Proceedings of The Institute of Acoustics 1A2

LIPREADING WITH FUNDAMENTAL FREQUENCY INFORMATION

S.M. ROSEN[1], B.C.J. MOORE[2] AND A.J. FOURCIN[1]

[1]DEPT. PHONETICS AND LINGUISTICS, UNIVERSITY COLLEGE LONDON

[2]DEPT. PSYCHOLOGY, CAMBRIDGE UNIVERSITY

I. Introduction: We have, over the last few years, been developing a single channel extra-cochlear implant for the totally deaf (Fourcin, et al:1979). The signal used for electrical stimulation is designed to be matched both to the deaf lipreader's needs and to his new restricted auditory ability. The basis of our stimulation is the fundamental frequency component of speech, which is able to evoke a sense of pitch. Psychophysical data gathered from our deaf patients make it possible to simulate in a normal hearer the auditory sensations elicited by such electrical stimulation in the deaf. We have found that the ability of normals to discriminate frequency changes in pulse trains or sinusoids is much superior to that of the electrically stimulated implantee. In consequence, we artificially worsen the discriminability of the pitch changes by presenting the pitch contour in the form of high-pass filtered pulse trains or amplitude-modulated noise.

II. General Methods: All tests, except connected discourse tracking, are recorded on a Philip's video cassette recorder with two audio tracks. On the first is recorded the speech signal from a high quality microphone while on the second is the output from a laryngograph (Fourcin: 1974) which by sensing impedance changes acroos the neck, gives a direct indication of vocal fold vibration and thus fundamental frequency. For testing purposes, the tapes are copied with the laryngograph vocal fold closure signal replaced by 2 mS pulses triggered from the original laryngograph waveform. These pulses, which carry the fundamental frequency, are then used to trigger other equipment for presentation to the subjects.

Three conditions are usually used in our experiments:
1) Lipreading with varying pitch (PV) - While the subject watches the video, the pulses representing the fundamental are used to trigger 20 μs pulses which are then band-pass filtered between 4 and 20 kHz, amplified and presented at a comfortable level over headphones. In one test, we used amplitude-modulated (AM) noise as the auditory signal. The pitch pulses triggered 1 mS pulses which were used to amplitude modulate broad band noise. Finally, to prevent any of the pulses leaking through, the AM noise was high-pass filtered at 2 kHz. In either case, the subjects hear the pitch contour in a degraded form. These conditions simulate the deaf implantee with relatively good frequency discrimination.
2) Lipreading with constant pitch (PC) - A free-running oscillator at 160 Hz is used to trigger the 20 μs pulses which are then filtered and amplified as before. The pulses representing the fundamental are used to gate this signal on and off, such that a sound is heard when voicing occurs and there is silence during periods of vocal fold inactivity. Thus the subject gets information about presence or absence of voicing, but not about pitch changes. This condition simulates the deaf implantee with minimal or no frequency discrimination.
3) Lipreading alone (LA) - This control condition simulates the unaided, totally deaf subject.

III. Results: We divide our tests into three main categories: segmental, suprasegmental and connected discourse. In this summary, we report the results of a single test in each category.

A. Segmental tests: Our main results involve the identification of intervocalic consonants, /aCa/where C = /b,m,p,v,f,d,n,z,s,t,g,k/. Two speakers were recorded, a male and a female. Sessions consisted of 48 trials (4 presentations per stimulus) in a random order. Five different sessions were recorded by each speaker. Two subjects ran each of the 10 test sessions in each of the 3 conditions (LA, PC and PV). Feedback sessions were given through the course of the experiment, although never in the actual testing sessions. The results summed over the two subjects are shown in the 3 confusion matrices. Certain summary statistics are compared for the 3 conditions in Table I. The supplementation of the visual by pitch information, whether constant or varying, has the following effects: 1) Overall performance is much improved. 2) Place errors, perhaps surprisingly, are reduced. 3) Voicing errors are highly reduced. 4) Errors which involve only manner are slightly increased. We have also performed tests with identification of CVC's and words.

TABLE I	LA	PC	PV
Percent correct	43.9	71.9	73.6
Place Errors	90	46	60
Voicing Errors	372	80	67
Manner Errors (not voicing or place)	116	148	134

B. Suprasegmental tests: We have tested the ability of our subjects to name the stressed word in simple three word sentences when the stress is placed on the three different words. There were two speakers, a male and a female, each of whom recorded one session (consisting of 30 presentations) with a sentence with continuous voicing ("They run well.") and one with gaps in voicing between each of the words ("Knives stay sharp."). Four subjects were tested on each of the four sessions in each of three conditions. Averaged over subjects, speakers and sentences, subjects scored 58.7% correct in condition LA, 97.7% correct in condition PV (with pitch signalled by amplitude-modulated noise) and 97.5% correct in condition PV with sound alone (no video). Similar results have been reported by Risberg & Agelfors (1978). We have also performed tests where subjects judge whether a short phrase or sentence is a question or a statement.

