THE PERCEPTION OF THE PHONEME *ALEPH* (/'/) IN MODERN HEBREW*

By

CHRISTOPHER FARRAR AND YEHIEL HAYON

Ohio State University

1. INTRODUCTION

1.1 Historical Background

THE HEBREW CONSONANT *Aleph* was historically realized as a glottal stop ([']), and originally could occur in any position within the word. However, as early as

*This study was supported by a generous grant from the Graduate School of The Ohio State University.

It would not have been possible without the help of a great many people, both in the United States and Israel. In particular, we would like to thank the following people, all at The Ohio State University: Professor Rob Fox for his help with the statistical analysis of the data; Professor Ilse Lehiste for her inestimable assistance at every stage of the project from its inception to its conclusion; Judy Moglia for her support and willingness to help with the legwork; and Professor Arnold Zwicky, for his encouragement in the project and for his willingness to take time from a very busy schedule for the purpose of reading numerous drafts of this paper.

At Haskins Laboratories in New Haven, Connecticut, where the crucial work of synthesis was done, special thanks are extended to Professor Arthur Abramson for his help in getting the nasals right; Dr. Terry Halwes for instruction in the fine art of mounting disks as well as initiation into the arcane mysteries of Deedee; Professor Alvin Liberman for extending to us the fabulous facilities of Haskins; and to Patti Price for help with some of the subtleties of speech synthesis.

In Israel we wish to express our gratitude to Ms. Nili Mor for her indefatigable efforts in finding subjects for the experiment; Professor Shelomo Morag for his encouragement and for putting us in touch with people who would help us in our endeavor; and Ms. Edna Poles at the Hebrew Gymnasia, for persuading her students to sit in school after hours to serve as subjects.

Finally, we wish to express our extreme gratitude to all of our subjects, who sat so patiently through what must have been a rather tedious experience for them. *Toda rabba!*

the Biblical Hebrew period the glottal stop tended not to appear phonetically when it occured at the end of a syllable.[1] In Modern Hebrew (hereafter MH) the loss of the glottal stop is very common and can be observed in all positions in the word. For example: /niš'ár/→ [nišár] 'he stayed'; /šo'él/→ [šoél] 'he asks.'

Since the glottal stop has phonemic status in Hebrew, its frequent loss may create a very large number of homophonous words, e.g.:

/ša'áv/ 'he pumped'→ [šav]← /šav/ 'he returned'
/qar'á/ 'she read' → [kará] ← /qará'/ 'he read'

There is evidence, however, that speakers of MH often perceive a glottal stop where none occurs phonetically. For example, Semiloff-Zelasko (1975) reports a study which she conducted in which Hebrew words were heard in isolation by native speakers of MH. In many cases, a glottal stop was perceived in words which had none phonetically, having instead a period of laryngealization in a position corresponding to that of the underlying glottal stop. This finding and similar, more informal ones, raise two intriguing questions about the production and perception of the glottal stop by speakers of MH. First: By what phonetic and/or stylistic factors is the loss of the glottal stop conditioned? Second: What are the phonetic cues which facilitate the perception of a glottal stop in cases where it is not physically present?

1.2 Preliminary Research

An early attempt on our part at answering the first question quickly turned out to be an unmanageably large undertaking. In consequence we concentrated our efforts on investigating the second question. To this end we constructed a listening test which foreshadowed the current study. For this earlier experiment we elicited productions of the words *galá* and *kav* from a phonetically sophisticated adult male speaker of MH. The word *galá* is potentially ambiguous between /ga'alá/ 'she redeemed' and /galá/ 'he was exiled'; *kav* between /ka'áv/ 'it hurt' and /kav/ 'line.' In each production the speaker varied a single phonetic feature which was chosen because we suspected it might be a significant perceptual cue. Vowel length, intonation contour and laryngealization were manipulated in this way. The resulting productions were analyzed spectrographically to ascertain that the utterances did indeed vary in the fashion desired. The best ones were spliced out, copied, spliced again and then randomized for presentation to native speakers. For each token which the

1. In Biblical Hebrew the loss of the glottal stop was coupled with the lengthening of the preceding vowel, e.g.,/qārá'/→[qārā] 'he read.' For a discussion of glottal stop loss in Biblical Hebrew and earlier, see Blau (1972, pp. 50-51, 160, 176).

subjects heard they were required to indicate which of the two possibilities they thought it was. The test was administered to three subjects. The results were suggestive but not conclusive, due in part to the small number of subjects but more importantly to problems inherent in the use of natural speech for a test of this kind. The difficulty is that in aural speech (that is, speech produced by a human being) many of the fine phonetic details are beyond voluntary control. Thus, even with sophisticated training a speaker cannot produce two identical utterances. Variation will show up in the lengths of the various segments, in the formant frequencies of the vowels, in the length of aspiration of stop consonants, and in fact in every other aspect of the speech signal. Using natural speech it is impossible in principle to vary only a single feature at a time, and it is precisely this kind of control which is essential in a study which aspires to be scientifically rigorous. The stringent control necessary can be achieved only with a speech synthesizer. This device assembles a speech signal precisely according to the instructions of the operator, thereby making it possible to overcome the limitations imposed by the use of natural speech. For this reason, in the present study the tokens presented to subjects for judgment were produced exclusively with a speech synthesizer.

1.3 Research Hypotheses

The preliminary research described above led us to suspect that at least six kinds of modification of the phonetic signal could serve as perceptual cues for /'/; that is, could affect a listener's judgment as to whether or not a speaker intended an underlying glottal stop. These were: (1) the length of the vocalic nucleus in a CVC syllable; (2) the presence of a period of laryngealization (creaky voice) in a location corresponding to that of the underlying glottal stop (Semiloff-Zelasko, 1975); (3) the occurrence of an actual glottal stop; (4) variations in the suprasegmental contour (pitch and amplitude) in the vocalic portion of a CVC syllable; (5) length of a consonant in the position immediately preceding that of the underlying glottal stop (Semiloff-Zelasko, 1975); and (6) the presence of a stressed syllable following the syllable containing the underlying glottal stop. These six types of modification served as the independent variables in the experiment. They will be discussed in considerable depth in Section 2.2.

We do not wish to claim here that perceptual cues which identify an underlying glottal stop in Hebrew do so in all languages which include a glottal stop in their phonemic inventory. Such cues identify only phones (that is, discrete phonetic events). The equation of a particular phone with a certain phoneme is governed by linguistic organizing principles which may differ from language to language.

We also wish to point out that the ability of a person to make a decision about two possible underlying forms (one with /'/, one without) does not necessarily imply that the person actually "hears" a glottal stop, in the sense than an English speaker "hears" an *n* in *can't* (where none exists phonetically, for most dialects). For the purposes of this paper, the term *cue for an underlying glottal stop* should be understood as referring to a phonetic event which allows a native speaker to decide between two possible underlying representations.

