOT is a fairly stable predictor, accounting for about 60% of the variance in subjects' categorizations in both experiments.

When the time at which the RMS amplitude of a token reaches a critical percentage of maximum (perhaps about 75%) is taken into consideration along with VOT, more accurate predictions of subjects' categorizations of a set of French /p/-English /b/ stimuli can be made than on the basis of VOT alone. This finding confirms to some extent the findings of Darwin and Pearson (1982), who used synthetic stimuli to test the perceptual significance of the point at which a critical intensity level is reached in a stimulus. Their results suggested that an overall intensity measure should be a more accurate predictor of when voicing is perceived to onset than VOT alone, because such an intensity measure more closely corresponds to perceptual events than does VOT. In other words, an intensity measure may be thought of as a refinement of the notion of VOT. However, the *d* values calculated for the tokens used in Experiments 3 and 4 were *not* found to be better predictors of subjects' categorizations than VOT alone. One possible explanation for such a finding is that while voicing is perceived to onset when the amplitude of the signal reaches some critical value, as Darwin and Pearson argue, the methods of measurement and calculation of the *d* values used here may be too imprecise. The experiments performed here were very exploratory. Perhaps a larger or smaller integrating window should have been used when the amplitudes of the stimuli were measured, or perhaps the peak amplitudes of the periodic portions of the stimuli should have been calculated in a different manner.

Nonetheless, when amplitude and VOT measurements are used *together* as predictors of subjects' performance, more accurate predictions are obtained than with VOT or *d* values alone. Therefore, a second possibility is that when the VOT and *d* values are taken together, the result is a more accurate measure of rise time as it is perceived by the listener than are the *d* values alone. It may be that the time of voice onset serves as a reference point within a signal which when combined with an amplitude measure provides a fairly accurate measure of the rise time of the periodic portion of the stimulus. If the VOT values are subtracted from the d75 values, for instance, the result is the length of time it takes for the amplitude of the *vowel* to increase from about 0 to 75% of maximum. The *d* values alone, however, indicate rise time in a much cruder sense: they simply indicate the time at which the stimulus reaches some critical amplitude regardless of the type of excitation of the vocal tract. Thus, the amplitude characteristics of a stop + vowel token may still be a plausible perceptual cue to whether the stop is a voiced or voiceless one, except that, contrary to the findings of Darwin and Pearson, listeners may focus mainly on the rise time of the periodic portion of a stimulus. The variance in the subjects' identification scores in the two experiments which is not attributable to VOT and d75 is still fairly large and may in part be the result of spectral or other stimulus characteristics which were not measured in Experiments 3 and 4.

Once again, the fact that an amplitude measure such as the d measure used here is useful in accounting for listeners' categorizations of natural tokens does not prove conclusively that listeners necessarily use amplitude characteristics as a basis for their judgments; however, the findings of the present study in which natural tokens have been used, taken together with the findings of Darwin and Pearson (1982), provide encouraging evidence that rise time may play an important role in distinguishing perceptual categories. These results also suggest that the simple physical measure VOT does not necessarily have a simple perceptual correlate. Obviously, considerably more research into this problem is required. In particular, it may be possible to manipulate natural speech tokens in such a way as to dissociate VOT and one of the d measures, for instance, in a systematic way rather than rely on chance differences which occur in naturally produced tokens in the present study.

Also, an interesting question arising from the perceptual experiments performed here, which deserves serious attention, is that of how /p t k/ are distinguished from /b d g/ in whispered speech in both English and French. Since whispered syllables are produced with no periodic energy, temporal cues other than VOT (such as overall intensity) may play a role in the perception of such sounds. To date there has been very little discussion of whispered speech (and especially of whispered consonants) in the phonetics literature. Yet the hypotheses presented here and those of Darwin and Pearson make some interesting predictions about how whispered speech may be perceived.

The fact that aspiration noise duration also seems to figure in subjects' categorizations of stimuli in Experiment 3 is not in itself surprising; however, the direction of the correlation in the regression equations was unexpected. Nine of the thirty-two tokens used in Experiment 3 included some short but measurable aspiration noise, and according to the regression equations, the duration of this noise is *negatively* correlated with the number of /p/ identifications for a particular stimulus. The reasons for this phenomenon are not clear, although one possible explanation is that it is not the aspiration noise durations of the stimuli which are important here, but the burst durations, which are equal to the difference between VOT and aspiration duration, both of which were included in the regression analysis. As mentioned earlier, aspirated consonants in both languages tend to have shorter bursts than unaspirated ones. Furthermore, in both languages /p/ bursts are significantly longer than /b/ bursts, and it is possible that this difference has some limited value as a perceptual cue. If so, the subjects may have tended to categorize the tokens with short bursts (i.e. the aspirated ones) as /b/s. Nevertheless, aspiration noise duration was *not* found to play a significant role as a predictor in Experiment 4. The reasons for this discrepancy are not clear.

Implications for the Perception of Categories from a Foreign Language

The results of the perceptual experiments performed here can be viewed as confirming to some extent Flege's (1984) claim that humans are highly sensitive to "divergences from the phonetic norms of their native language." Although the French /p/ category is very close in several respects to English /b/, most subjects were able to attend to relatively small divergences from the normal English /b/. This may be, as he suggests, because they had access to a "phonetic prototype" against which they were able to compare stimulus tokens. In this case, the stimulus items may have been "measured" against a prototypical English /b/ and categorized according to how well they correspond to that prototype. In the case of the experienced listeners, a prototype of French /p/ may also have been used, while in the case of the naive listeners, an English /p/ prototype may have been accessed.

However, it should be pointed out that such an explanation is not a unique one for the results. It is possible that the subjects' categorizations were in part the result of anchoring or range effects (Simon and Studdert-Kennedy, 1978; Brady and Darwin, 1978; Keating et al., 1981). Therefore, comparisons may have been made not only against some internalized standard but against tokens from within the stimulus set itself. That is, some tokens clearly differed from each other in terms of rise time, VOT characteristics, or both, and the subjects were able to reliably separate the stimulus set into two categories on the basis of these properties. The fact that the naive subjects were willing to classify some of the tokens as /p/s at all suggests that this is the case. If these listeners, who had no knowledge of French, had compared the stimulus items exclusively against prototypes of both English /b/ and English /p/, they might well have categorized all tokens as /b/ because the majority of them were much closer to English /b/ than to English /p/ in terms of VOT and other characteristics. That is, there were no "good English /p/" tokens in the stimulus set. On the other hand, because different patterns of response bias were observed among the experienced listeners and the successful naive ones, range and anchoring effects cannot completely account for the results. In other words, two groups of listeners hearing the same stimulus set, showed slightly different categorizing strategies, presumably because of differences in their linguistic experience. Among the experienced listeners in both experiments, there was a slight bias in favour of /p/ responses while among the successful inexperienced listeners in Experiment 4, there was a preponderance of /b/ responses. The experienced listeners were more willing to label some of the ambiguous items as /p/ perhaps because they were able to compare the tokens with a prototypical French /p/, which is characterized by a relatively short VOT and rise time in comparison with English /p/. The inexperienced listeners, on the other hand, may have felt that most tokens were quite close to English /b/ and therefore may have preferred to categorize as /p/s only those tokens with fairly long VOTs and fairly slow rise times (i.e. the tokens closest to English /p/).

From the perspective of an English speaker learning French, the results of the perceptual experiments discussed here have some significance. The fact that at least some naive listeners can reliably hear differences between the categories indicates that they are sensitive to relatively small cross-linguistic differences in superficially similar categories. An inability on the part of English-speaking learners to actually produce such differences therefore would not necessarily stem from an inability to perceive them (cf. Flege, 1984). It should be pointed out that if the subjects in both experiments had received training, including feedback on their performance, they might have performed even better. Pisoni et al. (1982) have in fact demonstrated that laboratory training procedures can lead to significant improvements in listeners' performance on tasks similar to these. It also seems plausible that if subjects are able to perceive small phonetic differences between cross-linguistic categories, there is a good chance that they can learn to produce them. At this point it is tempting to speculate about the future possibility of applying laboratory training procedures to facilitate subjects' learning to produce a contrast such as the /p/ - /b/ distinction investigated here. Flege (1984) has found evidence that a foreign accent can be detected in very short segments of speech and that such an accent can sometimes be detected on the basis of VOT. If rapid feedback could be given to second-language learners indicating the accuracy of voice onset time and rise time characteristics of consonants produced by them, perhaps they could be trained to modify their production appropriately.

A few shortcomings of the present study should be noted. First, there were some difficulties encountered in Experiment 1 because of the equipment used. In particular, the use of different tape recorders and microphones for the two language conditions was found to have effects on the amplitude measurements made in that experiment. However, no effects on the duration measurements were found to occur. These observations suggest that comparisons of the results of experiments performed by different researchers using different equipment may be acceptably valid insofar as they entail durational measures such as VOT, burst durations, and so forth. Other comparisons involving spectral and intensity measurements should be made with caution. As well, there are some difficulties in drawing firm conclusions about the results of Experiment 2 because of the small sample size. The fact that one subject's data seemed to differ somewhat from the others may be due to dialectal or idiolectal differences.

Tables

Table 3-1: Mean durations of segments (in msec.)

Consonant	n	Sil. Per.	V. Bar	Burst	Noise	VOT
French						
/p/ (unasp.)	36	110.7	0.0	13.6	0.0	13.6
/p/ (asp.)	44	89.6	0.0	7.6	26.5	34.3
/p/ (total)	80	99.1	0.0	10.4	14.6	25.0
/b/	80	0.0	107.8	7.6	0.0	-107.8
English						
/p/	56	105.0	0.0	6.9	56.2	63.1
/b/ (sil. p.)	42	61.1	65.4	5.6	2.9	8.5
/b/ (no s. p.)	14	0.0	111.5	7.0	1.5	-111.5
/b/ (total)	56	45.9	76.9	5.9	2.5	—

Table 3-2: Mean relative amplitudes of segments*

Consonant	n	V. Bar	Burst	Noise
French				
/p/ (unasp.)	36	0.00	0.16	0.00
/p/ (asp.)	44	0.00	0.13	0.13
/p/ (total)	80	0.00	0.14	0.07
/b/	80	0.32	0.27	0.00
English				
/p/	56	0.00	0.26	0.37
/b/ (sil. p.)	42	0.13	0.24	0.06
/b/ (no s. p.)	14	0.15	0.25	0.04
/b/ (total)	56	0.14	0.24	0.05

* Amplitudes are expressed as proportions of the amplitudes of the following vowels.

Table 3-3: Aspirated French /p/ in Experiment 1

Vowel	Proportion of aspirated /p/s	Mean duration of aspiration when present (msec.)
/a/	5/20 (25%)	11.8
/i/	16/20 (80%)	30.2
/ɛ/	6/20 (30%)	11.1
/u/	17/20 (85%)	30.5
Total	44/80 (55%)	26.5

Table 3-4: Mean aspiration duration by subject

Subject	n (asp. tokens)	mean asp. dur. when present (msec.)
S1	2	11.6
S2	5	19.9
S3	2	20.8
S4	7	17.2
S5	4	26.6
S6	6	36.2
S7	3	33.1
S8	6	31.3
S9	4	18.4
S10	5	39.0

Table 3-5: VOT in French /p/ by vowel

CV	Mean VOT (msec)	VOT range (msec.)
/pa/	14.1	7.1 - 30.1
/pi/	32.6	8.7 - 69.6
/pɛ/	14.5	7.3 - 26.3
/pu/	38.6	11.2 - 65.5

Table 4-1: Aspirated French /p/ in Experiment 2.

Vowel	Proportion of aspirated /p/s	Mean duration of aspiration when present (msec.)
/a/	3/6 (50%)	8.4
/i/	6/6 (100%)	36.0
/ɛ/	2/6 (33%)	8.2
/u/	6/6 (100%)	28.7
Total	17/24 (71%)	25.3

Table 4-2: Mean amplitude of aspiration noise by subject

Subject	Aspirated tokens	Amp. of asp.
S1	4/8	0.33
S2	7/8	0.14
S3	6/8	0.60

Table 4-3: Mean amplitudes of segments

Consonant	n	V. Bar	Burst	Asp.
/p/ (asp.)	17	0	0.43	0.35
/p/ (unasp.)	7	0	0.10	0
/p/ (total)	24	0	0.34	0.25
/b/	24	0.15	0.40	0.12

Table 4-4 (a): Mean durations of segments for S2 (in msec.)

	/p/			/b/	
	s.per	burst	asp.(n)	v. bar	burst
Exp. 1	89.9	9.5	33.1(3)	116.8	4.7
Exp. 2	96.4	7.1	23.3(7)	110.9	4.4

Table 4-4 (b): Mean amplitudes of segments for S2

	/p/		/b/	
	burst	asp.(n)	v. bar	burst
Exp. 1	0.13	0.08(3)	0.29	0.33
Exp. 2	0.27	0.14(7)	0.22	0.38

Table 4-5: Mean relative amplitudes for both samples

	burst	asp.	v. bar
Exp. 1	0.2067	0.0692	0.3167
Exp. 2	0.3676	0.2450	0.1505

Table 5-1: Correct identification scores (%) by subject

Subject	English	French	Total
S1	69.1	74.4	71.7
S2	58.1	58.1	58.1
S3	79.4	82.5	80.9
S4	57.5	50.6	54.1
S5	58.8	85.6	72.2
S6	68.8	83.4	76.1
S7	65.9	78.8	72.3
S8	61.7	72.7	67.2
Total	65.0	73.3	69.1

Table 5-2: Means and ranges of selected predictors

Predictor	Mean /p/	Mean /b/	Range /p/	Range /b/
VOT (msec.)	12.00	6.90	7.10-25.50	0-17.50
d75 (msec.)	24.14	12.94	16.36-34.61	9.95-17.65
t25 (msec.)	11.93	4.89	4.13-19.30	2.56-10.52
asp. (msec.)	3.23	1.73	0-14.90	0-12.00
na8*	0.2418	0.5338	0.0577-0.5357	0.0942-0.7291

*expressed as a proportion of mean peak

Table 5-3: Correlations among selected predictors and /p/-id scores

	asp. dur.	t25	d75	na8	VOT
t25	0.444				
d75	0.210	0.645			
na8	-0.492	-0.885	-0.605		
VOT	0.737	0.675	0.443	-0.761	
/p/-id	0.320	0.634	0.649	-0.713	0.767

Table 5-4: Characteristics of selected stimulus items.

