325

Journal of Phonetics (1980) 8, 491–499

Short Communication

A note on perceptuo-motor adaptation of speech

Quentin Summerfield,* Peter J. Bailey† and Donna Erickson
Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, *U.S.A.*

Received 16*th May* 1979

Abstract: Existing data on perceptuo-motor adaptation in speech are briefly reviewed, and an unsuccessful attempt to replicate the effect using a modified procedure is described.

Introduction

In both its evolution and its ontogeny human facility with speech has involved the co-development of the abilities to produce and perceive. Since the infant's speech matures to produce the contrasts of his or her native language, some influence whereby perception modifies production is implied. The elucidation of this influence has significance for theoretical accounts of the commonality underlying perception and production (for example, Liberman, Cooper, Shankweiler & Studdert–Kennedy, 1967) as well as for the design of programs of speech therapy (for example, McReynolds, Kohn & Williams, 1975).

Empirical support for a link between perception and production can be found in three types of demonstration: first, that perceptual sensitivity to variations in the acoustic properties of speech relates logically to their covariation in production (for example, Summerfield & Haggard, 1977); secondly, that productive and perceptive capabilities correlate within groups of individuals (for example, Bremer & McGovern, 1977); and thirdly, that an individual's immediate perceptual experience can exert measurable influences on his productions (for example, Lane & Tranel, 1971; Cooper, 1974). The last of these is the most direct and has been shown in several experiments by W. E. Cooper. In these experiments voice onset times (VOTs; Lisker & Abramson, 1964) in subjects' productions of voiced and voiceless initial stop consonants were measured after repeated exposure to an adapting sound. The basic result was that VOTs in productions of [pʰi] were shorter following perceptual adaptation with the voiceless adapter [pʰi] than following adaptation with the isolated vowel [i]. No effect on VOT was found in voiced productions (for example, [bi]) or following exposure to the voiced adapter [bi] (Cooper, 1974). Cooper & Lauritsen (1974) showed that the adaptation effect was not dependent on adapter and test syllable sharing place of articulation: repetitions of [pʰi] shortened VOTs in productions of [tʰi]. The effect could not be ascribed to mimicry because, overall, subjects produced shorter VOTs following adaptation rather than VOTs approximating that of the adapter. Cooper & Nager (1975) demonstrated that the bisyllabic adapter [rəpʰi] could shorten VOTs in productions in both [rəpʰi] and [rətʰi]. No systematic effects were observed on the duration of either the

*Present address: M.R.C. Institute of Hearing Research, The University of Nottingham, England.
†Present address: Department of Psychology, The University of York, England.

0095–4470/80/040491 + 09 $02.00/0

closure interval or the final stressed vowel, thereby ruling out the possibility that the effects on VOT were due to changes in speech rate or stress. The results of these experiments are summarized in Table I where it can be seen that, although they are systematic, the sizes of the changes in VOT are small. The changes correspond to less than 10% of the total duration of produced VOTs and are only about half the size of the shifts that can be induced in the phoneme boundary on a VOT continuum by perceptual adaptation (for example, 10.0 ms following adaptation with a voiceless aspirated stop; Eimas & Corbit, 1973).

Table I Perceptuo-motor adaptation effects

Test syllable uttered	Adapters				
	[bi]*	[pʰi]*	[rəbi]*	[rəpʰi]*	[sti]+ [stʰi]+ [tʰi]+
[bi][a]	+ 5.50 9/16 n.s.	+ 0.30 6/8 n.s.			
[pʰi][a]	− 0.40 4/8 n.s.	+ 5.60 13/16 $P < 0.05$			
[di][b]	n.s.	n.s.			
[tʰi][b]	n.s.	+ 3.20 23/32 $P < 0.01$			
[rəpʰi][c]				+ 2.73 12.5/20 $P < 0.05$	
[rətʰi][c]			+ 2.33 13/18 n.s.	+ 6.50 18/22 $P < 0.001$	
[sti][d]				n.s.	

Mean reduction in VOT (ms), proportion of subjects who showed a mean reduction, and significance levels of reductions reported in the sources indicated. Reductions in VOT are indicated by positive values.
[a]Cooper (1974)
[b]Cooper & Lauritsen (1974).
[c]Cooper & Nagar (1975).
[d]Cooper, Ebert & Cole (1976).
*Change in VOT with respect to [i]-adaptation.
+Change in VOT with respect to [si]-adaptation.