1.4 Purpose and Organization

Our purpose in conducting this study was two-fold. First, we wished to investigate what seemed to us to be an interesting problem and to make the outcome known to the scholarly community. Second, we wished to promote our methodology among scholars doing research in Hebrew linguistics. This methodology relies heavily on data obtained by instrumental means (e.g., through the use of sound spectrograms) and through perceptual testing with synthetic stimuli.

The organization of the paper reflects these aims. Wherever technical information figures in the study we have provided two partially overlapping discussions. In each case the first is a non-technical overview of the matter at hand. This is intended for scholars who wish to understand the matter in principle but lack either the inclination or the technical background to pursue it in depth. In these discussions, technical terms will be avoided where possible and explained where unavoidable.

The second discussion presupposes some technical background, as it is intended for scholars potentially interested in replicating the study or in pursuing similar investigations of their own.

The essentials of the study can be appreciated through the non-technical overviews alone.

2. THE EXPERIMENT

In 2.1 we provide a non-technical overview of the experiment. Full technical detail is supplied in 2.2

2.1 Overview

In order to test the variables listed in Section 1.3 we chose the three stimulus prototypes *šav*, *šavá* and *nasá*, each of which is potentially subject to two interpretations: For *šav* /šav/ 'he returned' and /ša'áv/ 'he drew (water)'; for *šavá* /šavá/ 'he imprisoned' and /ša'avá/ 'she drew (water)'; for *nasá* /nasá'/ 'he carried' and /nas'á/ 'she carried.' Each of these three prototypes was synthesized with the help of a computer and then modified in such a way as to incorporate the variables for which we were testing. Because this experiment was to our knowledge the first of its kind in Hebrew linguistic research, we were faced with a significant problem. In the *šav* and *šavá* series vowel length was one of our independent variables. The other relevant modifications (presence of a glottal stop, laryngealization, pitch-amplitude contour and presence of a following stressed syllable) had to be manifested on the same vowel whose length we were interested in manipulating. There were two possibilities. We could have produced a series of stimuli for *šav* and *šavá* which varied only in the length of the vowel. In order to test the other variables, then, we would have had to choose one of those lengths and to have performed the modifications on it. This would have had the advantage of severely limiting the number of stimuli which subjects would have had to judge. A shorter and less tedious experiment would have resulted. The disadvantage was that any length we chose for the other modifications could have turned out to be a sufficient cue for an underlying glottal stop. Had we received confirming results for all of our modifications we would have had no way of knowing if it was the length of the vowel which was relevant or the other modifications. Alternatively we could have produced the whole series of modifications for each vowel length we wanted to investigate. This would have resulted in a very long, monotonous test. We compromised by repeating all modifications for each degree of length in the *šav* series, varying length only in the *šavá* series. The resulting experiment was still long, but not to the extent that performance was affected by fatigue.[2]

In the *šav* and *šavá* series we tested six vowel lengths. These vowel length categories were 120 milliseconds, 160 ms, 200 ms, 240 ms, 280 ms and 320 ms. In the *šav* series each of these length categories was associated with nine experimental conditions (modifications). These were BASE, in which vowel length was the only variable; GLOTTAL STOP, in which a glottal stop was inserted; ADJUSTED GLOTTAL STOP, which was identical to GLOTTAL STOP except that the pitch-amplitude contour (intonation) was modified to make the stimulus sound more natural; LARYNGEALIZED, in which a period of laryngealization (creaky voice) was introduced; DROP FØ, in which the pitch on part of the vowel was dropped; RAISED FØ, in which the pitch on the latter part of the vowel was raised; DROP AND RAISE FØ, which combined the two pitch modifications; RAISE AO, in which the amplitude (loudness) of the latter part of the vowel was raised; and COMBINATION, which combined all the pitch-amplitude modifications (i.e., DROP FØ, RAISE FØ, RAISE AO).

In the *nasá* series there were three experimental conditions. These were

2. The question of fatigue is taken up in Section 5.2.

LENGTH OF S, in which the length of the consonant *s* varied from 40 ms to 190 ms in increments of 15 ms (11 lengths in all); GLOTTAL STOP, in which a glottal stop was inserted immediately after the *s*; and LARYNGEALIZED, in which a period of laryngealization was introduced immediately after the *s*.

Once the stimuli had been produced they were randomized and assembled on a tape for presentation to subjects. Answer sheets were prepared. For each stimulus which the subject heard, a choice between two Hebrew words was offered, one containing an underlying glottal stop and the other not. The answer sheets were printed exclusively in Hebrew. The subjects, all of whom were native speakers of Hebrew, had to indicate which of the two possibilities they thought they heard. Data from 71 subjects were collected and transferred to computer cards. Tallying was carried out by computer, with spot checks being made by hand to ensure accuracy. The results of the tallying are presented in Section 2.2.4 (Tables 1 and 2). For a detailed discussion of the various modifications see Section 2.2.2.

2.2 Technical Information

2.2.1 *Equipment*

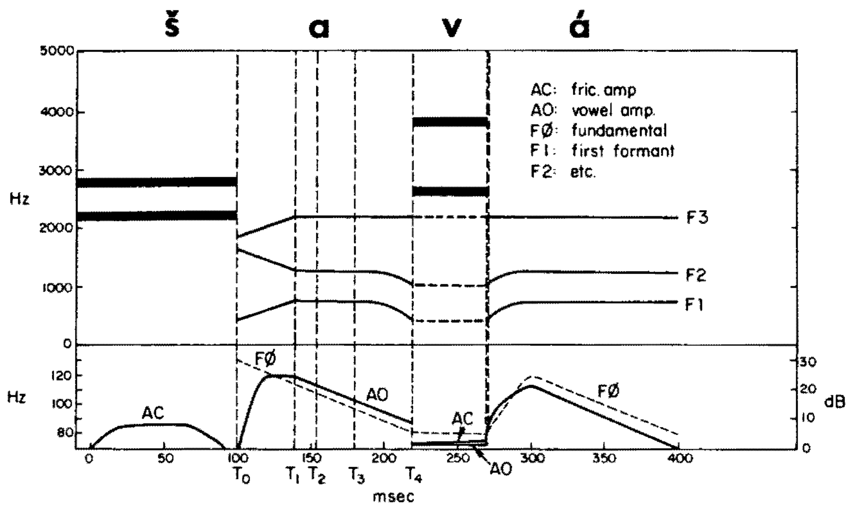
Stimuli for the experiment were generated at Haskins Laboratories in New Haven, Connecticut, using a DDP 224 computer in conjunction with the OVE III 15-parameter series synthesizer. The output from the synthesizer was recorded on magnetic tape and later copied using Ampex reel-to-reel tape recorders. The final copy of the tape was presented to subjects using a portable Uher 2400 Report-L tape recorder equipped with a junction box which enabled two subjects to listen simultaneously through identical Realistic Pro-20 headphones. The headphones had an advertised frequency response range of from 10 Hz to 16,000 Hz. The headphones were modified for binaural rather than stereo presentation.