Stim.	VOT	Nat. (msec.)	/p/-id Cat.	d75 (%)	t25 (msec.)	na8* (msec.)
16	6.1	/b/	22	12.15	3.97	0.6622
7	6.1	/b/	46	17.65	5.27	0.5239
15	7.0	/b/	17	10.12	4.92	0.4187
24	7.1	/p/	58	16.36	5.53	0.3703
32	8.0	/p/	45	25.65	15.19	0.2734
29	8.0	/p/	83	34.61	4.72	0.5357
23	8.4	/p/	46	17.16	4.53	0.4589
30	8.4	/p/	84	24.66	4.13	0.5268
25	13.0	/p/	44	17.49	14.44	0.1916
18	12.8	/p/	82	23.00	12.62	0.1832

*expressed as a proportion of mean peak

Table 5-5: Regression equations

Equation	R2 (%)
1. 0.02 + 0.0695 VOT - 0.0377 asp. dur + 0.0163 d75	77.6
2. 0.162 + 0.084 VOT - 0.0445 asp. dur.	69.9
3. 0.074 + 0.432 VOT + 0.0197 d75	68.7
4. 0.076 + 0.0478 VOT + 0.0128 d90	66.0
5. 0.221 + 0.432 VOT + 0.0186 d50	63.3
6. 0.140 + 0.494 VOT + 0.00939 t90	61.3
7. 0.322 + 0.0555 VOT	57.4
8. 0.866 - 0.884 na8 + 0.0176 d75	55.4

Table 6-1: Correct identifications by listener (%)

Subject	English /b/	French /p/	Total	/p/-id
Naive				
S1	85.4	51.0	68.2	33
S2	92.7	58.3	75.5	33
S3	72.4	59.4	65.9	43
S4	22.9	71.9	47.4	74
S5	33.9	65.6	49.7	66
Total	61.5	61.2	61.4	50
Experienced				
S6	60.9	81.3	71.1	60
S7	63.0	76.6	69.8	57
S8	59.4	75.0	67.2	58
S9	58.3	74.0	66.1	58
S10	77.1	68.2	72.7	46
Total	63.7	75.0	69.4	56
Total (overall)	62.6	68.1	65.4	53

Table 6-2: Comparison of results (Exp. 3 and 4)

Equation	*s* in Exp.3	RMS error in Exp. 4
VOT/asp. dur./ d75	0.1731	0.2779
VOT/asp. dur.	0.2005	0.2962
VOT/d75	0.2047	0.2463
VOT/d90	0.2132	0.2534
VOT/d50	0.2214	0.2425
VOT/t90	0.2275	0.2453
VOT	0.2386	0.2545
na8/d75	0.2444	0.8398

Table 6-3: Correlations among selected predictors and /p/-id scores

	asp. dur.	t10	d50	d75	VOT
t10	0.574				
d50	0.319	0.689			
d75	0.285	0.792	0.660		
VOT	0.707	0.587	0.529	0.228	
/p/-id	0.345	0.773	0.567	0.464	0.793

Figures

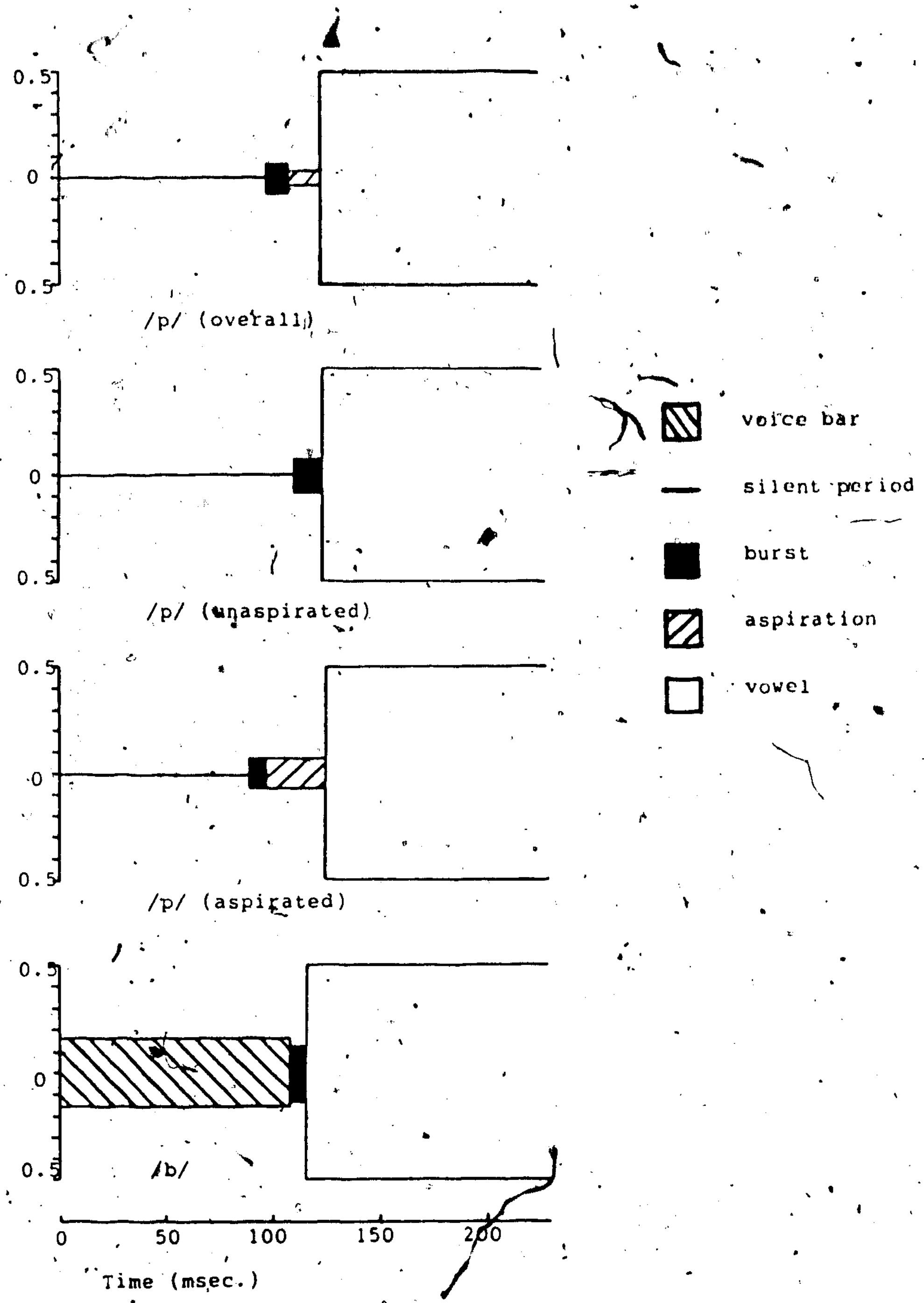


Figure 3-1: Schematic representation of average cases (French)

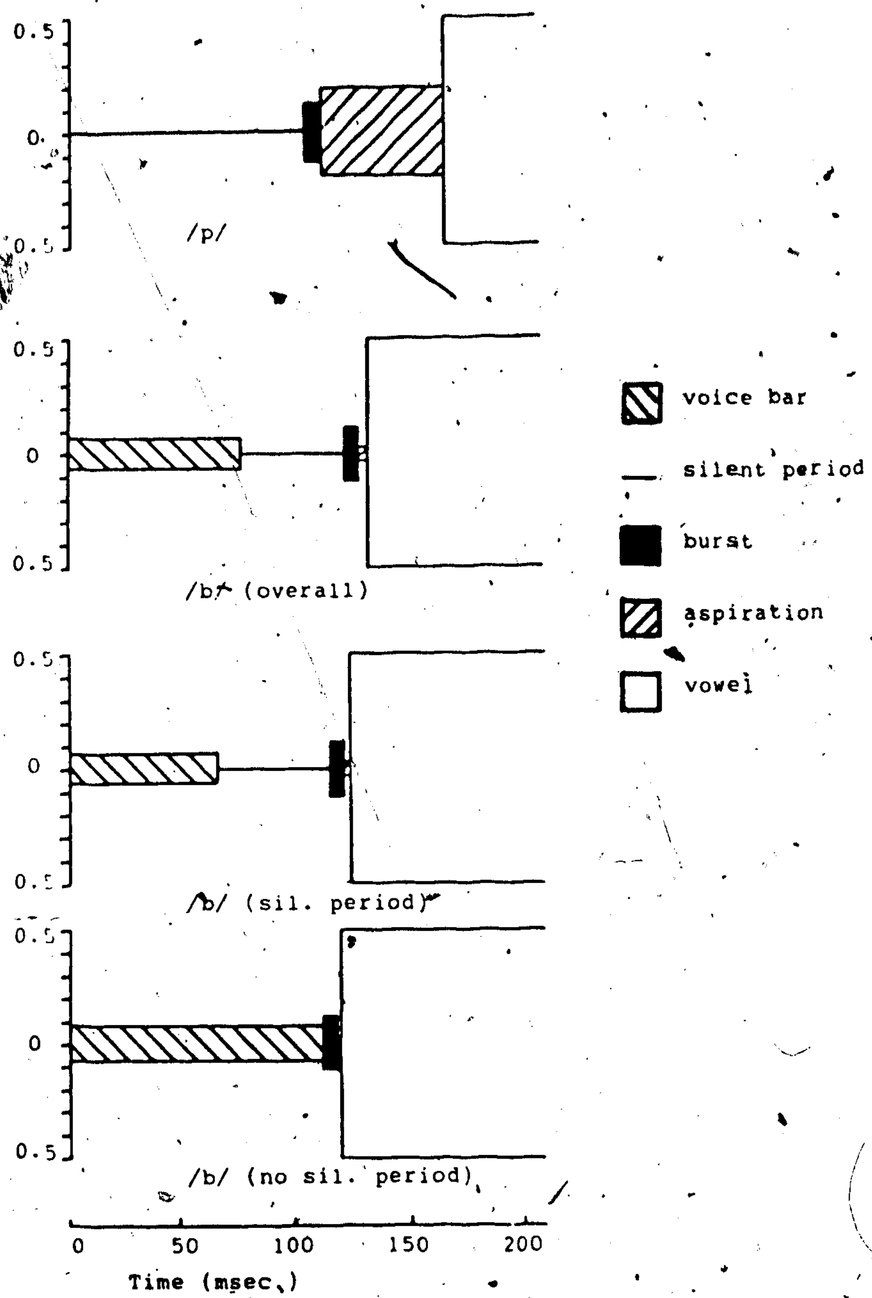


Figure 3-2: Schematic representation of average cases (English)

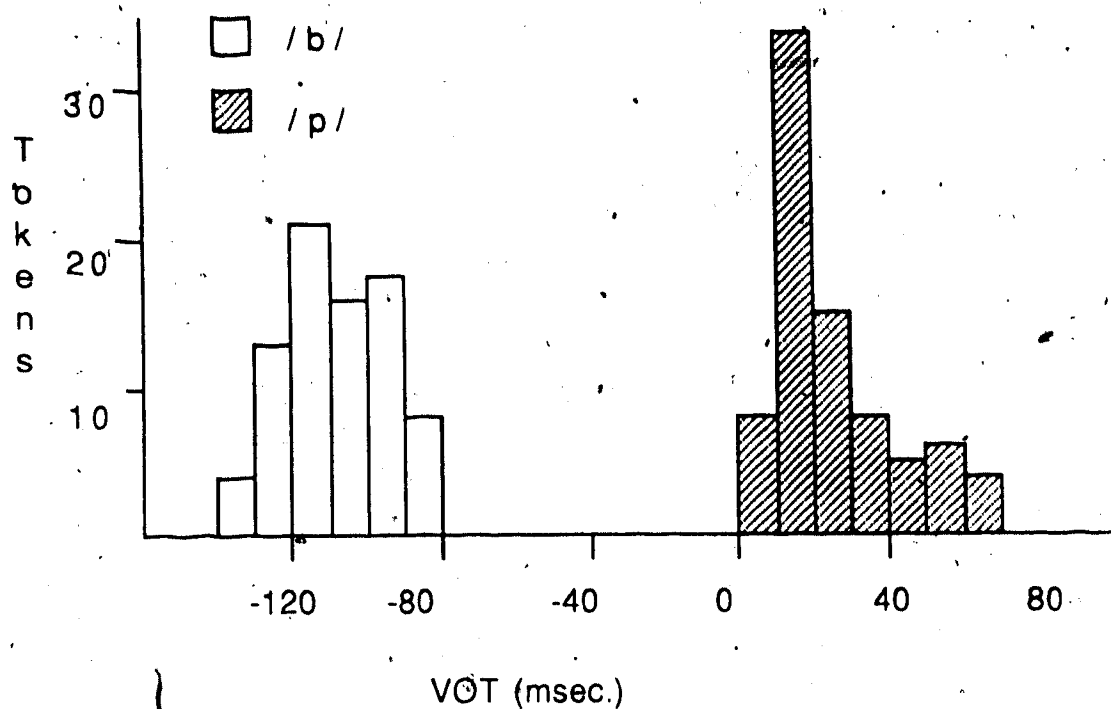


Fig. 3-3: Voice onset time in French (Experiment 1)

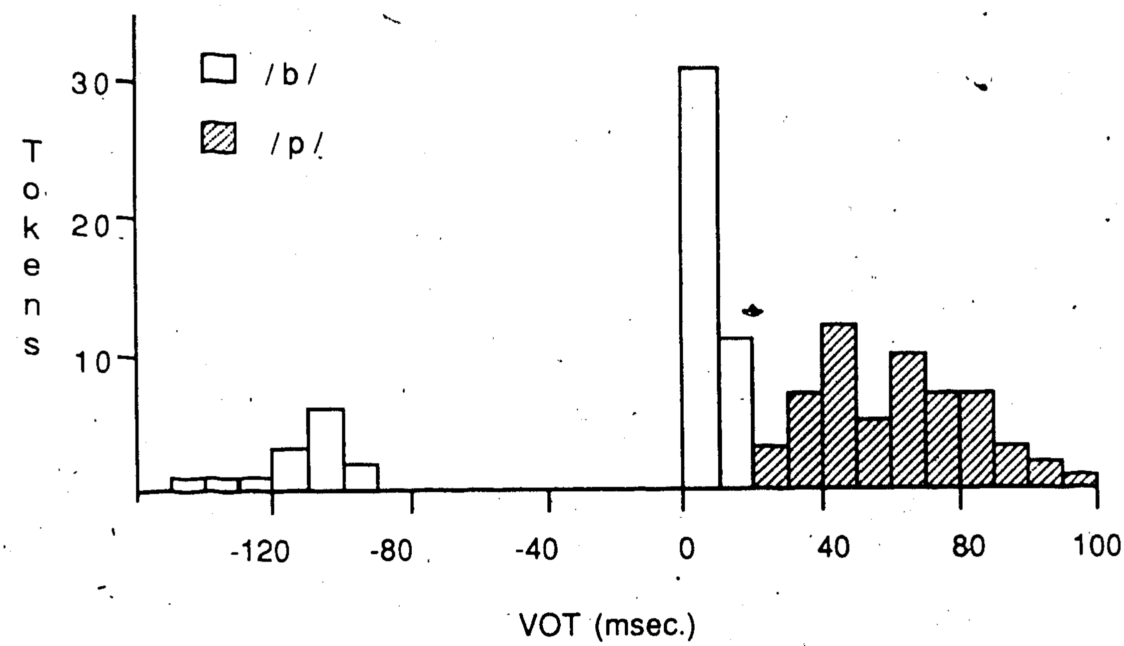


Fig. 3-4: Voice onset time in English (Experiment 1)