Cooper & Nager (1975) concluded that there was a genuine perceptuo-motor adaptation effect and that it was to be explained by the fatigue of neural elements which were presumed to be involved both in the perception of voiceless stops and in the abduction of the vocal cords. This "neural model" accounted for the absence of any effect following voiced adaptation and is consistent with the failure of voiceless adapters to change the VOT in voiced productions. The null results of Cooper, Ebert & Cole (1976) are also consistent with the model: none of the adapters [si], [sti], [sthi] and [thi] systematically affected the VOTs in productions of [sti]. These results were anticipated by the model since the vocal cords are adducted at the moment of stop release in [sti] (Ohala,[1]). Thus the data in Table I shows that systematic perceptuo-motor adaptation effects have been obtained for the utterances [phi], [thi], [rəphi] and [rəthi] only after perceptual adaptation with [phi] or [rəphi], when compared with the effect of the isolated vowel adapter [i]. One of the objectives of the present study was to determine whether a perceptuo-motor adaptation effect could be obtained with an alveolar adapter for stops produced at all three places of production.

A strong test of the Cooper & Nager (1975) model would require concurrent fiberoptic examination of the vocal cords and electromygraphic recordings from the intrinsic laryngeal musculature. The model would predict that, following adaptation with a voiceless aspirate stop adapter, the command to abduct the vocal cords would be modified in either or both of two ways. The command could either be weakened so that the cords were abducted to a smaller extent, or occur later relative to the adduction command. In both cases, the vocal cords would be less abducted at the moment when adduction commenced, so that the state of approximation required for voicing to onset would be achieved earlier. Preliminary results from a pilot study in which electromyographic recordings were made from the intrinsic laryngeal musculature are mentioned below, but the cost of such research rendered it necessary to demonstrate that Cooper's basic effect could be obtained reliably and efficiently with a small number of subjects. This was another objective of the present study. Also, since it is desirable to keep testing sessions as short as possible in electrophysiological experiments, we sought in this preliminary study to determine whether the basic perceptuo-motor effect could be obtained with a smaller number of repetitions of the adapter than Cooper had used.

There were three major differences between our procedure and that used by Cooper:

(1) We adapted subjects with tokens of their own speech so that each subject had a different natural speech adapter rather than the same synthetic adapter. We reasoned that any perceptuo-motor link would be most likely to be accessed by a speaker's own speech.

(2) We used the bisyllable [rəthi] as the adapter, rather than [phi] (Cooper, 1974; Cooper & Lauritsen, 1974) or [rəphi] (Cooper & Nager, 1975).

(3) After an initial period of one minute during which 60 repetitions of the adapter were presented we maintained perceptual adaptation with 10 repetitions prior to each production, following Bailey[2] and Ganong[3]. Cooper had always presented 70 repetitions of the adapter prior to each production.

This revised procedure did not provide a successful replication of Cooper's basic perceptuo-motor adaptation effect. We shall discuss the extent to which our failure may have resulted from the differences in procedure described above. We present our data to advise others against taking the same course.

Method

Subjects

Three female and three male undergraduates served in two experimental sessions lasting about 40 minutes each. They were tested individually and were paid $5.00 for taking part in the experiment.

[1] Ohala, J. J. (1970). Aspects of the control and production of speech. *UCLA Working Papers in Phonetics*, 15.

[2] Bailey, P. J. (1974). Procedural variables in speech adaptation. *Speech Perception: Report on research in progress in the Department of Psychology*, Belfast: The Queen's University of Belfast. 2.3.

[3] Ganong, W. F. (1975). III. An experiment on "Phonetic" adaptation. *RLE Progress Report*, 116, 206–210. Cambridge Massachusetts: MIT.

Procedure

Both sessions of the experiment were conducted in a sound-attenuating booth. Subjects' productions were recorded on a magnetic tape with an Ampex AG500 tape recorder and a Shure Model 51 microphone. At the start of the first session each subject practised uttering the bisyllables [rəpʰi], [rətʰi] and [rəkʰi] in a natural voice with stress on the second syllable. A randomized list was prepared which included 20 instances of each of the three syllables; the list was arranged in columns of 22 items of which the first and last would not be included in subsequent analyses. Each subject was instructed to begin at the top of each column and to utter one token of each item of the list with 10 s between each item in time with a visual metronome. A brief rest was taken at the end of each column. At the end of the session each subject produced a series in instances of the isolated vowel [i].