2.2.2 *Experimental Design*

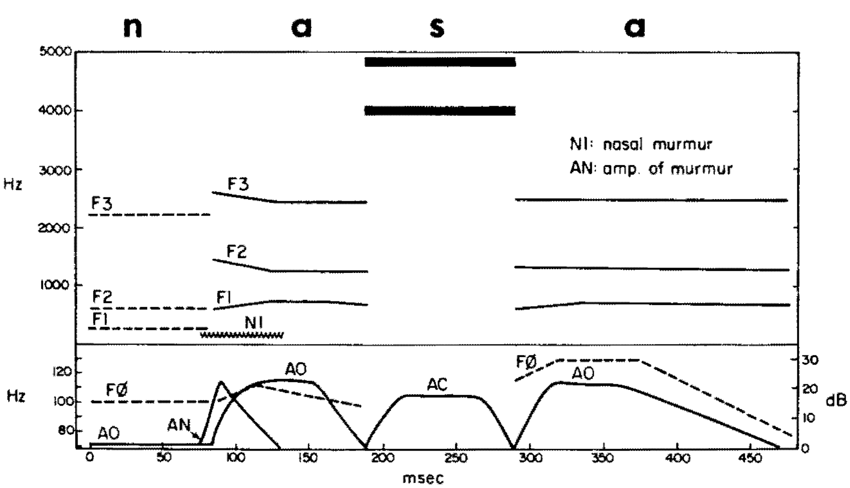
2.2.2.1 The Stimuli: Prototypes

The stimulus prototypes *šav*, *šavá* and *nasá* were created in the following fashion: Actual utterances of the three Hebrew words /šav/, /šavá/ and /nasá/ were recorded on magnetic tape by an adult male native speaker of Hebrew. These were analyzed by the Haskins Ubiquitous Spectrum Analyzer under the control of the DDP 224 computer. The end result of the analysis was that the acoustic information from the magnetic tape was converted to a representation in computer memory. This was presented as a spectrogram on a video display scope and converted to OVE III synthesizer parameter values by means of a light-pen. The tokens which resulted were painstakingly modified until the most natural-sounding forms were produced. These served as the prototypes for all of the test items.

Figure 1 provides graphic representations of the most important parameter values for the prototypes.



1a.



1b.

Fig. 1: Synthesizer parameters for the prototypes

2.2.2.2 The Stimuli: Variations

Šav

Visual inspection of spectrograms of actual Hebrew words revealed that in a word of the form /Ca'áC/ the phonetic realization of the underlying glottal stop was centered on a point about one third through the vocalic portion of the word. For this reason, a point one third through the vowel of the synthesized prototype was selected as the center of reference for all modifications introduced (except for the length modification, which had no center of reference, the nucleus simply being extended the required amount).

The vocalic nucleus was divided into intervals as shown in Figure 2. T_0 is the beginning of the vowel; T_4 is the end. The interval *a* is equal to *d* which is equal to *b* plus *c*, and the interval *b* is equal to one third of *a*.

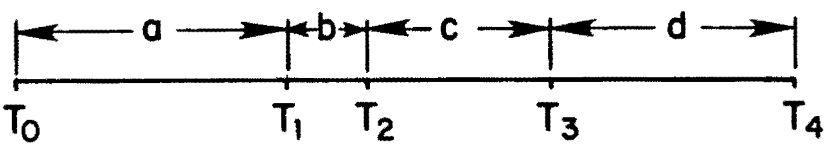


Fig. 2

In all the stimuli in the *šav(á)* series, formant transitions reach their steady-state values 40 ms into the vowel, as shown in Figure 1a.

The following paragraphs detail the modifications made on the prototype *šav*. The diagrams provide information about the suprasegmental contour of the item in question. The value of FØ at the points marked A, B, C, D and E is 130 Hz, 100 Hz, 112 Hz, 125 Hz and 80 Hz, respectively.

Figure 3 is provided as an example. In Figure 3, the FØ contour starts at 130 Hz, drops to 100 Hz at T_1, rises to 125 Hz at T_3 and falls to its terminal value of 80 Hz at T_4. Values intermediate between those specified were interpolated by the computer.

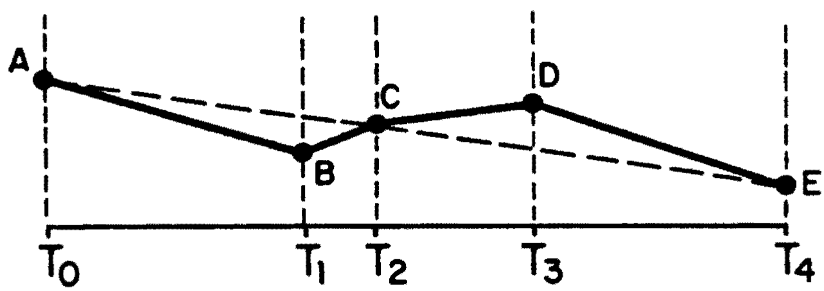
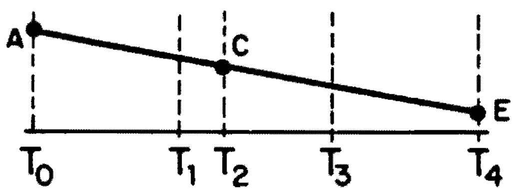


Fig. 3

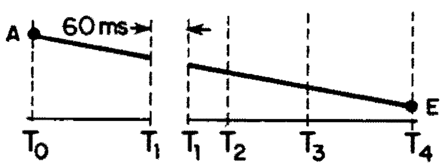
The amplitude contour for the majority of the stimuli is the same as that depicted in Figure 1a. The only modification consisted of raising it to 30 dB between T_2 and T_3, allowing it to fall off subsequently to its terminal value of 8 dB at T_4, as shown in Figures 6, 11 and 12.

Following are the variants of the prototype *šav*. Each of them is manifested over vocalic nuclei of lengths 120 ms, 160 ms, 200 ms, 240 ms, 280 ms and 320 ms.