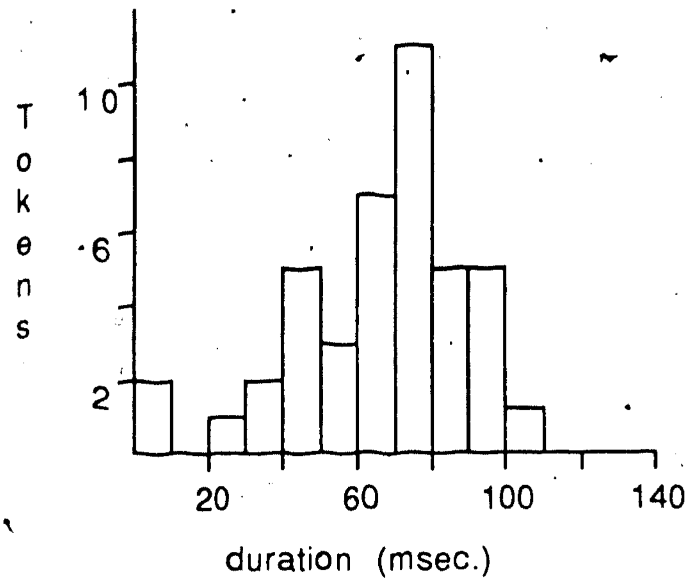


Fig. 3-5: Durations of "interrupted" voice bars in English /b/

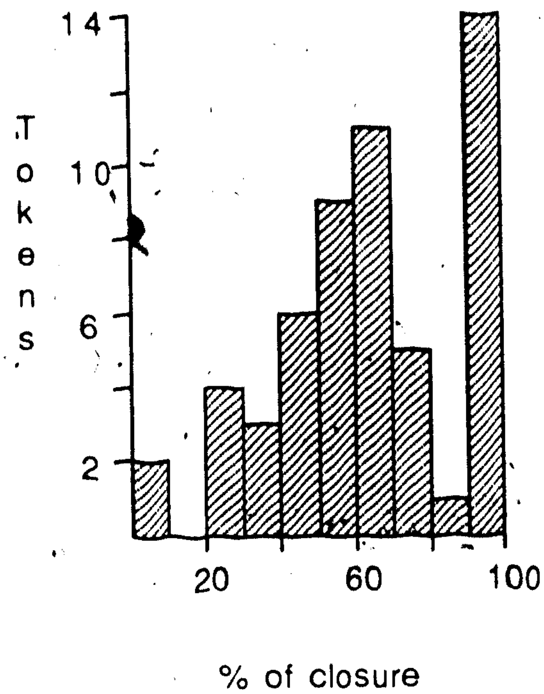


Fig. 3-6: % of closure interval which is voiced in English /b/

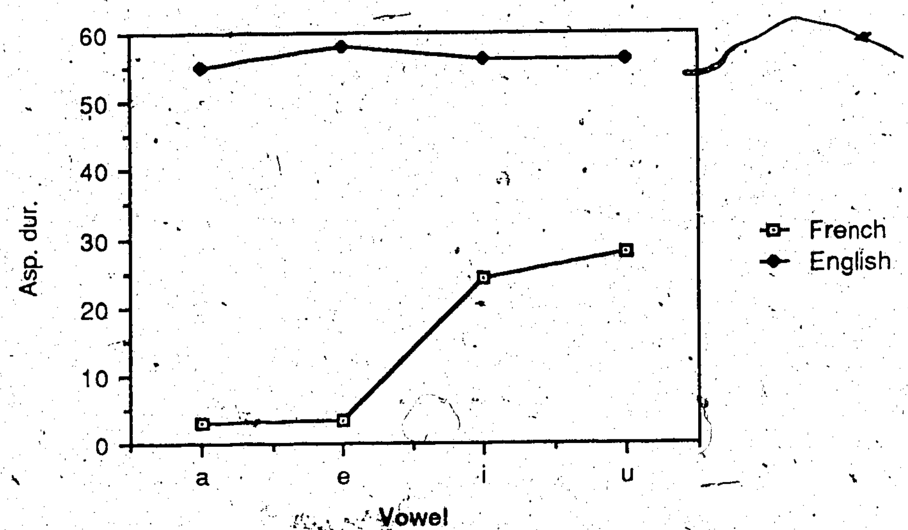


Fig. 3-7: Aspiration duration (msec.) — VL interaction

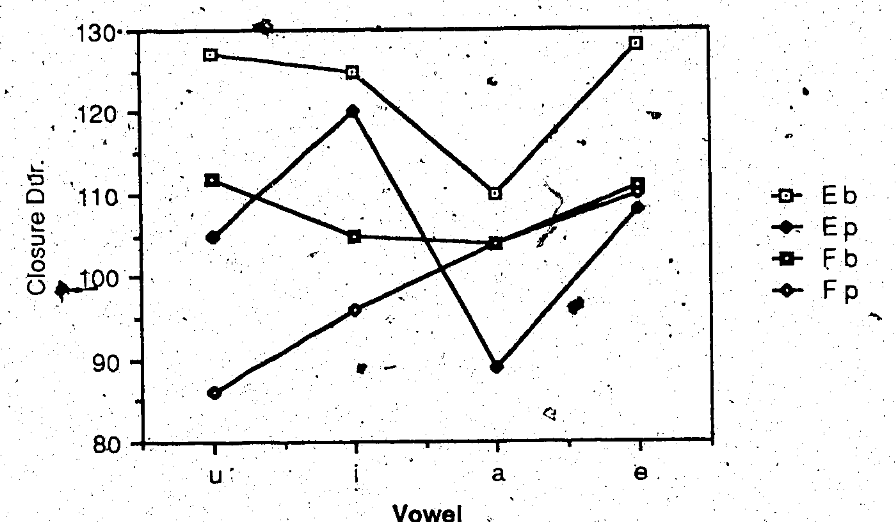


Fig. 3-8: Closure durations (msec.) — VCL interaction

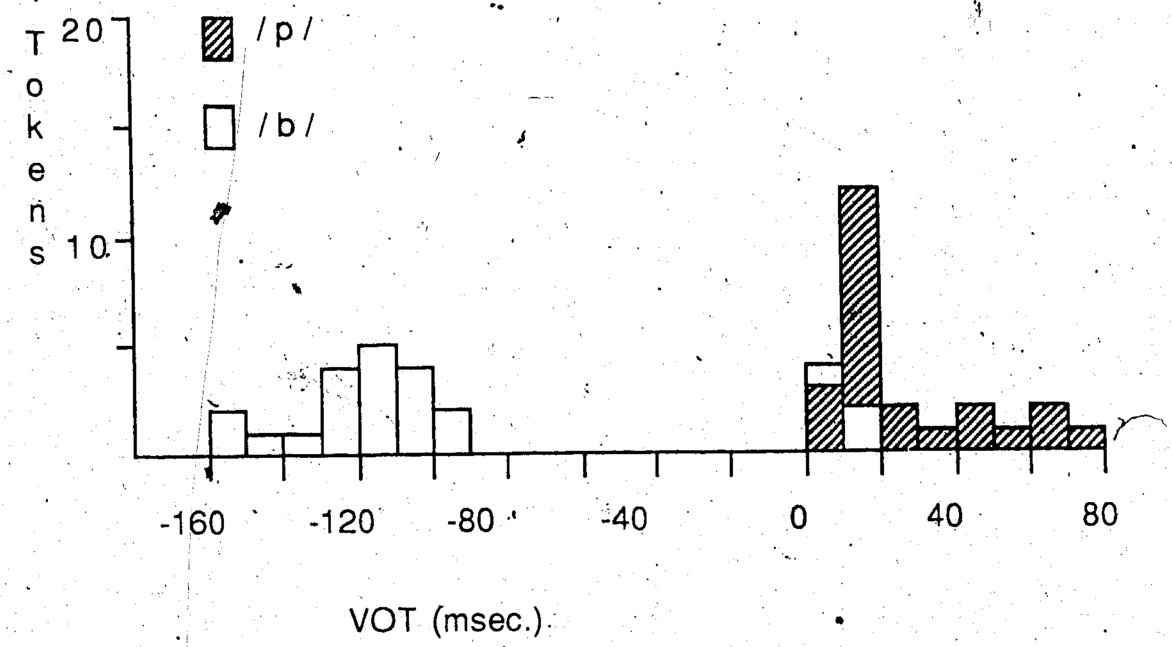


Fig 4-1: Voice onset time in French (Experiment 2)

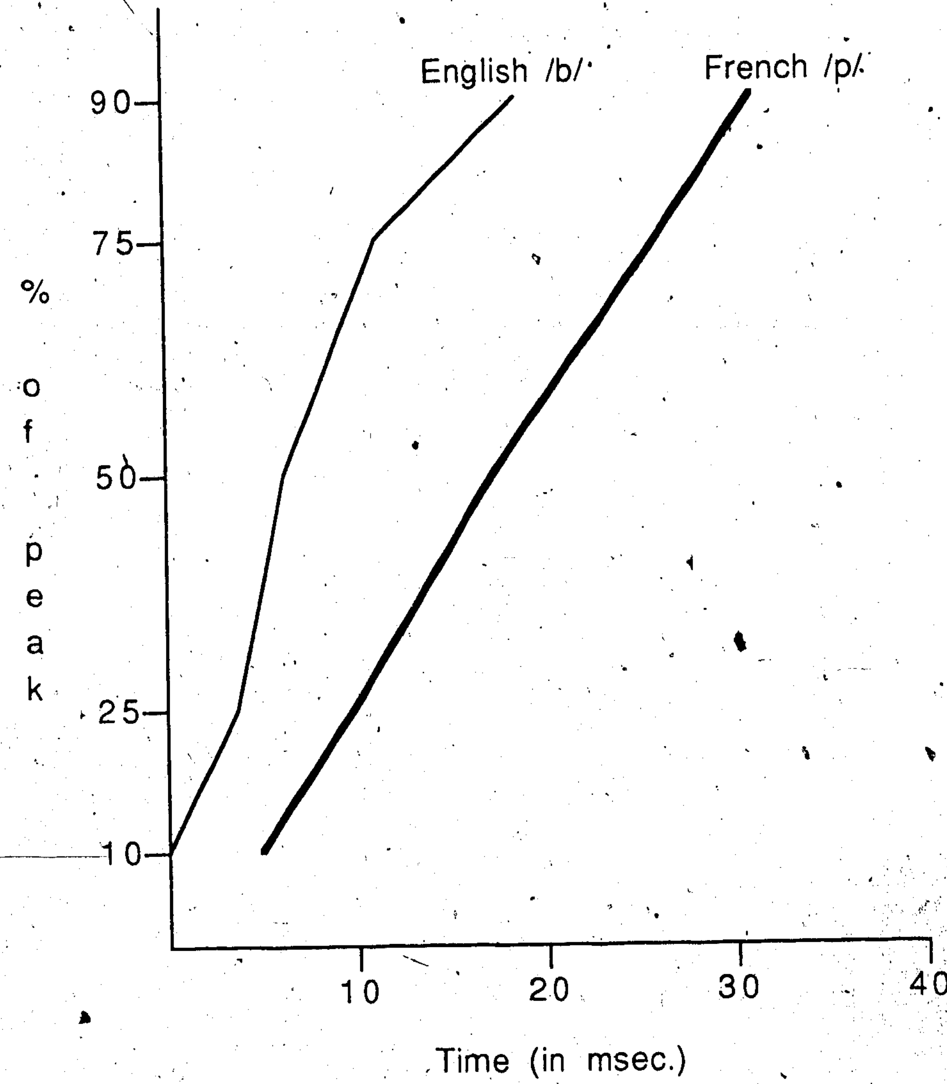
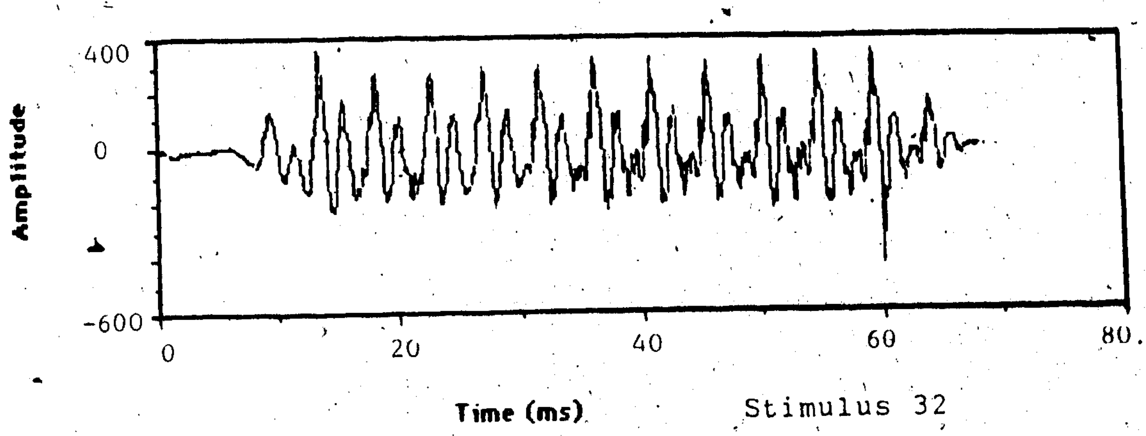


Figure 5-1: *d* values



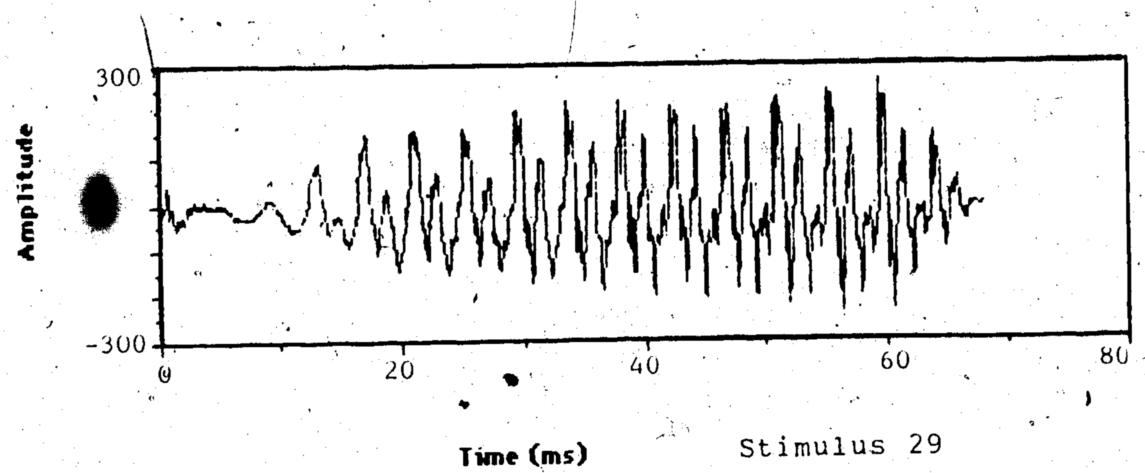
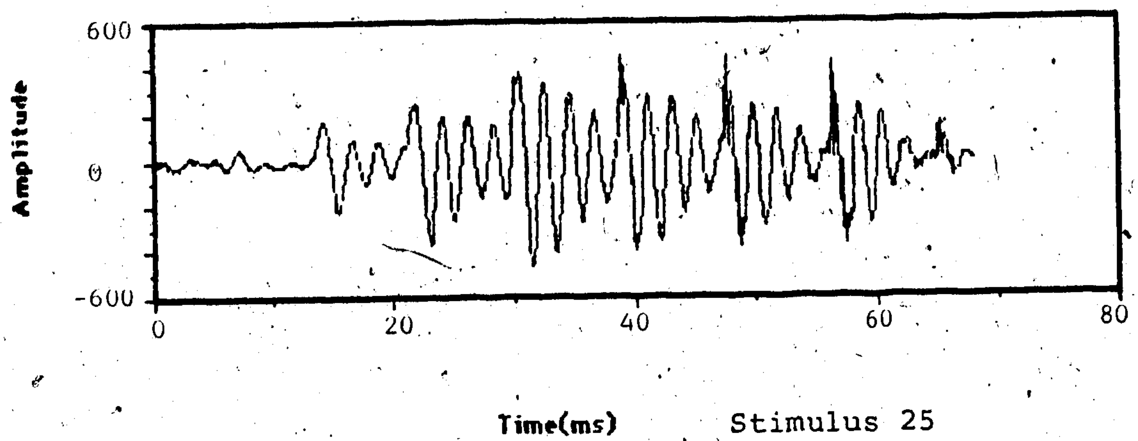


