Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 227–230
(S. Karger, Basel/New York 1965).

The Perceptual Value of Sibilant Transitions

By Edward Carney, Manchester

Spectrograms of clusters made up of voiceless sibilant + voiceless stop show clearly marked shifts in the latter part of the noise spectrum which seem analogous to the formant inflections of vowels in such a position; they would seem therefore to have potential perceptual value. Sporadic tests in the classroom showed that the perception of an unreleased post-sibilant voiceless stop fluctuated considerably. It seemed therefore worthwhile to test this perceptibility on a good statistical basis. A series of perception tests was devised for this purpose, using tape-cutting, to show the extent to which the following stop might be predicted from the allophonic variation in the sibilant, and to show the contextual variations in this probability.

Material

A series of nonsense words (English non-words) was constructed for the purpose of the test in which the six consonant groups [sp, st, sk, ʃp, ʃt, ʃk] were placed after and varied with the five short vowels [ɪ, ɛ, a, ɒ, ʊ], thus providing 30 different basic combinations. An initial [n] and a final [ə] were added to form the "words" (e.g. naspə, nɔʃkə, nʊstə, etc.). The 30 words were recorded by the author 6 times each in normal, fairly quick, unexaggerated speech, with pauses between words and no catenation features. The pronunciation was standard English except for [a], where a fully open vowel was used. The vowels were chosen to provide a range of allophonic variation in the sibilants. The [n] and the [ə] were added to provide phonetic bulk. Nonsense words were used so as to avoid semantic loading and to ensure statistical symmetry.

Cuts were made during the hold phase of the plosive without including any part of the release and without encroaching on the sibilant. Jockeying the tape past the playback head was too un-

reliable a method of locating the cut. Instead, preliminary cuts, located by jockeying, were made to fall within the [ə] just after the release, and leader tape inserted. Oscillograms, intensity displays and spectrograms were then made and the place of the final cut located by linear measurement. The resultant monosyllables consisted of [n] + V + [s/ʃ], with the sibilant terminated by the closure phase of the "unreleased" stop.

Testing

Each of the 180 resultant segments was then inserted on a loop and re-recorded on testing tapes four times in succession at 4 second intervals and with 10 seconds between items. The items were thoroughly randomised in the process. There were four separate test tapes, each with 45 items, since this number could be heard at a sitting without fatigue. These four tests were presented to two groups of Swedish listeners: a) a group of 45 language students at the University of Lund (= "Gen. Gp.") and b) a group of 7 qualified and experienced teachers of the deaf from Dövskolan, Lund, plus one postgraduate student of phonetics (= "T.D. Gp."). The phonological restrictions on cluster types in Swedish are not such as to affect the test by linguistic prejudice or interference. The Gen. Gp. heard each test tape twice (in several small sub-groups) – total 16,200 responses; the T.D. Gp. heard each test list three times – total 4,320 responses. Tests were given at weekly intervals. Playback was via a single good loudspeaker in ordinary classrooms.

Since the material was strictly symmetrical and non-meaningful, there was no risk of influencing judgements by explaining the purpose of the tests, and so this was done. Listeners were asked to "predict" the following "unreleased" plosive. They had in effect to identify one of three allophonic types and a choice was compulsory. Absolutely no information about performance results was given until the series was complete.

Results

With three variables one would expect a chance distribution of 33 %. The overall percentage of correct responses was 59.99 % (Gen. Gp. 58.63 %; T.D. Gp. 65.09 %) and is clearly significant. The distribution of both correct and incorrect responses in table I is remarkably even, but table II shows that there is an underlying skewed distribution, with a tendency to favour [t] in –s– contexts, largely at

Table I

Confusion Matrices of Overall Distribution

Gen. Gp.				*T.D. Gp.*			
%	p	t	k	%	p	t	k
p	*58*	23	19	p	*59*	25	16
t	22	*58*	20	t	20	*70*	10
k	24	16	*60*	k	19	15	*66*
	35	32	33		33	37	30

Table II

Confusion Matrices Showing Distribution in –s– and –ʃ– Contexts

Gen. Gp.	*–s– contexts*			*T.D. Gp.*	*–s– contexts*		
%	p	t	k	%	p	t	k
p	*55*	32	13	p	*61*	34	5
t	17	*71*	12	t	19	*77*	4
k	24	18	*58*	k	20	17	*63*
	32	40	28		34	42	24
Gen. Gp.	*–ʃ– contexts*			*T.D. Gp.*	*–ʃ– contexts*		
%	p	t	k	%	p	t	k
p	*61*	14	25	p	*57*	17	26
t	27	*46*	27	t	21	*64*	15
k	25	14	*61*	k	18	13	*69*
	38	24	38		32	31	37

the expense of [k]. The Gen. Gp. shows a corresponding disclination to hear [t] in –ʃ– contexts, though this is not so in the T.D. Gp. This can best be seen in the bottom line of the matrices, which shows the extent to which all responses, correct and incorrect, vary from a 33 % norm. This unevenness may have some connection, though not an immediate one, with the known fact that an abrupt tape-cut made during the course of a fricative gives the auditory impression of the homorganic stop.

The contextual variations in correct prediction do not always form an explicable pattern, but one pattern to emerge is that in –s– contexts the closer the front vowel, the more efficient the transitions.

Table III

Distribution (in %) of Correct Responses in Front Vowel + –s– Contexts

	Gen. (T.D.)		Gen. (T.D.)		Gen. (T.D.)
nɪsp	75 (84)	nɛsp	57 (67)	nasp	48 (56)
nɪst	77 (80)	nɛst	68 (67)	nast	57 (66)
nɪsk	72 (76)	nɛsk	69 (81)	nask	59 (60)

Perhaps the chief interest of the results of the tests is to show that in calculating the perceptual efficiency of a given factor, one should give full and symmetrical allophonic coverage in the material.

The hiss transitions probably play a minor part in the perception of sibilant + stop clusters and may well prove necessary in an adequate synthesis.

Author's address: Mr. Edward Carney, Department of General Linguistics, University of Manchester, *Manchester 13* (England).

Discussion

Penchoen (Paris): Question: Did you try to put in control items without "unexploded stops" at the end? It would be interesting to see if the percentages corresponded with the overall distribution responses of [p, t, k].

Answer *Carney:* The use of a control group in this particular experiment would be difficult. What would the control material consist of? Presumably of sibilants without p–t–k transitions; but the "plain" hiss must terminate somehow. A vertical cut without attenuation would be a quasi-transition to [t]. On the other hand, the final sibilant of an independent syllable would differ markedly in length and intensity from the final sibilants in the segments. A vertical cut could be made with imposed attenuation so as to produce a syllable equivalent in length and other respects to the segments, but its validity as control material would be questionable. Moreover, I doubt the basic validity of giving listeners an unreal choice of this kind. Presumably the result would be a chance distribution of 33%, or some degree of divergence from this. Given either result, I would not know how to interpret it and I do not see how it would constitute a control basis for this experiment.