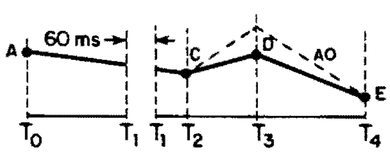
Name: BASE

Fig. 4



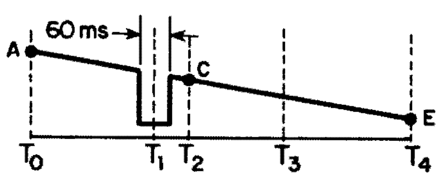
Name: GLOTTAL STOP

Fig. 5



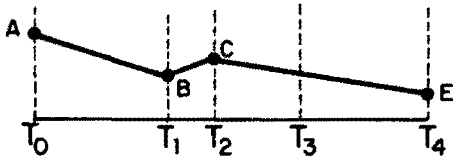
Name: ADJUSTED GLOTTAL STOP

Fig. 6



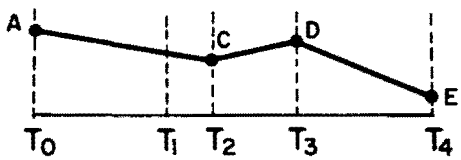
Name: LARYNGEALIZED

Fig. 7



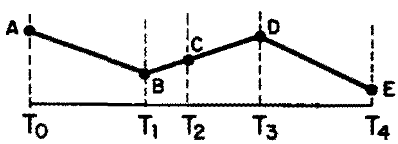
Name: DROP FØ

Fig. 8



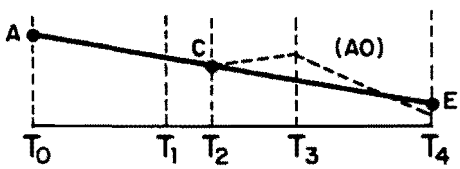
Name: RAISE FØ

Fig. 9



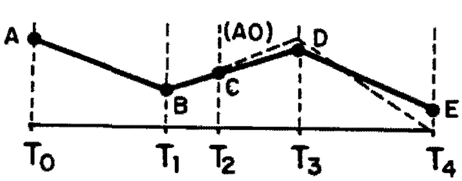
Name: DROP & RAISE FØ

Fig. 10



Name: RAISE AO

Fig. 11



Name: COMBINATION

Fig. 12

Comments:

1. In the shortest length condition (120 ms) the BASE form is identical to the prototype.
2. In the GLOTTAL STOP condition 60 ms of silence have been inserted at T_1 to simulate a glottal stop. Note that the entire nucleus is longer by 60 ms than that of the BASE form.
3. In the ADJUSTED GLOTTAL STOP condition 60 ms of silence have been inserted at T_1 to simulate a glottal stop. In addition the amplitude has been raised to 30 dB at T_3 and the FØ contour modified as indicated to produce a more natural sounding stimulus.
4. In the LARYNGEALIZED condition, FØ has been dropped to 50 Hz for a 60 ms interval centered on T_1. This simulated laryngealization. Note that the vocalic nucleus has the same length as in the BASE condition.
5. Conditions represented in Figures 8 through 12 test modifications of the pitch-amplitude contour.
6. In the RAISE AO condition the amplitude (AO) is raised to its peak value of 30 dB between T_2 and T_3. It then falls to its terminal value at T_4.
7. The COMBINATION condition combines all pitch-amplitude modifications. The intention is to provide a very natural sounding stimulus.

Šavá

The *šavá* series was produced by adding the stressed vowel *a* to the BASE form for each length category. This series thus contains only six distinct stimuli.

Nasá

The prototype for the *nasá* series was modified in three ways:
1. LENGTH OF S. The *s* ranged from 40 ms to 190 ms in length, in increments of 15 ms. There were thus 11 length categories.
2. GLOTTAL STOP. 60 ms of the vowel following the *s* were deleted to simulate a glottal stop.
3. LARYNGEALIZED. Beginning immediately after the end of the *s* FØ was dropped to 50 Hz to simulate laryngealization.

2.2.2.3 Number of Stimuli and Number of Randomizations

The number of stimuli which appeared in the final version of the experiment is as follows:

šav series:	9 conditions times 6 lengths, giving 54 stimuli
šavá series:	6 lengths, each occurring 3 times, giving 18 stimuli

nasá series:	11 lengths, 1 with glottal stop, 1 laryngealized, each occurring 3 times, giving 39 stimuli
Total:	111 stimuli

Two separate randomizations of these stimuli were presented to the subjects, bringing the grand total to 222. The interval between presentations of successive stimuli was 4 seconds.

Ten stimuli were copied onto the beginning of the test as examples, to allow the subjects to grow accustomed to the voice of the "speaker."

The 222 stimuli were arranged in groups of 10 by splicing in the words (in Hebrew) "Group No.______" spoken by an adult native speaker. Twenty-two groups were formed in this fashion, the last containing 12 rather than 10 stimuli. This grouping was reflected on the answer sheets. The purpose of grouping the stimuli was to prevent subjects from losing their place and to break the monotony of the presentation.

In order to prevent possible fatigue effects, two three-to-five minute breaks were provided, one after group 7 and the other after group 15. During the breaks subjects filled out personal information sheets and chatted with the experimenter. The entire experiment took 35-40 minutes.

Instructions to the subjects were presented in writing on the answer sheets, explaining the format of the experiment and the task required. This task was to mark with an X the word they thought they heard, leaving no blanks and guessing where necessary. The examples were then announced and presented, followed by the body of the test.

A sample of the answer sheet is presented in Figure 13.

2.2.3 *Subjects*

The subjects for the experiment were 50 students from the Neri Bloomfield Vocational High School and Teacher Training College in Haifa, and 23 from the Hebrew Gymnasia in Jerusalem. Ages ranged from 16 to 24, the majority being under the age of 18. Data from 71 of these subjects were processed. The responses of one subject were not included in the tallying because his personal data sheet (discussed below) was missing. The other subject was excluded because he indicated that he was not a native speaker of Hebrew.

The ideal subject for this experiment would have been a second generation native speaker of Israeli Hebrew, 16 to 18 years of age, linguistically naive and absolutely monolingual. Judgments from such a subject would be the least likely to reflect interference from other languages or from intellectual understanding of the aims of the experiment. Due to the immigrant nature of most of the population of Israel, such subjects tend to be very rare. In order to run the experiment at all it was necessary to accept subjects who did not have all the desirable characteristics. At the same time we felt we needed some kind of control which would allow us to isolate any possible effects of undesirable subject characteristics. To this end we had the subjects fill out a questionnaire which requested such information as the date and place of birth; the national

דף-תשובות 2

קבוצה מס' 3

31. נשא ___ נשאה ___
32. נשא ___ נשאה ___
33. שב ___ שאב ___
34. שב ___ שאב ___
35. נשא ___ נשאה ___
36. שב ___ שאב ___
37. שבה ___ שאבה ___
38. נשא ___ נשאה ___
39. שב ___ שאב ___
40. שב ___ שאב ___

קבוצה מס' 4

41. שב ___ שאב ___
42. שב ___ שאב ___
43. נשא ___ נשאה ___
44. שבה ___ שאבה ___
45. נשא ___ נשאה ___
46. נשא ___ נשאה ___
47. שב ___ שאב ___
48. שב ___ שאב ___
49. שב ___ שאב ___
50. נשא ___ נשאה ___

קבוצה מס' 5

51. נשא ___ נשאה ___
52. נשא ___ נשאה ___
53. נשא ___ נשאה ___
54. נשא ___ נשאה ___
55. נשא ___ נשאה ___
56. נשא ___ נשאה ___
57. שבה ___ שאבה ___
58. שבה ___ שאבה ___
59. שב ___ שאב ___
60. שב ___ שאב ___

קבוצה מס' 6

61. נשא ___ נשאה ___
62. שב ___ שאב ___
63. שבה ___ שאבה ___
64. שב ___ שאב ___
65. שב ___ שאב ___
66. שב ___ שאב ___
67. נשא ___ נשאה ___
68. נשא ___ נשאה ___
69. שב ___ שאב ___
70. שב ___ שאב ___

קבוצה מס' 7

71. שב ___ שאב ___
72. שבה ___ שאבה ___
73. שבה ___ שאבה ___
74. נשא ___ נשאה ___
75. שב ___ שאב ___
76. שב ___ שאב ___
77. שב ___ שאב ___
78. שב ___ שאב ___
79. נשא ___ נשאה ___
80. שב ___ שאב ___

קבוצה מס' 8

81. נשא ___ נשאה ___
82. שב ___ שאב ___
83. נשא ___ נשאה ___
84. נשא ___ נשאה ___
85. שבה ___ שאבה ___
86. שב ___ שאב ___
87. נשא ___ נשאה ___
88. שבה ___ שאבה ___
89. שב ___ שאב ___
90. שב ___ שאב ___

Fig. 13. Sample answer sheet

origin and native language of the subject; languages with which he was familiar and the degree of competence in those languages; whether or not the subject had lived outside of Israel (also where and for how long). Subject categories were determined on the basis of information from these questionnaires. The intent was to use these categories to investigate any contradictions or anomalies. As it turned out, neither could be discerned in the data so the matter was not pursued.