Figure 5-2: Sample waveforms



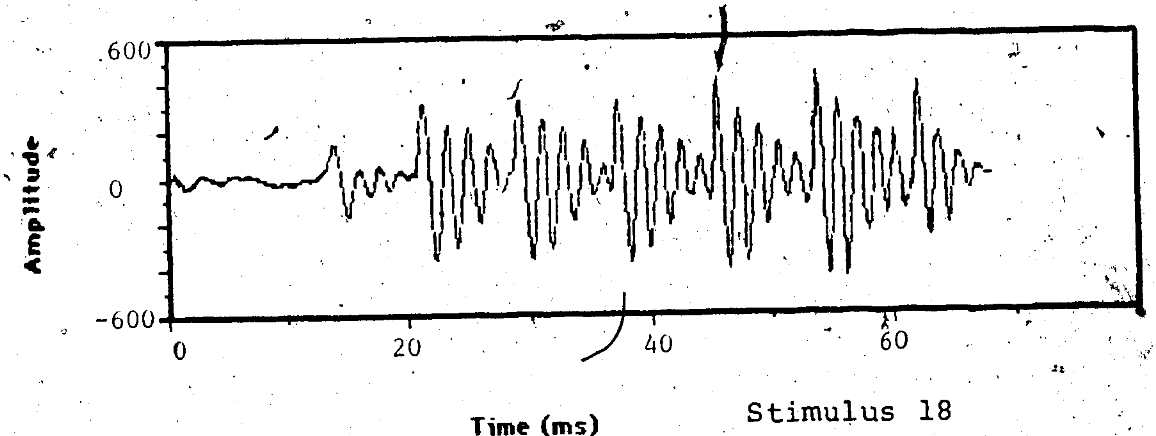
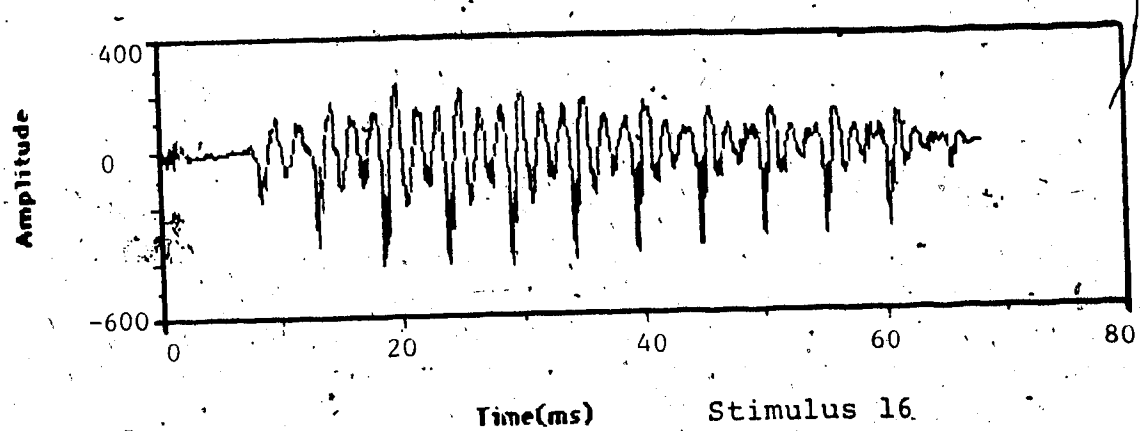


Figure 5-2 (continued)



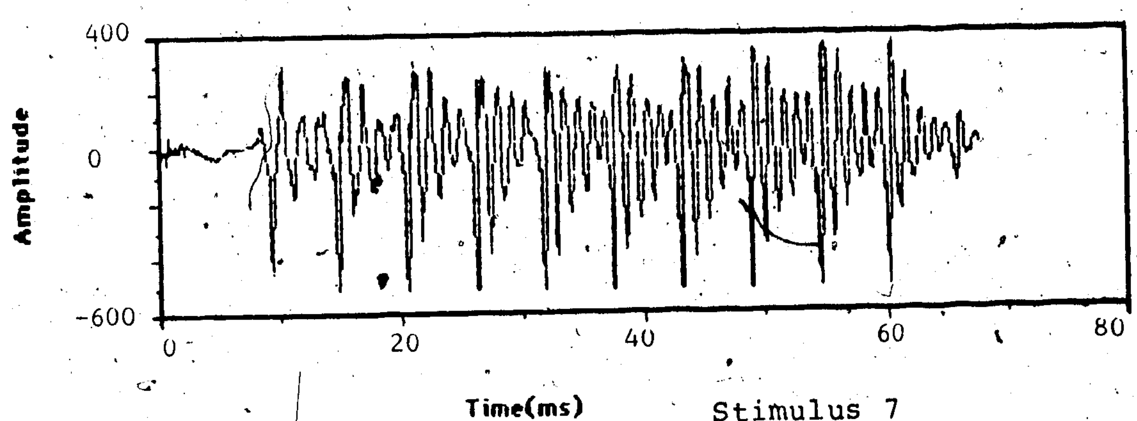


Figure 5-2 (continued)

104

BIBLIOGRAPHY

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica* **34**, 295-303.

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. (1978). Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica* **30**, 175-198.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *Journal of the Acoustical Society of America* **63**, 1556-1558.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics* **2**, 239-245.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* **54**, 421-428.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* **62**, 961-970.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* **24**, 597-606.

Darwin, C. J., & Pearson, M. (1982). What tells us when voicing has started? *Speech Communication* **1**, 29-44.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America* **74**, 1126-1135.

Delattre, P. (1951). *Principes de phonétique française a l'usage des étudiants anglo-américains*. Middlebury, Vermont: The College Store.

Delattre, P. (1965). Comparing the consonantal features of English, German, Spanish and French. *International Journal of Applied Linguistics*. **2**, 155-203.

Delattre, P. (1966). *Studies in French and Comparative Phonetics*. The Hague: Mouton.

Enstrom, D. H., & Spörri-Bütler, S. (1981). A voice onset time analysis on initial Swiss-German stops. *Folia Phoniatrica* **33**, 137-150.

Ferguson, G. A. (1981). *Statistical Analysis in Psychology and Education*, Fifth Edition, New York: McGraw-Hill.

Fisher-Jørgensen, E. (1972). PTK et BDG français en position intervocalique accentuée. *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*. P. Valdman (Ed.) The Hague: Mouton, 143-200.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* **10**, 177-192.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America* **76**, 692-707.

Flege, J. E., & Brown, Jr., W. S. (1982). The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance. *Journal of Phonetics* **10**, 335-345.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* **43**, 155-171.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* **47**, 613-617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading FØ cues in the voiced-voiceless distinction. *Journal of Phonetics* **9**, 49-62.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America* **29**, 107-116.

Han, M. S., & Weitzman, R. W. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/. *Phonetica* **22**, 112-128.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics* **1**, 263-272.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1979). *Applied Statistics for the Behavioral Sciences*. Chicago: Rand McNally.

Jaeger, J. J. (1983). The fortis-lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics* **11**, 277-290.

Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* **11**, 277-290.

Keating, P., Mikos, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America* **70**, 1261-1271.

Kim, C.-W. (1970). A theory of aspiration. *Phonetica* **21**, 107-116.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* **18**, 686-706.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* **63**, 905-917.

Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics* **1**, 73-83.

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

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

Lisker, L. (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America* **57**, 1547-1551.

Lisker, L. (1978). Rapid vs. Rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Reports on Speech Research*, **SR-54**, 127-132.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384-422.

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

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language* **47**, 767-785.

Lisker, L., Abramson, A. S., Cooper, F. S., & Schvey, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* **45**, 1543-1546.

Lubker, J. F., & Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop cognates /p/ and /b/. *Journal of the Acoustical Society of America* **47**, 625-633.

Malécot, A. (1970). The lenis-fortis opposition: its physiological parameters. *Journal of the Acoustical Society of America* **47**, 1588-1592.

Marchal, A. (1983). The fortis-lenis distinction in stops. *Speech Communication* **2**, 115-118.

Myers, J. L. (1979). *Fundamentals of Experimental Design*, Third Edition. Boston: Allyn and Bacon, Inc.

Nicholson, G. G. (1927). *A Practical Introduction to French Phonetics*. London: Macmillan.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* **75**, 224-230.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America* **61**, 1352-1361.

Pisoni, D. B., Aislin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* **8**, 297-314.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* **55**, 328-333.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* **22**, 173-189.

Santerre, L., & Suen, C. Y. (1981). Why look for a single feature to distinguish stop cognates? *Journal of Phonetics* **9**, 165-174.

Serniclaes, W., D'Alimonte, G., & Alegria, J. (1984). Production and perception of French stops by moderately deaf subjects. *Speech Communication* **3**, 185-198.

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America* **64**, 1338-1357.

Smith, B. L. (1978). Effects of place of articulation and vowel environment on 'voiced' stop consonant production. *Glossa* **12**, 163-175.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America* **55**, 653-659.

Stevenson, D., & Stephens, R. (1979). *The Alligator Reference Manual.* Unpublished document.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America* **62**, 435-446.

Valdman, A., Salazar, R. J., & Charbonneaux, M. A. (1970). *A Drillbook of French Pronunciation*, Second Edition, New York: Harper and Row.

Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception and Psychophysics* **29**, 323-335.

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

Warren, D. W. (1976). Aerodynamics of speech production. In N.J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*. New York: Academic Press.

Waters, R. S., & Wilson, Jr., W. A. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics* **26**, 285-289.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics* **26**, 95-104.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*. Second Edition. New York: McGraw-Hill.

Winitz, H., Lariviere, C. & Herriman, E. (1975). Variations in VOT for English initial stops. *Journal of Phonetics* **3**, 41-52.

Zlatin, M. A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America* **56**, 981-994.

Appendix

A. Stimulus List (English Speakers)

Say pap again, please.

bab

pab

bap

peep

beeb

peeb

beep

pep

beb

peb

bep

poop

boob

poob

boop

pop

bob

bop

B. Stimulus List (French Speakers)

Répétez pape une fois.

pabe
bape
babe
pipe
pibe
bipe
bibe
pèpe
pèbe
bèpe
bèbe
poupe
poube
boupe
boube
pape
pabe

C. Instructions -- Perception Experiments

In this experiment you will hear a number of syllables consisting of a consonant and a vowel. Your task will be to determine whether the consonant is a French /p/ or an English /b/. Each stimulus will be repeated through the headphones until you make your choice by pressing one of the switches on the switchbox.

In making your decision you should not assume that there are an equal number of /b/ and /p/ stimuli. (You may or may not find this to be true.) Nor should you assume that a particular speaker always uses his or her native language. (You may hear an English speaker producing a French /p/, for example.)

You may listen to each item as many times as necessary. You should, however, try to make your decision as quickly as possible.

Occasionally you may hear a consonant which does not seem to fit either category. In such a case, you should choose the category which the sound is closest to.