We measured two durations in each token. One was the period of devoicing prior to the release of the stop, which we shall refer to as the closure interval; the other was the duration of the VOT. The measurements were made using spectrum and waveform manipulation routines available on the Haskins Laboratories PDP 11/45-GT40 computer system. Preliminary tests showed that the intervals described below were measured most easily when the higher formants had been filtered out. Thus, each token was low-pass filtered at 1.5 kHz. before being digitized with a sampling rate of 10 kHz. The durations were measured from the displayed waveforms by aligning a cursor first with the left-hand boundary and then with the right-hand boundary of the desired interval. The program displays the duration of the demarcated interval. The resolution of the display is variable; with maximum resolution the cursor may be positioned to the nearest 0.2 ms. Our criteria for measuring the two intervals were as follows. The beginning of the period of devoicing was defined as the peak of the last detectable pitch pulse in the segment [r] and its end was the beginning of the stop release transient. The VOT was measured from this point to the peak of the first detectable pitch pulse; the first pitch pulse was often difficult to specify. The measures were made on the basis of a consensus of at least two of the three experimenters. From these measures we computed the means and the standard deviations of the closure interval and the VOT for productions of each of the three utterances [rəpʰi], [rətʰi] and [rəkʰi] for each subject. For each subject we then selected that production of [rətʰi] whose VOT was closest to the mean VOT produced by that subject in instances of [rətʰi], and a production of the isolated vowel [i] whose duration was at least as long as the total duration of [rətʰi].[4] These tokens were low-pass filtered at 3.2 kHz and digitized with a sampling rate of 10 kHz. The token of [i] was edited to have the same overall duration as the token of [rətʰi]. A perceptual adaptation tape was constructed from each token in the following format: 60 presentations of the token at a rate of 1/s were followed by 22 trials each consiting of ten repetitions of the token followed by a 3-s pause. Thus, there were two tapes for each subject, one each for the adapters [i] and [rətʰi]. These tapes were used in the second session of the experiment.

In the second session each subject listened three times in alternation to each of the tapes constructed from his or her own speech. Subjects rested briefly after each block. The order of presentation of adapters was counterbalanced across subjects. The sequences of adapters were presented binaurally through Grason-Stadler TDH-39 headphones at a constant peak listening level of 80 dB SPL. Subjects were instructed to hold their tongues comfortably

[4] The closure durations and VOTs (ms) in the [rətʰi] adapters selected for subjects 1 to 6 were: S1: 102.2, 104.6, S2: 134.0, 82.3, S3: 202.3, 121.5, S4: 105.2, 72.9, S5: 95.5, 80.4, S6: 96.9, 105.9.

against their teeth and not to subvocalize during presentation of the adapters. Immediately after each block of ten adapter repetitions subjects uttered one of the syllables [rəpʰi], [rətʰi] or [rəkʰi] according to a printed randomization. These productions were recorded as in session 1. In total, each subject produced 20 tokens of each of the three syllables in both adapter conditions. [Cooper (1974) recorded 20 productions in each condition; Cooper and Nager (1975) recorded ten]. Periods of devoicing and VOTs were measured in the manner described above.

Results

The results of the experiment are summarized in Tables II and III. Table II shows means and standard deviations of VOTs produced by each subject in each syllable for (a) the pre-adaptation condition in session; (b) after adaptation with [i]; and (c) after adaptation with [rətʰi]. Table III shows analogous measures for the closure intervals.

Table II Voice onset times

Subject	Test syllable	Pre-adaptation Mean	s.d.	[i]-adaptation Mean	s.d.	[rətʰi]-adaptation Mean	s.d.	Z-score	P
1	rəpʰi	74.44	10.06	75.51	11.29	75.94	13.75	+ 0.11	
	rətʰi	103.80	7.95	107.40	9.80	103.80	7.38	− 1.32	
	rəkʰi	104.40	8.36	102.40	11.51	99.54	6.72	− 0.96	
2	rəpʰi	82.88	12.06	62.71	12.43	62.15	14.53	− 0.13	
	rətʰi	83.93	8.60	79.81	12.10	75.15	7.91	− 1.44	
	rəkʰi	113.10	12.77	101.50	8.36	106.90	16.71	+ 1.29	
3	rəpʰi	110.50	19.36	92.59	8.45	85.95	11.81	− 2.12	< 0.025
	rətʰi	123.00	18.99	110.50	11.38	106.70	8.07	− 1.21	
	rəkʰi	144.20	15.66	125.70	7.73	119.10	10.43	− 2.29	< 0.025
4	rəpʰi	57.58	10.67	75.86	10.76	80.65	13.59	+ 1.24	
	rətʰi	72.88	10.07	82.48	5.30	78.86	7.84	− 1.71	
	rəkʰi	88.33	9.82	101.20	9.42	98.83	10.45	− 0.75	
5	rəpʰi	56.44	7.32	54.44	10.03	55.44	11.18	+ 0.30	
	rətʰi	80.18	9.41	72.64	6.50	72.71	6.87	+ 0.03	
	rəkʰi	78.75	11.27	72.76	5.71	78.89	10.27	+ 2.33	< 0.010
6	rəpʰi	98.95	8.96	79.91	10.28	86.46	9.41	+ 2.10	< 0.025
	rətʰi	106.20	9.24	95.83	8.54	99.33	11.43	+ 1.10	
	rəkʰi	117.60	8.58	110.30	10.99	112.40	9.72	+ 0.64	
Means	rəpʰi	80.13	11.41	73.50	10.54	74.43	12.27		
	rətʰi	95.00	10.71	91.44	8.94	89.42	8.25		
	rəkʰi	107.70	11.00	102.30	8.95	102.60	10.72		

Means and standard deviations of VOTs in ms for each of six subjects in the pre-adaptation condition (session 1) and the [i]- and [rətʰi]-adaptation conditions (session 2). *Z*-scores and their probabilities of occurrence are shown for the comparison between the two means from session 2.