2.2.4 *Results from the Tallying*

Table 1 shows the results of the experiment. Table 1a presents results for the *šav* series. Table 1b presents results for the *šavá* series. Table 1c presents results for the *nasá* series. The number in each cell is the total number of positive responses for that cell (that is, the total number of judgments that an underlying glottal stop *was* present).

	120ms	*160*	*200*	*240*	*280*	*320*
BASE	1	4	35	112	100	104
G. STOP	95	133	140	141	141	141
ADJ. GL. STOP	103	140	141	142	142	142
LARYNGEALIZED	4	53	130	138	140	142
DROP FØ	1	7	100	124	138	142
RAISE FØ	1	18	83	115	141	140
DROP & RAISE FØ	1	8	103	134	139	141
RAISE AO	0	29	105	138	138	140
COMBINATION	1	75	137	142	142	141

1a. *šav*. 142 possible in each cell

120ms	*160*	*200*	*240*	*280*	*320*
238	343	403	405	392	381

1b. *šavá*. 426 possible in each cell

LENGTH OF S	*40ms*	*55*	*70*	*85*	*100*	*115*
	170	198	191	227	259	239
	130	*145*	*160*	*175*	*190*	
	223	193	206	182	160	
GLOTTAL STOP	422		LARYNGEALIZED			417

1c. *nasá*. 426 possible in each cell

Table 1: Number of positive responses in each test condition

Table 2 presents the same information in percentages.

	120 ms	*160*	*200*	*240*	*280*	*320*
BASE	1	3	25	79	70	73
GL. STOP	67	94	99	99	99	99
ADJ. GL. STOP	73	99	99	100	100	100
LARYNGEALIZED	3	37	92	97	99	99
DROP FØ	1	5	70	87	97	100
RAISE FØ	1	13	59	81	99	99
DROP & RAISE FØ	1	6	73	94	98	99
RAISE AO	0	20	74	97	97	99
COMBINATION	1	53	97	100	100	99

2a. *šav.*

120 ms	*160*	*200*	*240*	*280*	*320*
56	81	95	95	92	89

2b. *šavá*

LENGTH OF S	*40 ms*	*55*	*70*	*85*	*100*	*115*
	40	47	45	53	61	56
	130	*145*	*160*	*175*	*190*	
	52	45	48	43	38	
GLOTTAL STOP	99		LARYNGEALIZED			98

2c. *nasá.*

Table 2: Percent positive responses (rounded)

3. STATISTICAL ANALYSES OF THE DATA

3.0 Introduction

The discussion in Section 3.1 is intended primarily for readers whose research does not involve the use of statistical techniques for analyzing data. For those conversant with the theory and terminology of statistics, a more technical presentation is provided in Section 3.2.

3.1 Statistical Overview

Quantitative data lend themselves to the use of statistical tests which are designed to determine the probability that an observation or a series of observations is the result of chance. The lower this probability is, the more importance the results have for the researcher.

When the two statistical tests Chi-square (X^2) and Analysis of Variance (ANOVA) were applied to the data of Table 1 we were able to draw the following conclusions with a high degree of confidence:

(1) The pattern of positive responses to stimuli in the *nasá* series cannot be attributed to chance. Consequently, a closer look is warranted to determine what factors may have produced the observed results. See Section 4.1 for more discussion.

(2) The number of positive responses to each of the six versions of *šavá* (differing in the length of the first vowel) is significantly greater than can be accounted for by chance. An explanation for this fact is suggested in Section 4.2.

(3) When the data for the *šav* series are analyzed, the following conclusions can be drawn:

(a) The pattern of positive responses for the BASE condition is significantly different from the patterns for all other experimental conditions (that is, GLOTTAL STOP, RAISE AO, etc.). In particular, the number of positive responses rises (with length) more slowly and peaks lower (at around 74 percent) in the BASE condition than in the other conditions. Furthermore, the point at which positive and negative responses occur with equal frequency (the "random choice point") comes significantly later in the BASE condition (between lengths 200 ms and 240 ms) than in the other conditions. This is significant because the random choice point is that point at which positive responses will begin to predominate over negative ones. The effectiveness of a cue can be judged by the extent to which it produces positive responses. All things being equal, the more effective cue is the one which favors positive responses at a shorter vowel length. The fact that the number of positive responses increases with length for all conditions suggests that vowel length is the single most important cue to the perception of a glottal stop in this experiment. This conclusion is supported by our statistical testing. The fact that the random choice point occurs later in BASE (where only length is varied) than in the other conditions (where length and other factors vary) is quite reasonable, since every condition but BASE provides the subject with multiple cues which can be expected to be more effective than any single cue. This matter is taken up in Section 4.

(b) The patterns of responses for conditions other than BASE are significantly different from each other only in the three shortest length categories (120 ms, 160 ms and 200 ms). In other words, once the vowel reaches a certain length (240 ms) the other modifications become redundant and no longer affect perception in a measurable fashion. This matter is discussed in Section 4.

(c) The patterns of positive responses for the conditions GLOTTAL STOP and ADJUSTED GLOTTAL STOP do not differ significantly. Apparently the enhanced naturalness of the latter did not affect the subjects' perceptions.

3.2. Statistical tests

The data were analyzed using Pearson's Chi-square goodness-of-fit test and two-way ANOVA for non-replicated data. The findings are discussed in Section 4.

Chi-square tests

1. *Nasá*

Figure 14 plots the number of positive responses against the length of the consonant *s*. The resulting distribution is significantly different from chance ($p<.001$). Response frequencies for the GLOTTAL STOP and LARYNGEALIZED conditions were likewise significant ($p<.001$).