D. Windowing Programs (Experiments 3 and 4)

```
C FROM BURSTR (CURSOR 3)
      COMMON/LPCDEF/ NWIDE,NHOP,SAMRAT,LENMX,LPCORD,RNS=
C NWIDE NHOP AND SAMRAT SELF-EXPLANATORY.
C LENMX IS MAXIMUM NUMBER OF SAMPLES
C THAN CAN BE HANDLED
C LPCORD IS ORDER OF LPC THERE ARE LPCORD RC
C COEFFIEIENTS AND LPCORD+1 A COEFFICEINTS
      DATA SAMRAT/16000./
C.................................
      COMMON/SEGDOP/IALLIG,SEGFI(3),SEGNAM(2),
     $ IENBLK,IEXBLK,LENSEG
      COMMON/CONDOP/SCUR(20)
C--------------------------------------------------
C # IVFDEF 10,3 AS OF 4 SEPT 82
      COMMON/VFILES/NVFILS,IVFDEF(10,3)
      COMMON/VFDSRN/VFDSR(3,3)
      COMMON/VFLABS/ISIGO,IGLOT,ISIG1
C--------------------------------------------------
C
C--------------------------------------------------
      COMMON/RFILES/NRFILS,IRFDEF(7,3)
      COMMON/RFDSRN/RFDSR(3,3)
      COMMON/RFLABS/LPCO,LPF10,LPC1
C-------------------
C WINDOW STUFF
      INTEGER NWIN,MAXNW,MAXWL
C NUMBER OF WINDOWS <=MAXNW, MAXWINDOW LENGTH
      INTEGER ISTWIN(10)
      INTEGER LENWIN(10)
      DATA MAXNW/10/,MAXWL/256/
C RAMP KNOTS
      REAL RSAMP(10),RMUL(10)
C-------------------------
      REAL SEG1(2),SIG1(3),SIGOUT(3)
      REAL ALFILE(3),ALSGN(2)
C
      CALL TEKSET('ERASE')
      CALL LOCK(1)
C THIS WILL BE A LOOP
C
      CALL INWIN
      CALL VFDEF
C
      DO 995 I=1,20
995   SCUR(I)=0.
```

```
        WRITE(4,781)
781     FORMAT(' FULL AMP DURATION ? ',$)
        READ(4,782) TNOM
782     FORMAT(10I10)
C LENGTH OF RAMP
        RLEN=8.
        T2=TNOM
        T3=TNOM+RLEN
        NKNOTS=5
100     CONTINUE
        CALL NXTSIG(SIG1,SEG1)
C.........
        RSAMP(1)=1
        RMUL(1)=0
        RSAMP(2)=SCUR(3)-1
        IF(RSAMP(2).LT.1) RSAMP(2)=1
        RMUL(2)=0.
        RSAMP(3)=SCUR(3)
        RMUL(3)=1.0
        RSAMP(4)=T2*16+SCUR(3)
        RMUL(4)=1.0
        RSAMP(5)=T3*16+SCUR(3)
        RMUL(5)=0.0
C..........
        WRITE(4,99) SIG1,SEG1,SCUR(3)
99      FORMAT(' FILE; SEGMENT, SCUR(3):',3A6,2A6,1X,I6)
        IPRINT=0
        CALL GETSEG(SIG1,SEG1,'ALLIG',IPRINT,IERT)
        IF(IERT.NE.0.) CALL ERROR(' TRACK: SEGMENT NOT FOUND
C
C
        WRITE(4,3928) LENSEG
3928    FORMAT(' LENSEG',G15.5)
        CALL SIGXF(ISIG0)
C RESET OUTPUT LENTGH
        CALL VFLEN(ISIG1,0,1)
C
        CALL WINSUB(ISIG0,ISIG1,NWIN,ISTWIN,LENWIN)
C
C
        CALL DORAMP(ISIG0,ISIG1,NKNOTS,RSAMP,RMUL)
        CALL CHNGDV(SIG1,'D1:','W3:',SIGOUT)
        WRITE(4,101) SIG1,SIGOUT
101     FORMAT(1X,3A6,1X,3A6)
        AMP=1.
        DC=512.
C.....
        INORM=0
C
        IPRT=0
        CALL ALWRT(ISIG1,SIGOUT,SEG1,AMP,DC,INORM,IERR,IPRT)
        GOTO 100
C
        STOP
        END
```

```
      SUBROUTINE DORAMP(IN,OUT,NKNOTS,RSAMP,RMUL)
C MODIFIED 12 JUN 84 TO OMIT LEADING ZEROS
      INTEGER IN,OUT,NKNOTS
      REAL RSAMP(NKNOTS),RMUL(NKNOTS)
      REAL BUFI(257),BUFO(257)
      INTEGER NBUF
      DATA NBUF/257/
      CALL SETBF(IN,BUFI,NBUF)
      CALL SETBF(OUT,BUFO,NBUF)
      IF(NKNOTS.GE.2) GOTO 100
      WRITE(4,1)
1     FORMAT(' DORAMP: NEEDS AT LEAST 2 KNOTS')
      RETURN
C
100   CONTINUE
C SKIP OUTPUT
      NSKIPO=0
      IF(RMUL(1).EQ.0..AND.RMUL(2).EQ.0.) NSKIPO=RSAMP(2)
      IF(NSKIPO.LT.0) NSKIPO=0
C     WRITE(4,99) NSKIPO
99    FORMAT(' NSKIPO',G15.5)
      DO 200 I=1,NKNOTS-1
      IST=RSAMP(I)
      LAST=RSAMP(I+1)
      A1=RMUL(I)
      A2=RMUL(I+1)
      NPTS=LAST-IST+1
      ISTI=IST
      ISTO=IST-NSKIPO
C     ISTO=ISTI
C
C     WRITE(4,2) NPTS,ISTI,ISTO,A1,A2
2     FORMAT(' NPTS,ISTI,ISTO,A1,A2',3I7,2G15.5)
      CALL RAMP(NPTS,IN,ISTI,BUFI,OUT,ISTO,BUFO,A1,A2)
200   CONTINUE
      RETURN
      END
```

```
C  H. NRAMP 28-FEB 84
C  MOD 12 JUN 84 FOR M.GUNKO
C  ALLOWS FOR NEG OUTPUT FILE INDICES
C  (SO LEADING ZERO SIGNAL CAN BE TRUNCED)
         SUBROUTINE RAMP(NPTS,IN,ISTI,BUFI,OUT,ISTO,BUFO,
             AMP1, AMP2)
         INTEGER NPTS,IN,ISTI,OUT,ISTO
         REAL BUFI(1),BUFO(1)
         REAL AMP1,AMP2
C
         COMMON/VFILES/NVFILE,IVFDEF(10,5)
C  BE SURE TO CALL SETBUF BEFORE USING
C
C  USEFUL FUNCTIONS
         LOGICAL OUTV,OUTVHI
         INTEGER IBF
         OUTVHI(IVUT,IT)=IT.GT.IVFDEF(9,IVUT)
         IBF(IVUTT,ITT)=ITT-IVFDEF(5,IVUTT)
C......
C  GET THE INITIAL RECORDS
         IF(NPTS.LE.0) RETURN
         CALL GETVR(IN,ISTI,BUFI)
         CALL GETVR(OUT,ISTO,BUFO)
C  AMP DEL
         ADEL=(AMP2-AMP1)/FLOAT(NPTS-1)
         AMP=AMP1
         ITHI=ISTI
         ITHO=ISTO
         DO 10 I=1,NPTS
C  GETP(IN,ITHI,VAL,BUFI)
         IF(OUTVHI(IN,ITHI)) CALL GETVR(IN,ITHI,BUFI)
         VAL=BUFI(IBF(IN,ITHI))
         VAL=AMP*VAL
C  OADP(OUT,ITHO,VAL,BUFO)
C  IGNORE OUTPUT BEFORE BEGINNING OF FILE MOD 12JUN84
         IF(ITHO.LE.0) GOTO 109
         IF(OUTVHI(OUT,ITHO)) CALL GETVR(OUT,ITHO,BUFO)
         BUFO(IBF(OUT,ITHO))=VAL+BUFO(IBF(OUT,ITHO))
109      CONTINUE
         ITHI=ITHI+1
         ITHO=ITHO+1
         AMP=AMP+ADEL
10       CONTINUE
         MAXOUT=ISTO+NPTS-1
         IF(MAXOUT.GT.IVFDEF(4,OUT)) IVFDEF(4,OUT)=MAXOUT
         CALL PUTVR(OUT,BUFO)
         RETURN
         END
```

Table 4-1: Aspirated French /p/ in Experiment 2

Vowel	Proportion of aspirated /p/s	Mean duration of aspiration when present (msec.)
/a/	3/6 (50%)	8.4
/i/	6/6 (100%)	36.0
/ɛ/	2/6 (33%)	8.2
/u/	6/6 (100%)	28.7
Total	17/24 (71%)	25.3

Table 4-2: Mean amplitude of aspiration noise by subject

Subject	Aspirated tokens	Amp. of asp.
S1	4/8	0.33
S2	7/8	0.14
S3	6/8	0.60

Table 4-3: Mean amplitudes of segments

Consonant	n	V. Bar	Burst	Asp.
/p/ (asp.)	17	0	0.43	0.35
/p/ (unasp.)	7	0	0.10	0
/p/ (total)	24	0	0.34	0.25
/b/	24	0.15	0.40	0.12

Table 4-4 (a): Mean durations of segments for S2 (in msec.)

	/p/			/b/	
	s.per	burst	asp.(n)	v. bar	burst
Exp. 1	89.9	9.5	33.1(3)	116.8	4.7
Exp. 2	96.4	7.1	23.3(7)	110.9	4.4

Table 4-4 (b): Mean amplitudes of segments for S2

	/p/		/b/	
	burst	asp.(n)	v. bar	burst
Exp. 1	0.13	0.08(3)	0.29	0.33
Exp. 2	0.27	0.14(7)	0.22	0.38

Table 4-5: Mean relative amplitudes for both samples

	burst	asp.	v. bar
Exp. 1	0.2067	0.0692	0.3167
Exp. 2	0.3676	0.2450	0.1505

Table 5-1: Correct identification scores (%) by subject

Subject	English	French	Total
S1	69.1	74.4	71.7
S2	58.1	58.1	58.1
S3	79.4	82.5	80.9
S4	57.5	50.6	54.1
S5	58.8	85.6	72.2
S6	68.8	83.4	76.1
S7	65.9	78.8	72.3
S8	61.7	72.7	67.2
Total	65.0	73.3	69.1

Table 5-2: Means and ranges of selected predictors

Predictor	Mean /p/	Mean /b/	Range /p/	Range /b/
VOT (msec.)	12.00	6.90	7.10-25.50	0-17.50
d75 (msec.)	24.14	12.94	16.36-34.61	9.95-17.65
t25 (msec.)	11.93	4.89	4.13-19.30	2.56-10.52
asp. (msec.)	3.23	1.73	0-14.90	0-12.00
na8*	0.2418	0.5338	0.0577-0.5357	0.0942-0.7291

*expressed as a proportion of mean peak

Table 5-3: Correlations among selected predictors and /p/-id scores

	asp. dur.	t25	d75	na8	VOT
t25	0.444				
d75	0.210	0.645			
na8	-0.492	-0.885	-0.605		
VOT	0.737	0.675	0.443	-0.761	
/p/-id	0.320	0.634	0.649	-0.713	0.767

Table 5-4: Characteristics of selected stimulus items

Stim.	VOT	Nat. (msec.)	/p/-id Cat.	d75 (%)	t25 (msec.)	na8* (msec.)
16	6.1	/b/	22	12.15	3.97	0.6622
7	6.1	/b/	46	17.65	5.27	0.5239
15	7.0	/b/	17	10.12	4.92	0.4187
24	7.1	/p/	58	16.36	5.53	0.3703
32	8.0	/p/	45	25.65	15.19	0.2734
29	8.0	/p/	83	34.61	4.72	0.5357
23	8.4	/p/	46	17.16	4.53	0.4589
30	8.4	/p/	84	24.66	4.13	0.5268
25	13.0	/p/	44	17.49	14.44	0.1916
18	12.8	/p/	82	23.00	12.62	0.1832

*expressed as a proportion of mean peak

Table 5-5: Regression equations

Equation	R2 (%)
1. 0.02 + 0.0695 VOT - 0.0377 asp. dur + 0.0163 d75	77.6
2. 0.162 + 0.084 VOT - 0.0445 asp. dur.	69.9
3. 0.074 + 0.432 VOT + 0.0197 d75	68.7
4. 0.076 + 0.0478 VOT + 0.0128 d90	66.0
5. 0.221 + 0.432 VOT + 0.0186 d50	63.3
6. 0.140 + 0.494 VOT + 0.00939 t90	61.3
7. 0.322 + 0.0555 VOT	57.4
8. 0.866 - 0.884 na8 + 0.0176 d75	55.4

Table 6-1: Correct identifications by listener (%)

Subject	English /b/	French /p/	Total	/p/-id
Naive				
S1	85.4	51.0	68.2	33
S2	92.7	58.3	75.5	33
S3	72.4	59.4	65.9	43
S4	22.9	71.9	47.4	74
S5	33.9	65.6	49.7	66
Total	61.5	61.2	61.4	50
Experienced				
S6	60.9	81.3	71.1	60
S7	63.0	76.6	69.8	57
S8	59.4	75.0	67.2	58
S9	58.3	74.0	66.1	58
S10	77.1	68.2	72.7	46
Total	63.7	75.0	69.4	56
Total (overall)	62.6	68.1	65.4	53

Table 6-2: Comparison of results (Exp. 3 and 4)

Equation	*s* in Exp.3	RMS error in Exp. 4
VOT/asp. dur./ d75	0.1731	0.2779
VOT/asp. dur.	0.2005	0.2962
VOT/d75	0.2047	0.2463
VOT/d90	0.2132	0.2534
VOT/d50	0.2214	0.2425
VOT/t90	0.2275	0.2453
VOT	0.2386	0.2545
na8/d75	0.2444	0.8398

Table 6-3: Correlations among selected predictors and /p/-id scores

	asp. dur.	t10	d50	d75	VOT
t10	0.574				
d50	0.319	0.689			
d75	0.285	0.792	0.660		
VOT	0.707	0.587	0.529	0.228	
/p/-id	0.345	0.773	0.567	0.464	0.793

Figures

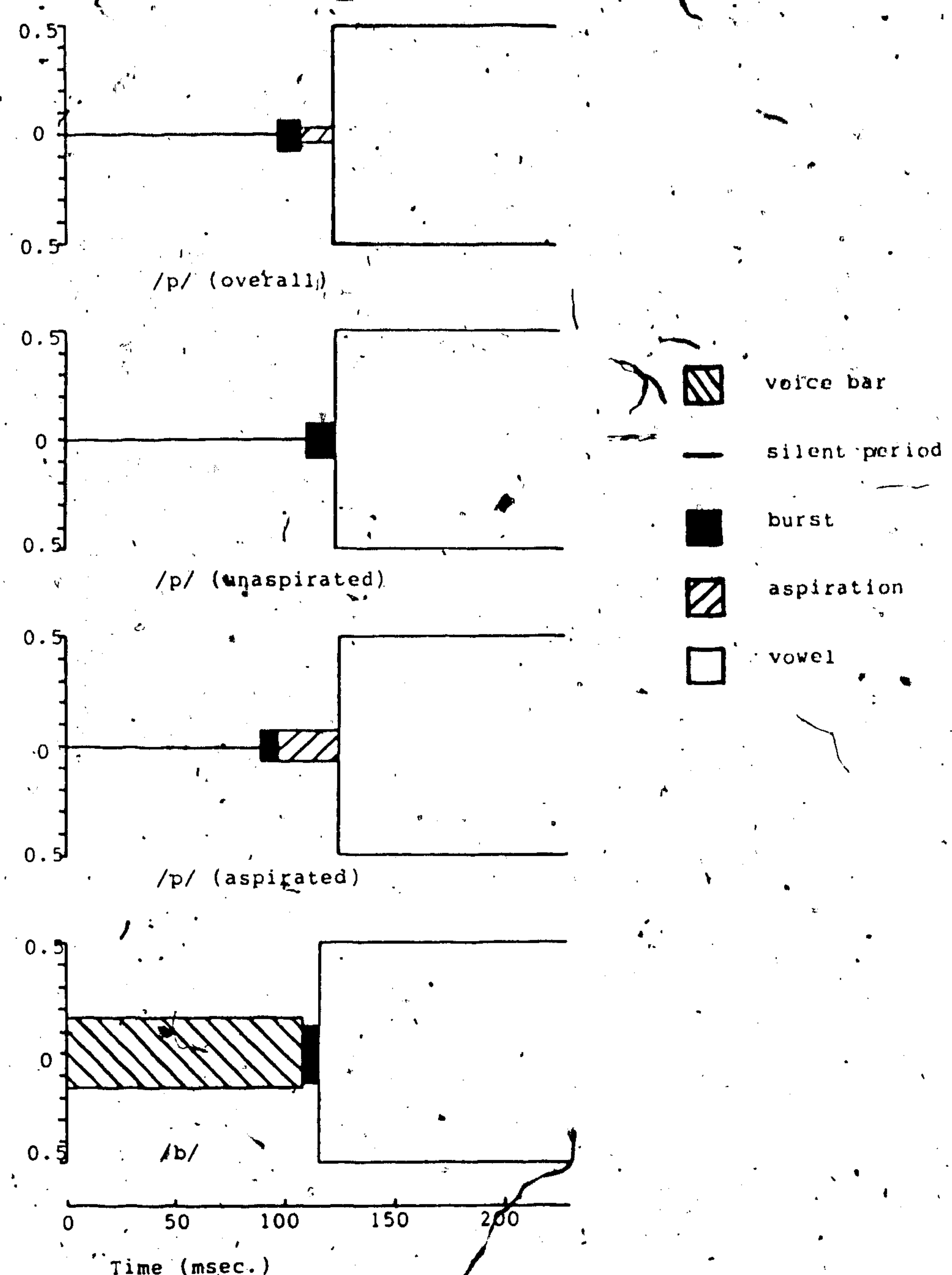


Figure 3-1: Schematic representation of average cases (French)