C. Connected Discourse tests: We have made use of a technique described by De Filippo & Scott (1978) known as connected discourse tracking (CDT). In CDT, one person, designated the receiver must repeat back verbatim what another person, the talker, has said. The talker hears the receiver as normal, but the receiver is constrained by performing in one of the three conditions, LA, PC, or PV. The talker wears laryngograph electrodes, and pulses to trigger equipment are obtained directly from the laryngograph processor. Acoustic isolation between talker and receiver is achieved by performing the task across two thick pieces of glass between two sound-proof rooms with broad-band masking noise introduced into the receiver's room. A novel was used as the text and performance is measured in words per minute. We have worked in session of 5-7 minutes and the results for one talker-receiver pair are shown in Figure 1. The lines are the result of three-point smoothing of the day-to-day data. For one receiver immediate improvement with fundamental information is noted, while the other receiver takes some days before a gain is shown. Receiver SR over the course of the experiment shows, relative to LA, a 107% improvement in condition PV and 25% in PC. Tactile aids in the same task show at most a 30% improvement (De Filippo & Scott: 1978).

LIPREADING WITH FUNDAMENTAL FREQUENCY INFORMATION

LIPREADING ALONE

RESPONSE

STIMULUS	B	M	P	V	F	D	N	Z	S	T	G	K		SUM
B	33	21	26										B	80
M	28	32	20										M	80
P	33	9	36		1	1							P	80
V				50	30								V	80
F				23	56	1							F	80
D			1			25	12	11	9	16	4	2	D	80
N						12	38		1	6	14	9	N	80
Z				1		3	1	28	25	20	1	1	Z	80
S						7		16	37	16	1	3	S	80
T						17	4	17	12	29		1	T	80
G						4	19				36	19	G	78
K						6	19			1	33	20	K	79
SUM	94	62	83	74	87	76	93	72	84	88	89	55		

LIPREADING WITH CONSTANT PITCH

RESPONSE

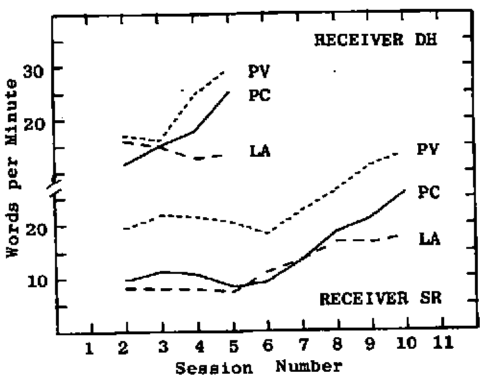
STIMULUS	B	M	P	V	F	D	N	Z	S	T	G	K		SUM
B	57	23											B	80
M	7	73											M	80
P	14		65									1	P	80
V				66	13								V	79
F				3	77								F	80
D						42	10	18	2	7	1		D	80
N						1	57	18			4		N	80
Z					1	20		40	14	3	1	1	Z	80
S						1			48	31			S	80
T						6			20	46	2	6	T	80
G						18	6				53	3	G	80
K						1			1	3	9	66	K	80
SUM	78	96	65	69	91	89	73	76	85	90	70	77		

LIPREADING WITH FUNDAMENTAL FREQUENCY INFORMATION

LIPREADING WITH VARYING PITCH

RESPONSE

STIMULUS	B	M	P	V	F	D	N	Z	S	T	G	K		SUM
B	61	15	2						1		1		B	80
M	6	73				1							M	80
P	8	1	71										P	80
V				62	16	1							V	79
F					80								F	80
D						49	10	12	6	2	1		D	80
N							57	14			6	3	N	80
Z				1		20	4	46	8	1			Z	80
S					1			1	46	31		1	S	80
T						6			22	47	1	4	T	80
G		1			1	19	9	1	1		46	2	G	80
K									1	5	7	67	K	80
SUM	75	90	73	63	98	96	80	74	85	86	62	77		



ACKNOWLEDGEMENTS

This work is done in conjunction with E.E. Douek and G.P. Clarke of Guy's Hospital and J.R. Walliker of UCL and is supported by the MRC, the Sir Halley Stewart Trust and the RNID. Thanks to J.Fisher, A.Parker, M.Ashby, D.Howard and especially E.Abberton.

REFERENCES

* De Filippo, C.L. and Scott, B.L. (1978) "A method for training and evaluating the reception of on-going speech", J. Acoust. Soc. Am. 63, 1186-1192.
* Fourcin, A.J. (1974) "Laryngographic examination of vocal fold vibration", 315-333 in Wyke B. (ed) Ventilatory and Phonatory Control Systems. London: Oxford U. Press.
* Fourcin, A.J., Rosen, S.M., Moore, B.C.J., Douek, E.E., Clarke, G.P. Dodson, H. & Bannister, L.H. (1979) External electrical stimulation of the cochlea: clinical, psychophysical, speech-perceptual and histological findings", The British Journal of Audiology, 13, 85-107.
* Risberg, A. and Agelfors, E. (1978) "On the identification of intonation contours by hearing impaired listeners", STL-QPSR 2-3/1978, 51-61.