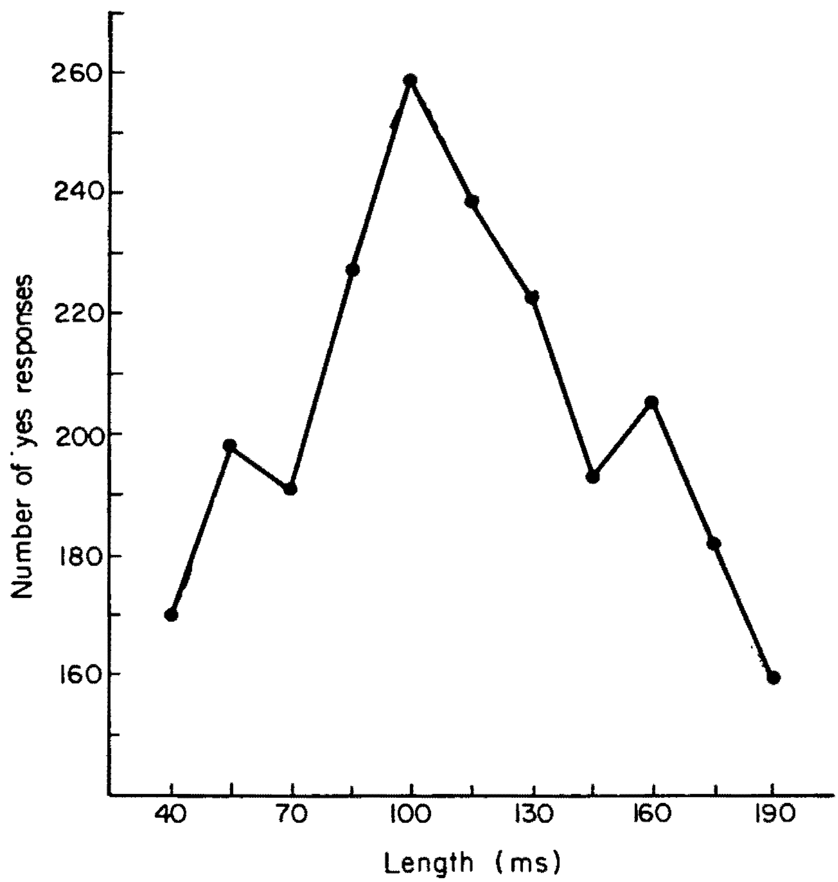


Fig. 14: Nasá: Number of positive responses

2. *Šavá*

Response frequency for the 120 ms length is significantly different from chance (p <.025). Since this is the shortest length category it is clear that response frequencies for the other five length conditions are even more significant.

3. *Šav*

a. Tests over All Length Categories

For these tests, response frequencies (across all length categories) for one experimental condition (e.g., BASE) were taken to represent the expected frequencies in the Chi-square test. The distributions of responses for other experimental conditions were then tested against this expected distribution. The results of these tests are summarized in Table 3. The distributions treated as being expected are given across the top of the table. The distributions which were compared against these are listed vertically. The table shows that the distributions for the GLOTTAL STOP conditions and the ADJUSTED GLOTTAL STOP condition did not differ significantly whereas all other comparisons did show significant differences.

	BASE	ADJ. GL. ST.	COMBIN.
BASE	———		
GL. ST.	p<.001	NS	
ADJ. GL. ST.	p<.001	———	
LARYNG.	p<.001		
DROP FØ	p<.001		p<.001
RAISE FØ	p<.001		p<.001
DROP & RAISE FØ	p<.001		p<.001
RAISE AO	p<.001		p<.001
COMBIN.	p<.001		———

Table 3: X^2 tests over all length categories

b. Tests over the Three Longest Length Categories

It appears from Tables 1 and 2 that the greatest variation in response frequencies comes in the three shortest length categories, the frequencies for the three longest lengths varying considerably less among themselves. Consequently the tests described in the preceding paragraph were repeated restricting the corpus to the response frequencies for the lengths 240 ms, 280 ms and 320 ms. The results are presented in Table 4. This table shows that significant differences appear only when the various experimental conditions are tested against the BASE condition.

	BASE	ADJ. GL. ST.	COMBIN.
BASE	———		
GL. ST.	p<.001	NS	
ADJ. GL. ST.	p<.001	———	
LARYNG.	p<.001		
DROP FØ	p<.001		NS
RAISE FØ	p<.001		NS
DROP & RAISE FØ	p<.001		NS
RAISE AO	p<.001		NS
COMBIN.	p<.001		———

Table 4: X^2 tests over 3 longest length categories

In a further attempt to isolate the variation in the data, an ANOVA was performed on the data in these three longest length categories. The data for the BASE condition were not included. This test is described below in the section devoted to analysis of variance.

c. Random Choice Point

Inspection of Table 2 reveals that barring minor variations, the frequency of positive responses consistently increases with the length of the vowel. This being the case, it is important to locate (for each experimental condition) that point at which positive responses begin to predominate, that is, the point where judgments that an underlying glottal stop exists begin to be favored over judgments that it does not. At this random choice point negative and positive responses will occur with equal frequency. With N equal to 142, the interval of responses *not* significantly different from chance (i.e., 71) is $59.3<n<82.7$. For the BASE condition, the random choice point falls between lengths 200 ms and 240 ms. For the LARYNGEALIZED condition and all of those below it in Table 1 (except COMBINATION) the random choice point lies between 160 ms and 200 ms. For GLOTTAL STOP and ADJUSTED GLOTTAL STOP the number of responses in the first length category is already significantly greater than can be attributed to chance. Figure 15 gives the locations of the random choice points.

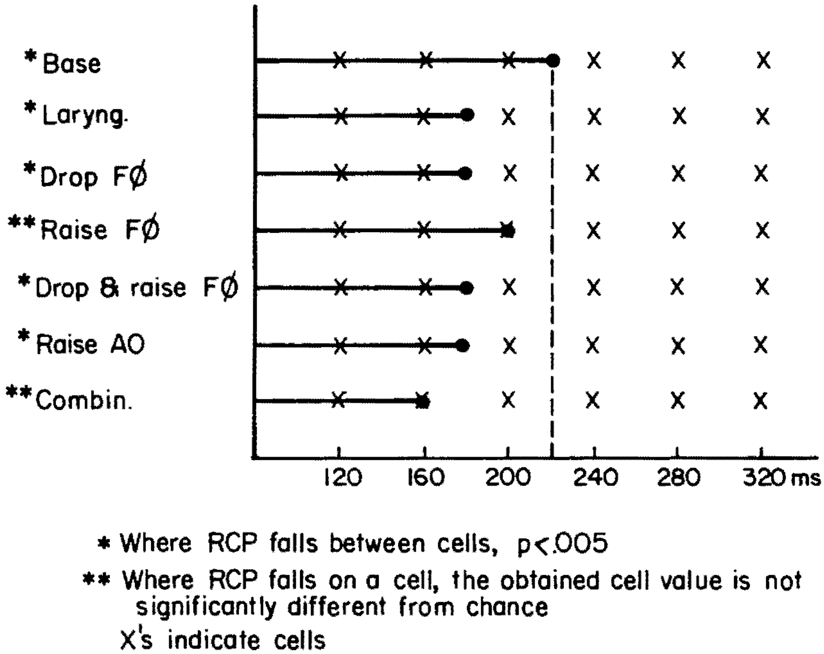


Fig. 15: Random choice points

Analysis of Variance

The data for *šav* were divided into three sets. Set I included all the data. Set II included only the data for lengths 240 ms, 280 ms and 320 ms. Set III was identical to set II except that it excluded the data from the BASE condition. This was done because the Chi-square test described earlier suggested that most of the variation in the three longest length categories could be attributed to the BASE condition.