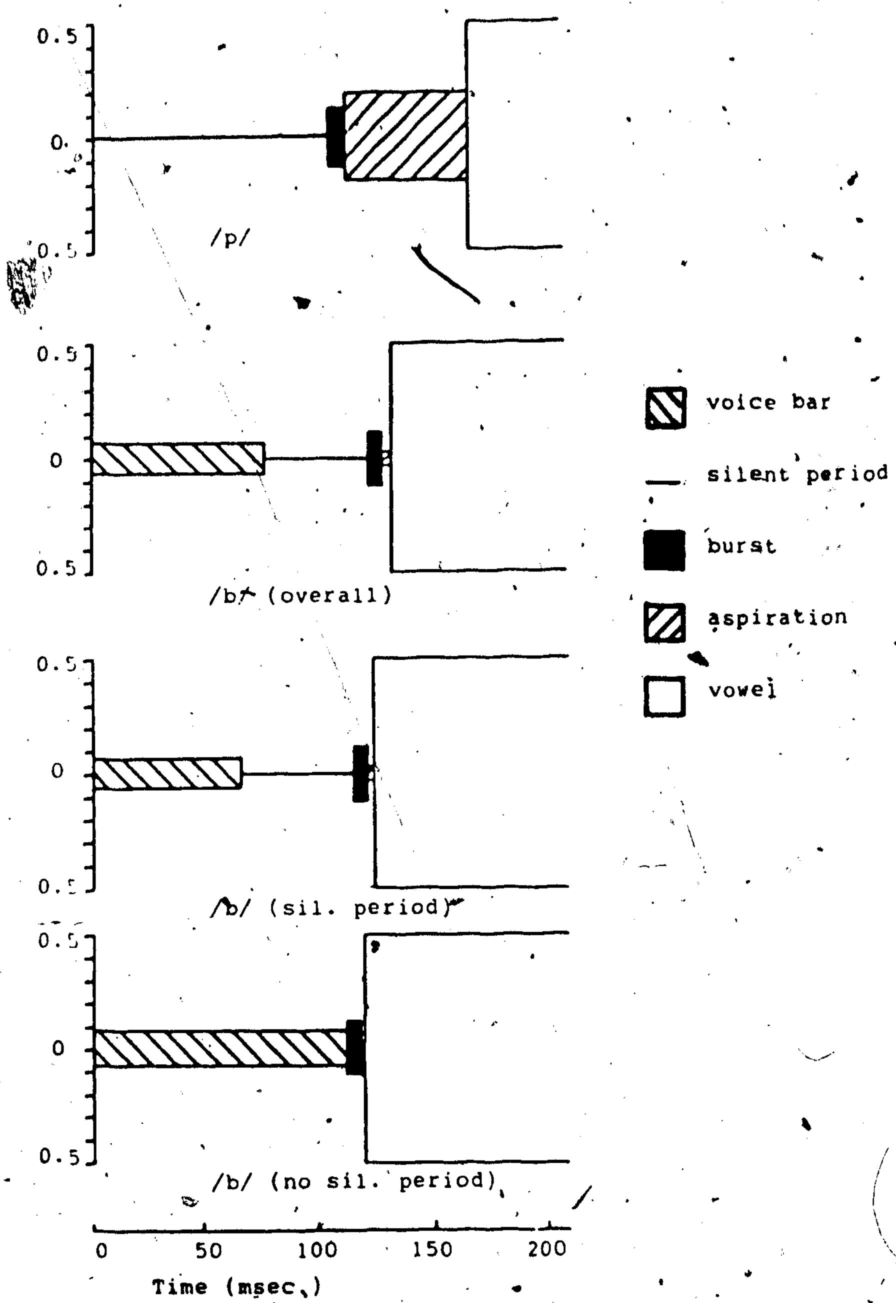


Figure 3-2: Schematic representation of average cases (English)

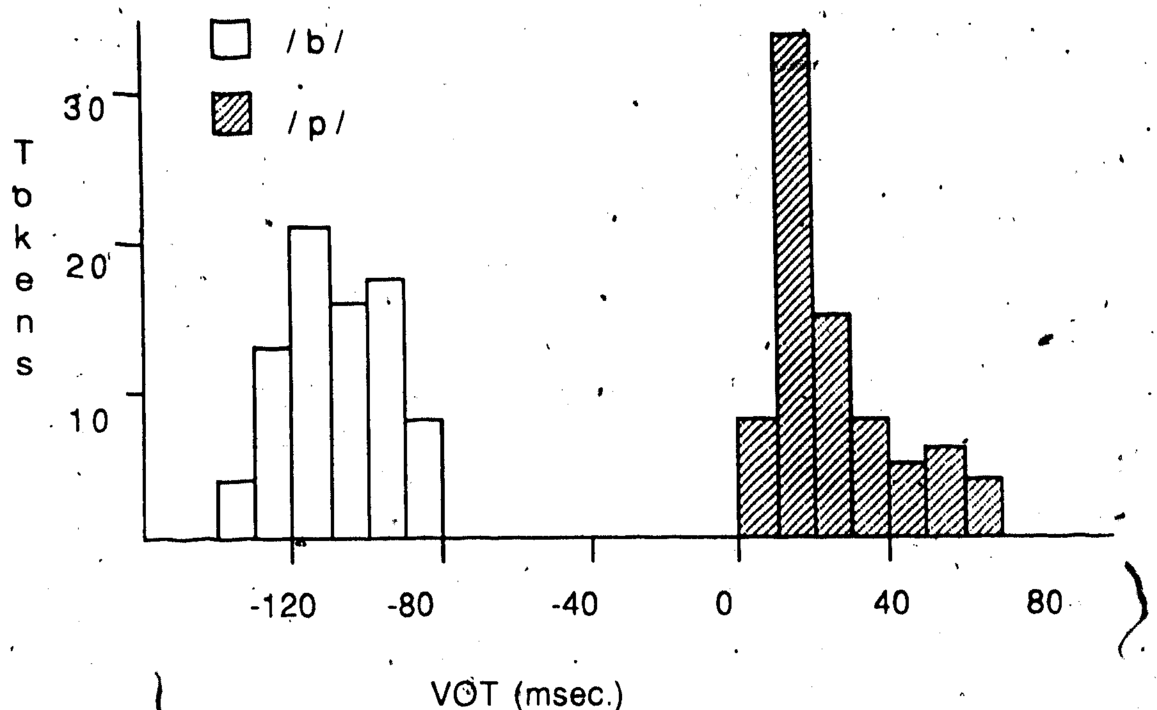


Fig. 3-3: Voice onset time in French (Experiment 1)

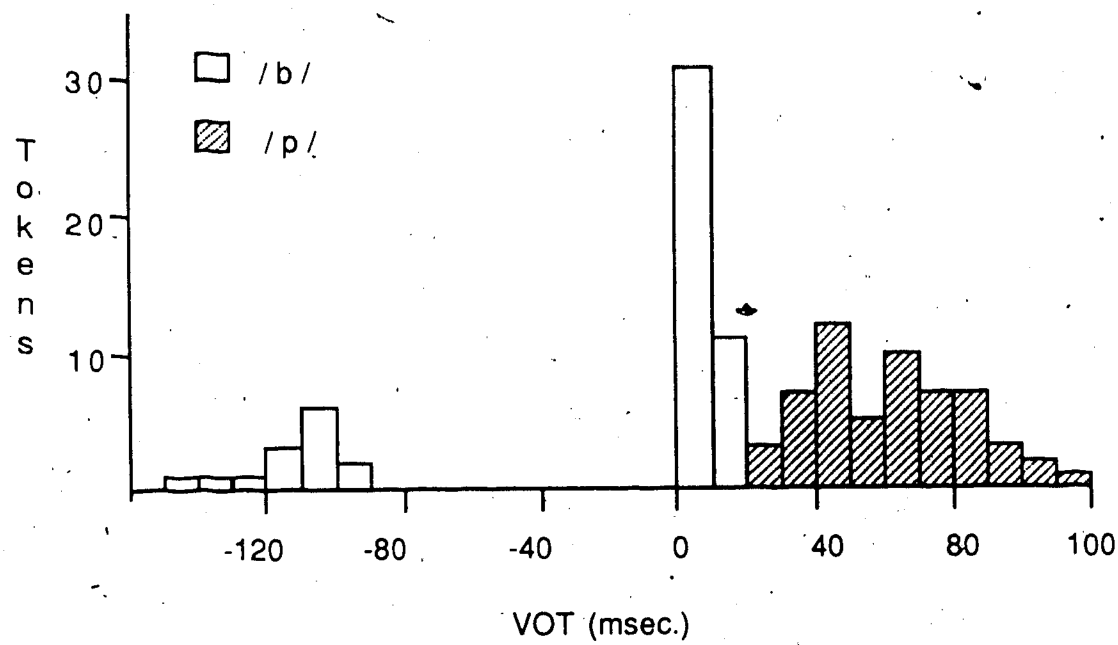


Fig. 3-4: Voice onset time in English (Experiment 1)