The mean VOTs were examined in an analysis of variance with the factors Subjects (6) × Conditions (3) × Test Syllables (3). The effect of conditions was not significant, $F(2, 10) = 1.256; P < 0.2$, indicating that there was no systematic perceptuo-motor adaptation effect. Overall, and for five of the six subjects, VOTs in [rəpʰi] were slightly longer after adaptation with [rətʰi] compared to adaptation with [i]. VOTs in [rəkʰi] were also slightly longer;

Table III Periods of devoicing

Subject	Test syllable	Pre-adaptation Mean	s.d.	[i]-adaptation Mean	s.d.	[rət^{h}i]-adaptation Mean	s.d.	Z-score	P
	rəp^{h}i	130.00	8.41	93.75	7.95	95.97	10.75	+0.74	
1	rət^{h}i	109.70	9.22	74.96	8.92	70.80	10.28	−1.37	
	rək^{h}i	105.60	5.95	89.30	10.13	89.66	9.98	+0.11	
	rəp^{h}i	129.90	10.37	100.40	9.84	108.00	12.17	+2.17	<0.025
2	rət^{h}i	136.50	8.39	105.80	15.04	123.00	14.63	+3.67	<0.001
	rək^{h}i	103.40	14.42	77.33	19.88	86.59	14.82	+1.67	<0.050
	rəp^{h}i	229.50	30.29	158.80	13.13	151.80	17.28	−1.44	
3	rət^{h}i	239.30	26.62	166.40	20.36	161.10	14.54	−0.95	
	rək^{h}i	216.50	31.50	141.40	17.56	127.30	13.60	−2.78	<0.010
	rəp^{h}i	94.95	11.09	95.68	13.38	96.70	7.13	+0.30	
4	rət^{h}i	93.43	13.01	104.50	10.14	104.20	6.56	−0.11	
	rək^{h}i	84.20	13.05	87.33	8.75	92.15	7.67	+1.85	<0.050
	rəp^{h}i	94.05	12.23	73.51	9.00	79.12	10.56	+1.81	<0.050
5	rət^{h}i	84.88	15.76	62.73	8.50	69.90	11.51	+2.24	<0.025
	rək^{h}i	87.03	9.65	65.50	7.53	69.70	9.27	+1.57	
	rəp^{h}i	125.70	17.71	120.20	12.11	118.60	11.01	−0.44	
6	rət^{h}i	130.70	20.69	123.90	13.90	130.60	13.05	+1.57	
	rək^{h}i	126.80	16.24	118.00	12.00	123.10	14.18	+1.23	
	rəp^{h}i	134.00	15.01	107.10	10.90	108.40	11.48		
Means	rət^{h}i	131.40	15.62	106.40	12.81	109.90	11.76		
	rək^{h}i	120.60	15.14	96.48	12.64	98.08	11.59		

Means and standard deviations of periods of devoicing in ms for each of six subjects in the pre-adaptation condition (session 1) and the [i]- and [rət^{h}i]-adaptation conditions (session 2). *Z*-scores and their probabilities of occurrence are shown for the comparison between the two means from session 2.

three subjects showed this tendency. Only VOTs in [rət^{h}i] were reduced overall by perceptuo-motor adaptation, but only in the productions of three of the subjects. Thus, of the eighteen possible comparisons between effects of [i] and [rət^{h}i] adaptation eight are in the predicted direction while ten are in the opposite direction. One subject (S3) produced the expected direction of change for all three syllables, while two others (S5 and S6) produced the reversed effect for all three syllables. *Z*-scores were computed for each pair of means between the [i] and [rət^{h}i] adaptation conditions. Four of these are sufficiently large to have occurred by chance with a probability less than 0.05; two (for S3 with production of [rəp^{h}i] and [rək^{h}i]) indicate that a significant reduction in VOT occurred, while two (for S5 with [rək^{h}i] and for S6 with [rəp^{h}i]) indicate that significant lengthening in VOT occurred.

Three other analyses of variance were carried out to examine the differences among standard deviations of the VOTs and among the means and the standard deviations of the closure intervals. A significant effect of condition was found for the mean closure intervals, $F(2, 10) = 4.15$; $P < 0.05$, indicating that significantly longer closures were produced in the preadaptation condition. This is consistent with a non-significant tendency for longer VOTs to have occurred in that session, and suggests that subjects spoke more quickly in the adaptation conditions. The standard deviations of neither the VOTs nor the closure intervals differed systematically between the three conditions.