A two-way ANOVA for non-replicated data was used. The results are as follows:

Set I (all data): Significant effects for length only ($p < .01$).

Set II (3 longest lengths): Significant effects for conditions only ($p < .01$).

Set III (set II less BASE): No significant effects for length or for conditions.

4. DISCUSSION

In this part of the paper we discuss the meaning of the experimental findings.

4.1 *Nasá*

It was established on the basis of a Chi-square test (Section 3.2) that the pattern of positive responses to *nasá* in the LENGTH OF S condition cannot be attributed to chance. This pattern certainly was not predicted by the hypothesis we were investigating, that shorter consonants would favor positive responses and longer consonants would favor negative responses.[3] Instead the data show that one length, 100 ms, optimally favors the interpretation /nas'á/ 'she carried.' At other lengths the number of positive responses decreases in direct proportion to the size of the difference between the actual length and the optimal length of 100 ms.

This proportional relationship between the length of the consonant *s* and the number of positive responses can be represented by an inverted V whose peak is centered on 100 ms and whose two feet are located at 40 ms and 190 ms. This function is represented by the dashed line ABC in Figure 16. A Chi-square test showed that the obtained distribution of results (solid line) does not differ significantly from this hypothetical function ($p < .05$).

An identification function of this type favors a perceptual model in which an incoming stimulus is matched against an idealized mental representation.[4] A stimulus which matches this "prototype" closely is more likely to be considered an instance of the prototype than one which is a more distant fit. A prototype model such as this has been proposed in Repp (1977). Although Repp's model is proposed in connection with the perception of segments and syllables, we see no reason why there should not be idealized images of words as well. In explaining our results we would then suggest that for speakers of Modern Hebrew the perceptual representation of the word /nas'á/ is very much like our synthesized *nasá* with an *s* approximately 100 ms in length.

For a fuller discussion of the prototype model, we refer the reader to Repp (1977).

4.2 *Šavá*

With the *šavá* series the problem is to explain why it is that in even the shortest length category subjects demonstrate a significant preference for /ša'avá/, whereas with the BASE condition of the *šav* series (identical except for the existence of a stressed suffix *-á*) a preference for /ša'av/ does not appear until the vowel length is 240 ms.

The work reported in Lehiste (1972) will shed some light on the matter. In this paper the author found that citation forms of word triplets such as *speed,*

4. We are grateful to Rob Fox for bringing this to our attention.
3. This idea was originally suggested in Semiloff-Zelasko (1975).

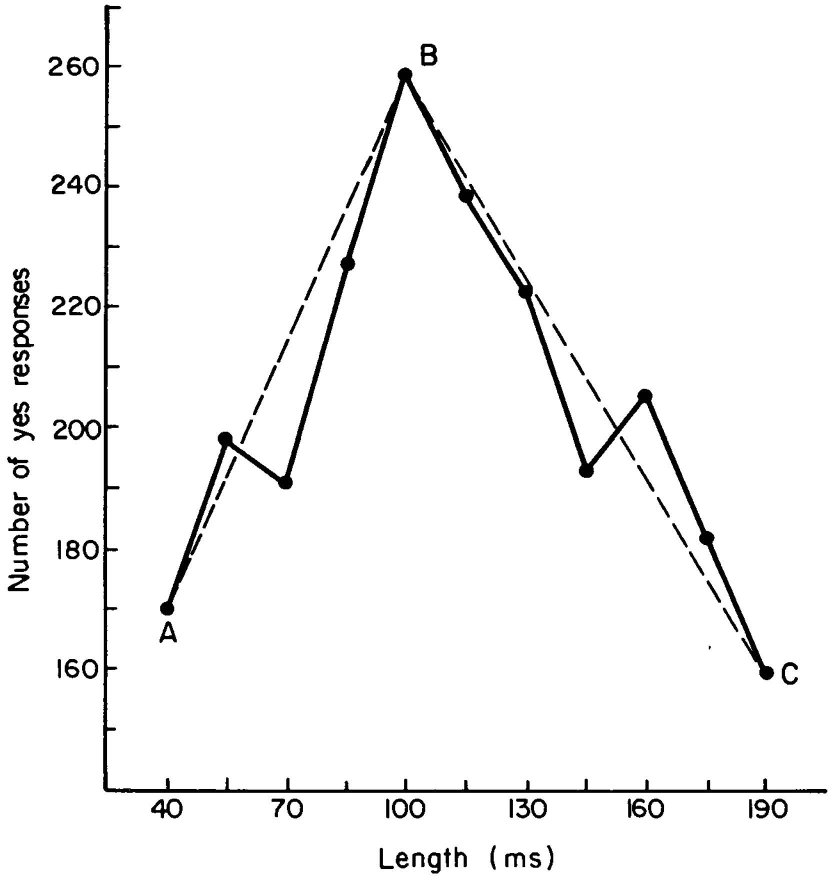


Fig. 16: Prototype model for nasá (dashed line)

speedy and *speedily* were all approximately equal in length, the first vowel being progressively shortened with the addition of each affix. Shortening of this type is characteristic of stress timed languages and is related to various phonological processes such as vowel reduction and vowel deletion. Modern Hebrew is clearly a stress timed language and can be expected to exhibit phenomena similar to those characteristic of such languages. Consider then the reaction of a native speaker of Hebrew upon hearing tokens of *šav* and *šavá* in which the first vowels of each were identical in length. This person would most likely associate *šavá* with an underlying form in which the nucleus is considerably longer, for example /ša'avá/. The reason is that if *šav* and *šavá* were both instances of the same underlying root /šav-/ the vowel of *šavá* should be considerably shorter than that of *šav*, just as the first vowel of *speedily* is shorter than that of *speed*. Therefore, where the two vowels are identical in length the word with the suffix will be associated with an intrinsically longer underlying form.

4.3 *Šav*

Careful consideration of the data and the statistical analyses presented so far for *šav* will reveal an interesting problem. The number of positive responses increases with length for all experimental conditions until the last three length categories, where it levels off. For every condition except BASE, the average values for the three longest lengths range from 92 percent to 100 percent. For BASE, however, the average value is 74 percent, which is significantly smaller than the values for the other experimental conditions. Furthermore, the ANOVA described in Section 3.2 revealed that the single most powerful experimental variable was vowel length. Why then should the percentage of positive responses for the BASE condition (where length alone was varied) peak at 74 percent while responses for the other conditions (where other cues co-occurred with length) peak much closer to 100 percent?

For the conditions other than BASE the stability of the values in the three longest vowel lengths can be explained in terms of a ceiling effect;[5] that is, no matter how effective a cue is, it cannot result in response frequencies higher than 100 percent. Consequently, differences in the effectiveness of experimental conditions cannot be measured past a certain point by the method which we used in this study. Other methods, e.g., reaction times, might of course be more revealing.