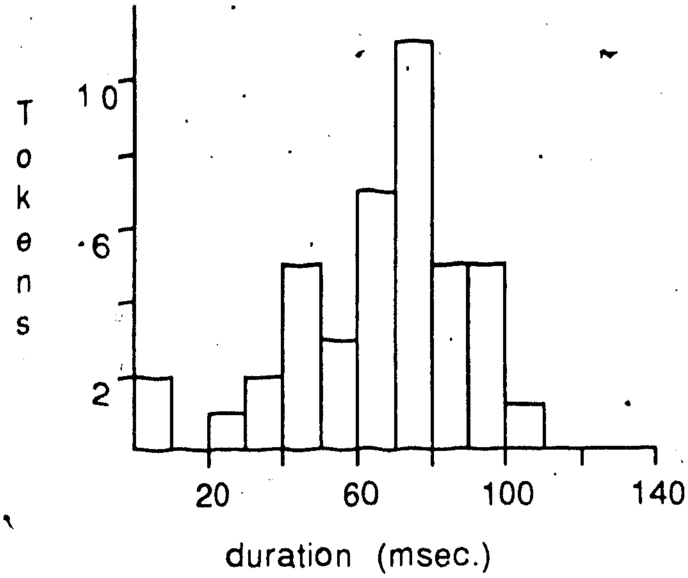


Fig. 3-5: Durations of "interrupted" voice bars in English /b/

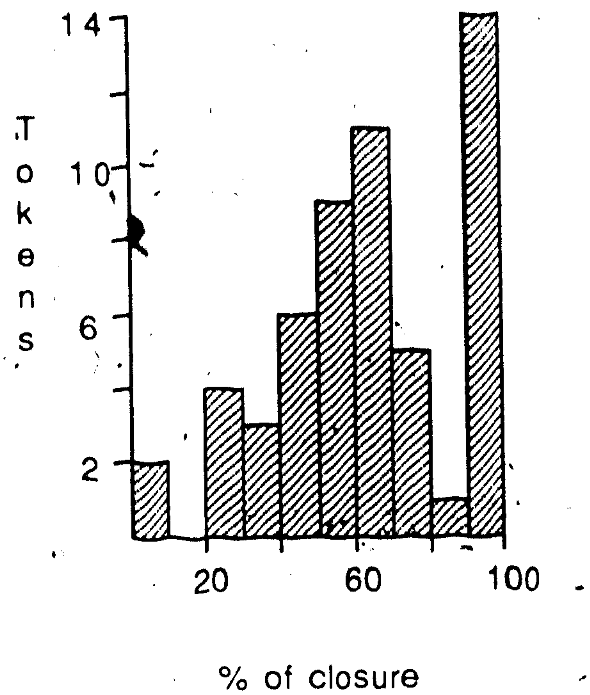


Fig. 3-6: % of closure interval which is voiced in English /b/

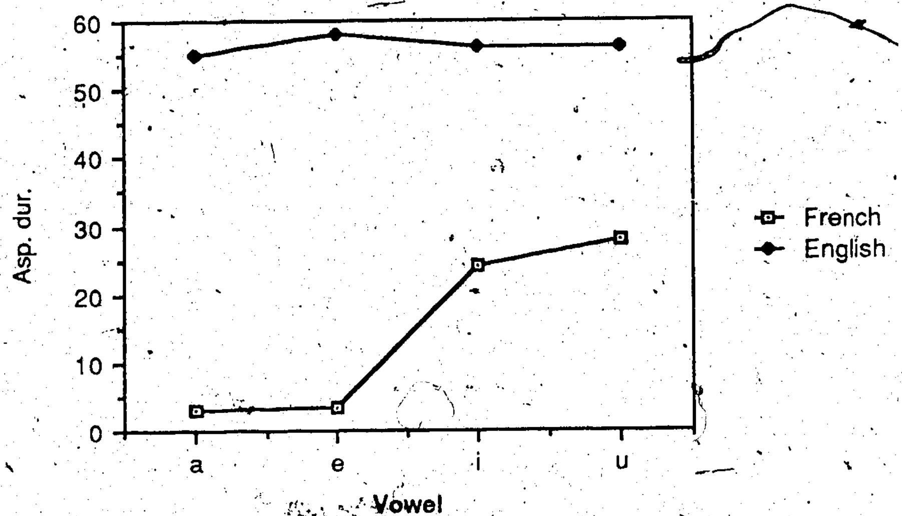


Fig. 3-7: Aspiration duration (msec.) — VL interaction

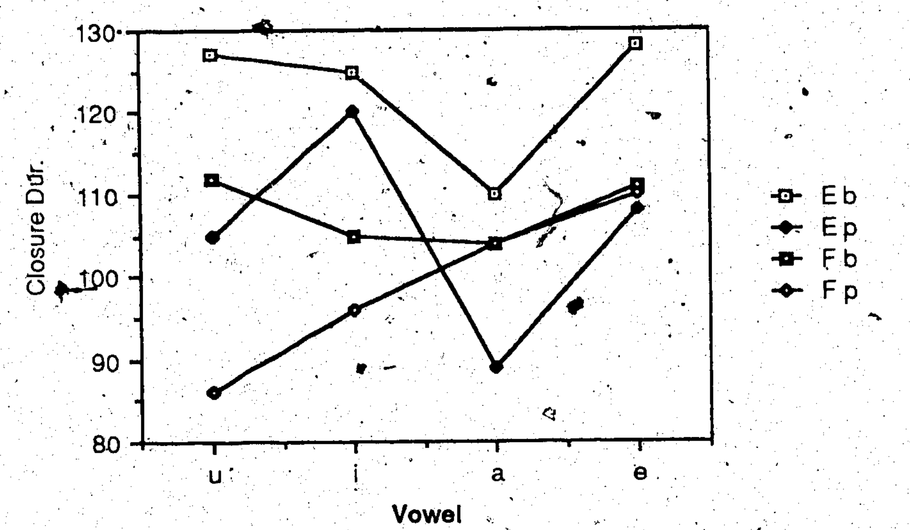


Fig. 3-8: Closure durations (msec.) — VCL interaction

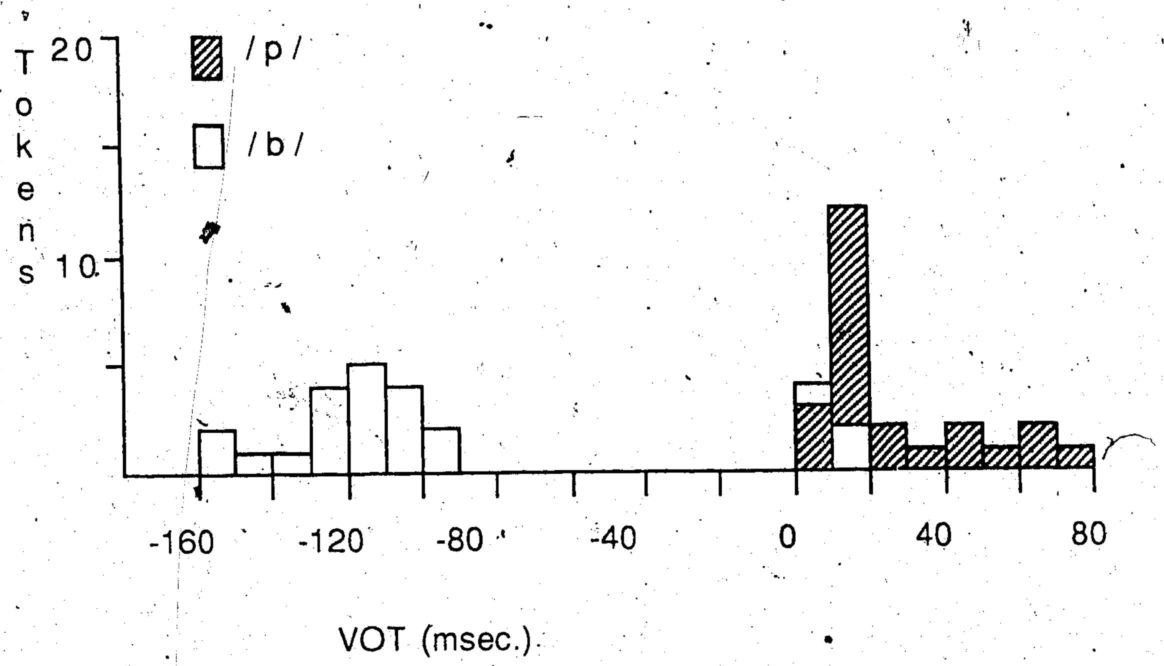


Fig 4-1: Voice onset time in French (Experiment 2)

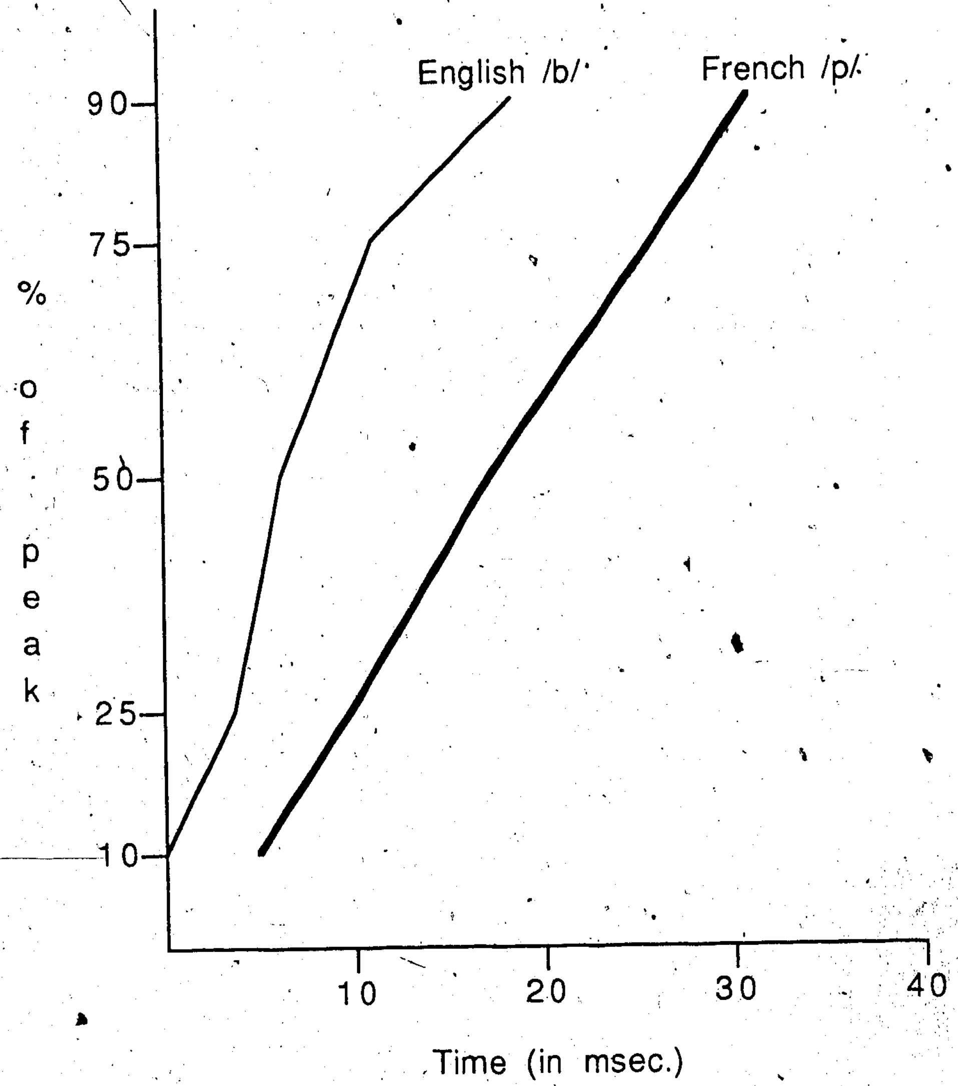
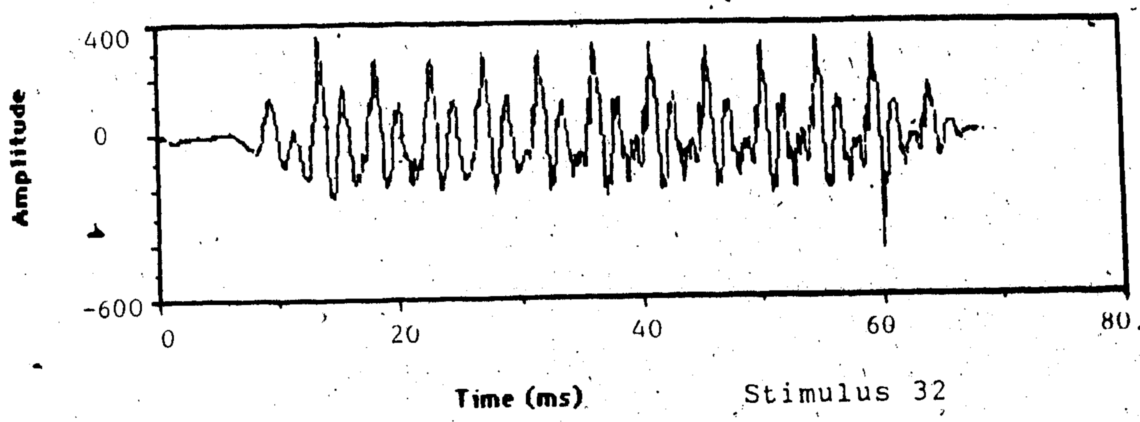


Figure 5-1: *d* values



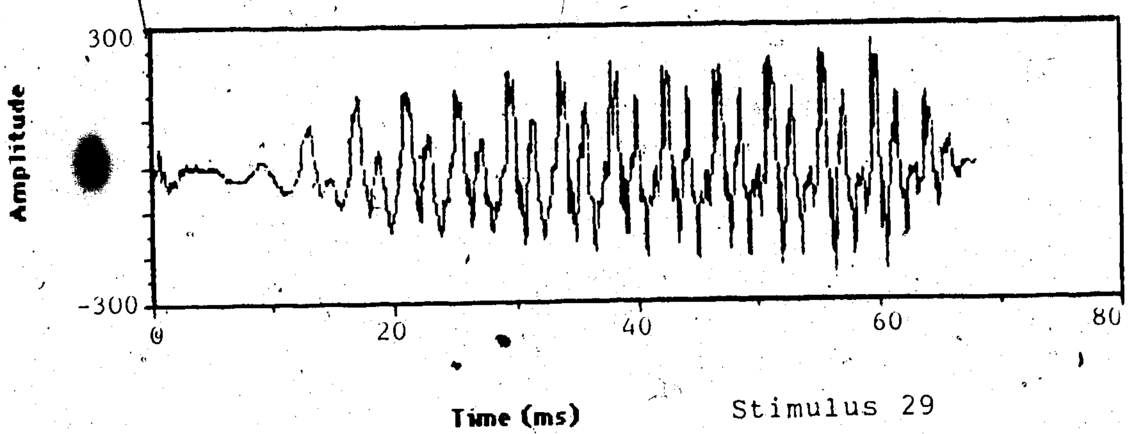
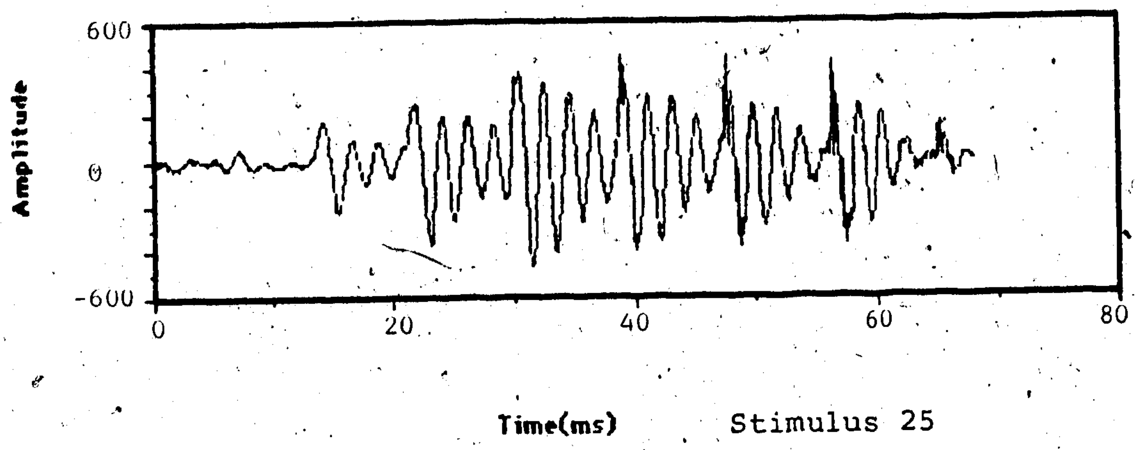


Figure 5-2: Sample waveforms



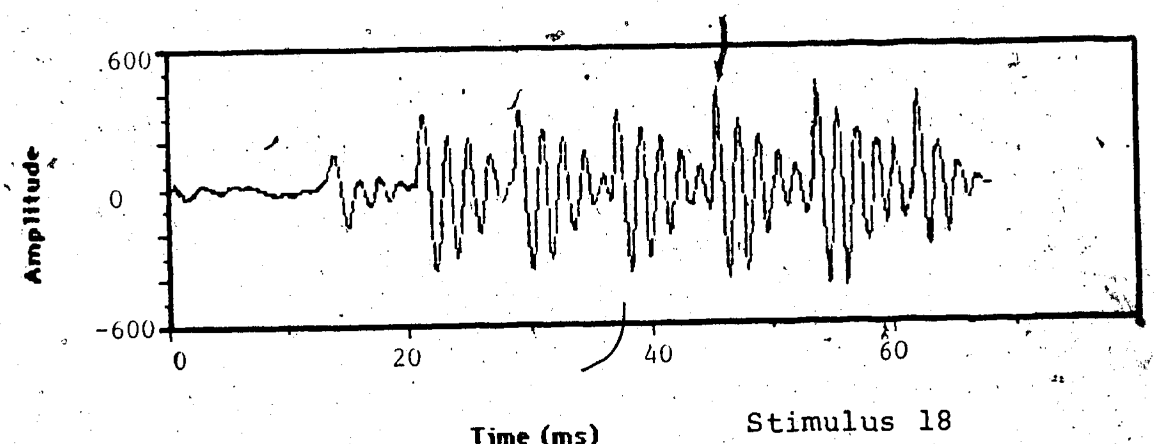
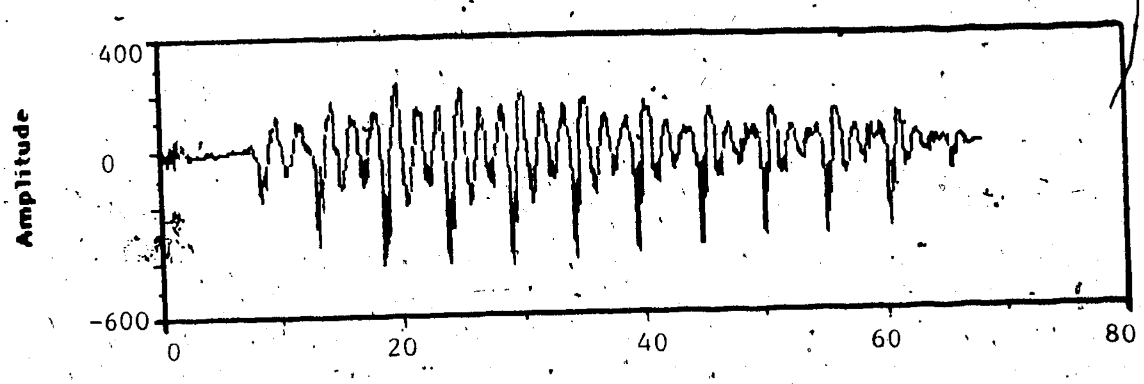
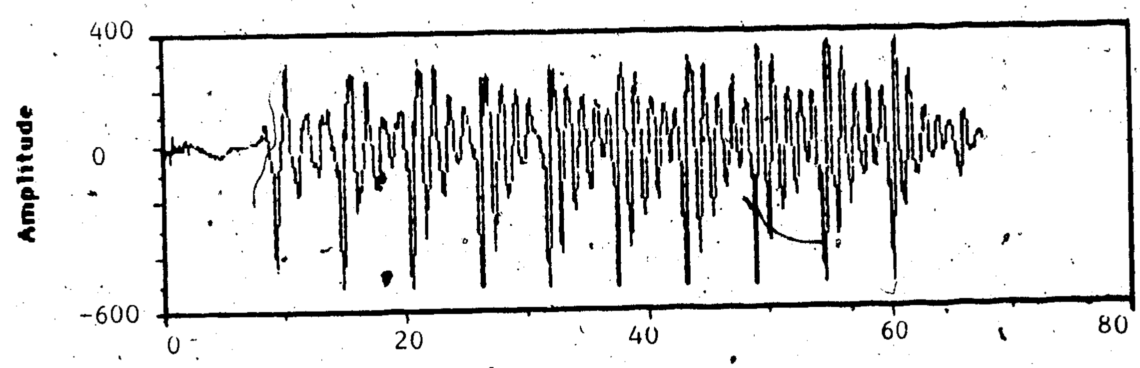


Figure 5-2 (continued)



Stimulus 16



Stimulus 7

Figure 5-2 (continued)

BIBLIOGRAPHY

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica* **34**, 295-303.

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. (1978). Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica* **30**, 175-198.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *Journal of the Acoustical Society of America* **63**, 1556-1558.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics* **2**, 239-245.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* **54**, 421-428.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* **62**, 961-970.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* **24**, 597-606.

Darwin, C. J., & Pearson, M. (1982). What tells us when voicing has started? *Speech Communication* **1**, 29-44.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America* **74**, 1126-1135.

Delattre, P. (1951). *Principes de phonétique française a l'usage des étudiants anglo-américains*. Middlebury, Vermont: The College Store.

Delattre, P. (1965). Comparing the consonantal features of English, German, Spanish and French. *International Journal of Applied Linguistics*. **2**, 155-203.

Delattre, P. (1966). *Studies in French and Comparative Phonetics*. The Hague: Mouton.

Enstrom, D. H., & Spörri-Bütler, S. (1981). A voice onset time analysis on initial Swiss-German stops. *Folia Phoniatrica* **33**, 137-150.

Ferguson, G. A. (1981). *Statistical Analysis in Psychology and Education*, Fifth Edition, New York: McGraw-Hill.

Fisher-Jørgensen, E. (1972). PTK et BDG français en position intervocalique accentuée. *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*. P. Valdman (Ed.) The Hague: Mouton, 143-200.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics* **10**, 177-192.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America* **76**, 692-707.

Flege, J. E., & Brown, Jr., W. S. (1982). The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance. *Journal of Phonetics* **10**, 335-345.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* **43**, 155-171.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* **47**, 613-617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading FØ cues in the voiced-voiceless distinction. *Journal of Phonetics* **9**, 49-62.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America* **29**, 107-116.

Han, M. S., & Weitzman, R. W. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/, and /p^h, t^h, k^h/. *Phonetica* **22**, 112-128.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics* **1**, 263-272.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1979). *Applied Statistics for the Behavioral Sciences*. Chicago: Rand McNally.

Jaeger, J. J. (1983). The fortis-lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics* **11**, 277-290.

Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* **11**, 277-290.

Keating, P., Mikos, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America* **70**, 1261-1271.

Kim, C.-W. (1970). A theory of aspiration. *Phonetica* **21**, 107-116.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* **18**, 686-706.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* **63**, 905-917.

Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics* **1**, 73-83.

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

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

Lisker, L. (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America* **57**, 1547-1551.

Lisker, L. (1978). Rapid vs. Rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Reports on Speech Research*, **SR-54**, 127-132.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384-422.

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

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language* **47**, 767-785.

Lisker, L., Abramson, A. S., Cooper, F. S., & Schvey, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* **45**, 1543-1546.

Lubker, J. F., & Parris, P. J. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop cognates /p/ and /b/. *Journal of the Acoustical Society of America* **47**, 625-633.

Malécot, A. (1970). The lenis-fortis opposition: its physiological parameters. *Journal of the Acoustical Society of America* **47**, 1588-1592.

Marchal, A. (1983). The fortis-lenis distinction in stops. *Speech Communication* **2**, 115-118.

Myers, J. L. (1979). *Fundamentals of Experimental Design*, Third Edition. Boston: Allyn and Bacon, Inc.

Nicholson, G. G. (1927). *A Practical Introduction to French Phonetics*. London: Macmillan.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* **75**, 224-230.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America* **61**, 1352-1361.

Pisoni, D. B., Aislin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* **8**, 297-314.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* **55**, 328-333.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* **22**, 173-189.

Santerre, L., & Suen, C. Y. (1981). Why look for a single feature to distinguish stop cognates? *Journal of Phonetics* **9**, 165-174.

Serniclaes, W., D'Alimonte, G., & Alegria, J. (1984). Production and perception of French stops by moderately deaf subjects. *Speech Communication* **3**, 185-198.

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America* **64**, 1338-1357.

Smith, B. L. (1978). Effects of place of articulation and vowel environment on 'voiced' stop consonant production. *Glossa* **12**, 163-175.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America* **55**, 653-659.

Stevenson, D., & Stephens, R. (1979). *The Alligator Reference Manual.* Unpublished document.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America* **62**, 435-446.

Valdman, A., Salazar, R. J., & Charbonneaux, M. A. (1970). *A Drillbook of French Pronunciation*, Second Edition, New York: Harper and Row.

Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception and Psychophysics* **29**, 323-335.

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

Warren, D. W. (1976). Aerodynamics of speech production. In N.J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*. New York: Academic Press.

Waters, R. S., & Wilson, Jr., W. A. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics* **26**, 285-289.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics* **26**, 95-104.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*. Second Edition. New York: McGraw-Hill.

Winitz, H., Lariviere, C. & Herriman, E. (1975). Variations in VOT for English initial stops. *Journal of Phonetics* **3**, 41-52.

Zlatin, M. A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America* **56**, 981-994.

Appendix

A. Stimulus List (English Speakers)

Say pap again, please.

bab

pab

bap

peep

beeb

peeb

beep

pep

beb

peb

bep

poop

boob

poob

boop

pop

bob

bop

B. Stimulus List (French Speakers)

Répétez pape une fois.

pabe
bape
babe
pipe
pibe
bipe
bibe
pèpe
pèbe
bèpe
bèbe
poupe
poube
boupe
boube
pape
pabe

C. Instructions -- Perception Experiments

In this experiment you will hear a number of syllables consisting of a consonant and a vowel. Your task will be to determine whether the consonant is a French /p/ or an English /b/. Each stimulus will be repeated through the headphones until you make your choice by pressing one of the switches on the switchbox.

In making your decision you should not assume that there are an equal number of /b/ and /p/ stimuli. (You may or may not find this to be true.) Nor should you assume that a particular speaker always uses his or her native language. (You may hear an English speaker producing a French /p/, for example.)

You may listen to each item as many times as necessary. You should, however, try to make your decision as quickly as possible.

Occasionally you may hear a consonant which does not seem to fit either category. In such a case, you should choose the category which the sound is closest to.

D. Windowing Programs (Experiments 3 and 4)

```
C FROM BURSTR (CURSOR 3)
      COMMON/LPCDEF/ NWIDE,NHOP,SAMRAT,LENMX,LPCORD,DNS
C NWIDE NHOP AND SAMRAT SELF-EXPLANATORY.
C LENMX IS MAXIMUM NUMBER OF SAMPLES
C THAN CAN BE HANDLED
C LPCORD IS ORDER OF LPC THERE ARE LPCORD RC
C COEFFIEIENTS AND LPCORD+1 A COEFFICIENTS
      DATA SAMRAT/16000./
C...............................
      COMMON/SEGDOP/IALLIG,SEGFI(3),SEGNAM(2),
     $ IENBLK,IEXBLK,LENSEG
      COMMON/CONDOP/SCUR(20)
C-----------------------------------------------------------
C # IVFDEF 10,3 AS OF 4 SEPT 82
      COMMON/VFILES/NVFILS,IVFDEF(10,3)
      COMMON/VFDSRN/VFDSR(3,3)
      COMMON/VFLABS/ISIG0,IGLOT,ISIG3
C-----------------------------------------------------------
C
C-----------------------------------------------------------
      COMMON/RFILES/NRFILS,IRFDEF(7,3)
      COMMON/RFDSRN/RFDSR(3,3)
      COMMON/RFLABS/LPC0,LFF10,LPC1
C-----------------
C WINDOW STUFF
      INTEGER NWIN,MAXNW,MAXWL
C NUMBER OF WINDOWS <=MAXNW, MAXWINDOW LENGTH
      INTEGER ISTWIN(10)
      INTEGER LENWIN(10)
      DATA MAXNW/10/,MAXWL/256/
C RAMP KNOTS
      REAL RSAMP(10),RMUL(10)
C-----------------
      REAL SEG1(2),SIG1(3),SIGOUT(3)
      REAL ALFILE(3),ALSGN(2)
C
      CALL TERSET('ERASE')
      CALL LOCK(1)
C THIS WILL BE A LOOP
C
      CALL INWIN
      CALL VFDEF
C
      DO 995 I=1,20
995   SCUR(I)=0.
```

```
      WRITE(4,781)
781   FORMAT(' FULL AMP DURATION ? ',$)
      READ(4,782) TNOM
782   FORMAT(10I10)
C LENGTH OF RAMP
      RLEN=8.
      T2=TNOM
      T3=TNOM+RLEN
      NKNOTS=5
100   CONTINUE
      CALL NXTSIG(SIG1,SEG1)
C.........
      RSAMP(1)=1
      RMUL(1)=0
      RSAMP(2)=SCUR(3)-1
      IF(RSAMP(2).LT.1) RSAMP(2)=1
      RMUL(2)=0.
      RSAMP(3)=SCUR(3)
      RMUL(3)=1.0
      RSAMP(4)=T2*16+SCUR(3)
      RMUL(4)=1.0
      RSAMP(5)=T3*16+SCUR(3)
      RMUL(5)=0.0
C..........
      WRITE(4,99) SIG1,SEG1,SCUR(3)
99    FORMAT(' FILE; SEGMENT, SCUR(3):',3A6,2A6,1X,I6)
      IPRINT=0
      CALL GETSEG(SIG1,SEG1,'ALLIG',IPRINT,IERT)
      IF(IERT.NE.0.) CALL ERROR(' TRACK: SEGMENT NOT FOUND
C
C
      WRITE(4,3928) LENSEG
3928  FORMAT(' LENSEG',G15.5)
      CALL SIGXF(ISIG0)
C RESET OUTPUT LENTGH
      CALL VFLEN(ISIG1,0,1)
C
      CALL WINSUB(ISIG0,ISIG1,NWIN,ISTWIN,LENWIN)
C
C
      CALL DORAMP(ISIG0,ISIG1,NKNOTS,RSAMP,RMUL)
      CALL CHNGDV(SIG1,'D1:','D3:',SIGOUT)
      WRITE(4,101) SIG1,SIGOUT
101   FORMAT(1X,3A6,1X,3A6)
      AMP=1.
      DC=512.
C.....
      INORM=0
C
      IPRT=0
      CALL ALWRT(ISIG1,SIGOUT,SEG1,AMP,DC,INORM,IERR,IPRT)
      GOTO 100
C
      STOP
      END
```

```
      SUBROUTINE DORAMP(IN,OUT,NKNOTS,RSAMP,RMUL)
C MODIFIED 12 JUN 84 TO OMIT LEADING ZEROS
      INTEGER IN,OUT,NKNOTS
      REAL RSAMP(NKNOTS),RMUL(NKNOTS)
      REAL BUFI(257),BUFO(257)
      INTEGER NBUF
      DATA NBUF/257/
      CALL SETBF(IN,BUFI,NBUF)
      CALL SETBF(OUT,BUFO,NBUF)
      IF(NKNOTS.GE.2) GOTO 100
      WRITE(4,1)
1     FORMAT(' DORAMP: NEEDS AT LEAST 2 KNOTS')
      RETURN
C
100   CONTINUE
C SKIP OUTPUT
      NSKIPO=0
      IF(RMUL(1).EQ.0..AND.RMUL(2).EQ.0.) NSKIPO=RSAMP(2)
      IF(NSKIPO.LT.0) NSKIPO=0
C     WRITE(4,99) NSKIPO
99    FORMAT(' NSKIPO',G15.5)
      DO 200 I=1,NKNOTS-1
      IST=RSAMP(I)
      LAST=RSAMP(I+1)
      A1=RMUL(I)
      A2=RMUL(I+1)
      NPTS=LAST-IST+1
      ISTI=IST
      ISTO=IST-NSKIPO
C     ISTO=ISTI
C
C     WRITE(4,2) NPTS,ISTI,ISTO,A1,A2
2     FORMAT(' NPTS,ISTI,ISTO,A1,A2',3I7,2G15.5)
      CALL RAMP(NPTS,IN,ISTI,BUFI,OUT,ISTO,BUFO,A1,A2)
200   CONTINUE
      RETURN
      END
```

```
C H. NFAME 26-FEB 84
C MOD 12 JUN 84 FOR M.BUNKO
C ALLOWS FOR NEG OUTPUT FILE INDICES
C (SO LEADING ZERO SIGNAL CAN BE TRUNCED)
      SUBROUTINE RAMP(NPTS,IN,ISTI,BUFI,OUT,ISTO,BUFO,
         AMP1, AMP2)
      INTEGER NPTS,IN,ISTI,OUT,ISTO
      REAL BUFI(1),BUFO(1)
      REAL AMP1,AMP2
C
      COMMON/VFILES/NVFILE,IVFDEF(10,5)
C BE SURE TO CALL SETBUF BEFORE USING
C
C USEFUL FUNCTIONS
      LOGICAL OUTV,OUTVHI
      INTEGER IBF
      OUTVHI(IVUT,IT)=IT.GT.IVFDEF(9,IVUT)
      IBF(IVUTT,ITT)=ITT-IVFDEF(5,IVUTT)
C......
C GET THE INITIAL RECORDS
      IF(NPTS.LE.0) RETURN
      CALL GETVR(IN,ISTI,BUFI)
      CALL GETVR(OUT,ISTO,BUFO)
C AMP DEL
      ADEL=(AMP2-AMP1)/FLOAT(NPTS-1)
      AMP=AMP1
      ITHI=ISTI
      ITHO=ISTO
      DO 10 I=1,NPTS
C GETP(IN,ITHI,VAL,BUFI)
      IF(OUTVHI(IN,ITHI)) CALL GETVR(IN,ITHI,BUFI)
      VAL=BUFI(IBF(IN,ITHI))
      VAL=AMP*VAL
C OADP(OUT,ITHO,VAL,BUFO)
C IGNORE OUTPUT BEFORE BEGINNING OF FILE MOD 12JUN84
      IF(ITHO.LE.0) GOTO 109
      IF(OUTVHI(OUT,ITHO)) CALL GETVR(OUT,ITHO,BUFO)
      BUFO(IBF(OUT,ITHO))=VAL+BUFO(IBF(OUT,ITHO))
109   CONTINUE
      ITHI=ITHI+1
      ITHO=ITHO+1
      AMP=AMP+ADEL
10    CONTINUE
      MAXOUT=ISTO+NPTS-1
      IF(MAXOUT.GT.IVFDEF(4,OUT)) IVFDEF(4,OUT)=MAXOUT
      CALL PUTVR(OUT,BUFO)
      RETURN
      END
```