Obviously, ceiling effects cannot be invoked to explain the much lower peak in the BASE condition. For this it is necessary to reconsider the nature of the cues used in this study. All of our experimental variables can contribute to the perception of an utterance as being bisyllabic.[6] The glottal stop and laryngealization variables of course provide something extra: a discrete physical event which can be directly associated with an underlying glottal stop. In other cases, however, the subject must base his decision on whether he hears one syllable or two. A perceived monosyllable indicates that there is probably no glottal stop underlyingly. A perceived bisyllable indicates to the subject that there very likely is. The occurrence of an extra long vowel favors a bisyllabic percept. The perception of a bisyllable is further reinforced by the superimposition of additional cues such as certain changes in the pitch-amplitude contour.

5. Suggested by Guy Carden, personal communication.
6. This is true of vowel length in languages which do not have a phonemic length distinction.

Having dealt with that, we may turn to another problem. This is the question of the quite different results obtained under the LARYNGEALIZED condition in the *šav* series as opposed to the *nasá* series. In the *šav* series laryngealization is not nearly as powerful a cue as is the presence of a glottal stop. In the *nasá* series the two cues are equally effective, resulting in positive responses of nearly 100 percent in both cases.

In this connection it should be mentioned that even for trained phoneticians laryngealization and a glottal stop can be difficult to distinguish. In many cases uncertainty can be resolved only with the help of a spectrogram. This being the case, the different effectiveness of larygealization in the two series needs to be explained.

It is perhaps relevant that the length of the *s* for the laryngealized version of *nasá* was 100 ms, which is the optimal length for positive responses, as discussed earlier. However, if we are dealing with another case of multiply occurring cues being more effective than those occurring singly, it is still unclear why there should be a difference between LARYNGEALIZED and GLOTTAL STOP, since both are identical in this respect.

No ready explanation presents itself.

5. SOME SPECIAL PROBLEMS IN THE METHODOLOGY

5.1 Naturalness

A scientific experiment is an attempt to isolate a bit of nature in such a way as to facilitate observations which could not be made in a natural setting. The validity of the experiment crucially depends on the experimental variables being identical to the natural variables or else being faithful copies. Thus we could not conclude automatically from the current study that an actually occurring glottal stop in natural speech is a sufficient cue for an underlying glottal stop. We must first show that our synthetic glottal stop is a reasonable copy of a natural one. The same observation can of course be made about our other variables. In addition to considering the naturalness of the synthesized variables we must consider the naturalness of the whole synthesized utterance, since grossly unspeechlike stimuli may affect the subjects' judgments in unpredictable ways.

A number of observations need to be made here.

First, our choice of variables was guided by our experience (both impressionistic and instrumental) with natural speech. Our hypotheses were formulated on the same basis. The systematic nature of the results argues that subjects accepted our experimental manipulations as being genuine instances of the phenomena under study.

Second, spectrograms of representative tokens were made and examined.

These revealed that the experimental manipulations which we used were indistinguishable from the natural speech phenomena which they represented. Thus, synthetic laryngealization had the same form as natural laryngealization and synthesized glottal stop, pitch rise and amplitude rise were identical to their natural counterparts. The tokens as a whole looked slightly less natural than samples of actual human speech would, because the synthesized formants were extremely regular in frequency; and pitch, when it was being held constant, was more regular than it would be in natural speech. However, these subtleties were not apparent in aural presentation, and only one subject deduced that the speech he heard was computer generated.

In addition to making aural and spectrographic judgments of naturalness we chose our experimental variables in such a way that we would be able to gauge the effects of unnaturalness on our results. The condition ADJUSTED GLOTTAL STOP differed from GLOTTAL STOP only in that the former had a more natural pitch-amplitude contour than the latter. In addition, the condition COMBINATION was the most natural of the pitch-amplitude modifications because it combines contour variations which normally co-occur in natural speech. Interestingly, whereas there was no significant difference between response distributions for GLOTTAL STOP and ADJUSTED GLOTTAL STOP, COMBINATION *was* more effective than the other contour modifications. This is probably because from the subject's point of view, the occurrence of any single pitch-amplitude cue could well be a random event brought about by the vagaries of human speech; the combination of such modifications was too much to be attributed to chance, however, and therefore served as a more potent cue. Presumably the occurrence of an actual glottal stop was such a powerful cue that the addition of natural pitch-amplitude variation was simply superfluous.

If these explanations are correct, the data provide no evidence that naturalness was a factor in the subjects' responses.

5.2 Fatigue

Because the experiment was rather long we were concerned that toward the end subjects might be responding randomly due to fatigue. To avoid this we interrupted the experiment twice for three to five minutes each time. Some subjects indicated after taking the listening test that toward the end they had been responding randomly. However, the results show no evidence that this was the case.

6. CONCLUSION

In addressing the question of naturalness, we hope to have shown that our results generalize beyond the experiment which we conducted. We would like to be able to claim that there are at least two kinds of cues which enable native speakers of Hebrew to judge whether or not an utterance should be associated with a word containing an underlying glottal stop. In one category are cues which are in fact allophones of /ʾ/. These would be an actual glottal stop [ʾ] and laryngealization. In the other category are syllabication cues which achieve importance when there is no glottal stop allophone present.

We hesitate to say much about the cue of consonant length. We tested only a single consonant, /s/, and the relationship between the length of this consonant in *nasá* and the number of positive responses turned out to be more complex than we had originally envisioned. The exact nature of this relationship is a subject worthy of further study. Such a study should begin with an attempt to determine the natural limits of length variation in consonants preceding an underlying glottal stop. Statistical comparisons would have to be made with consonants occurring in words not containing an underlying glottal stop to see if significant differences exist. A perception study using synthetic speech could then be conducted based on the findings of the instrumental work.

In conclusion we would like to say that we hope that research in Hebrew phonetics and phonology will benefit from this study. In particular we hope that future studies will make use of instrumental and experimental observation as ours has done. Impressionistic observations should never serve as other than a starting point because one's impressions are extremely vulnerable to the influence of preconceptions of all sorts. Moreover, such studies are in principle not replicable because there is no way to guarantee that impressions will not vary from observer to observer.

BIBLIOGRAPHY

Blau, J. 1972. *Torat hahege vəhaccurot.* Tel Aviv.

Lehiste, I. 1972. "The Timing of Utterances and Linguistic Boundaries." *Journal of the Acoustical Society of America,* Vol. 51, No. 6 (Part 2), pp. 2018–2024.

Repp, B. 1977. "Dichotic Competition of Speech Sounds: The Role of Acoustic Stimulus Structure." *Journal of Experimental Psychology: Human Perception and Performance,* Vol. 3, No. 1, pp. 37–50.

Semiloff-Zelasko, H. 1975. "An Acoustic-Perceptual Study of Modern Hebrew 'Alef and 'Ayin," *Journal of Phonetics* 3: 167–174.