INCORPORATION OF THE FORTIS-LENIS FEATURE IN A QUASIARTICULATORY SYSTEM OF TACTILE SPEECH SYNTHESIS BY ADDING TEMPORAL VARIATIONS

HANS GEORG PIROTH

Institut für Phonetik und Sprachliche Kommunikation
der Universität München
Schellingstr. 3, 8000 München 40, F.R.G.

ABSTRACT

A method for electrocutaneous speech synthesis was developed using pulse train sequences with variable intervals that are delivered to 16 electrode pairs along the forearm. The coding was 'quasiarticalatory' in that places of articulation (front - back, high - low)' were mapped quasi-isomorphically to the forearm (distal - proximal, dorsal - volar).

By varying the repetition rate of the pulse bursts (faster - slower) a tactile fortis-lenis equivalent was incorporated and, additionally, a plosive-fricative distinction was defined. So the inventory of tactile consonants was expanded to cover the whole range of the obstruent system of a language such as German. Exp. I uses tactile fricative-vowel equivalents, Exp. II plosive-vowel equivalents to test the learnability of such patterns.

INTRODUCTION

There is a long history of experimental investigations in the field of tactile speech transmission (e.g. [2,3,4,8]). Most of them used mechanical or electrical stimulation devices to transform the acoustic parameters of the speech signal into tactile patterns. The fact that most of these systems failed to reach the level of practical use, demands general reconsideration. According to our point of view in all these investigations the role of articulatory gestures for speech perception seems to be underestimated. Since tactile and proprioceptive reafferent control is present during the period of language acquisition, one may assume that normal speech perception is at least partially governed by the perception of articulatory guestures [1,9]. So it may be argued that a transformation of articulatory rather than acoustic information to the skin would provide a better opportunity to develop a successful tactile speech transmission system [6,7,10].

In its final state a quasiarticulatory system for tactile speech transmission would consist of four components:

(1) Registration of a speaker's acoustic signal.
(2) Analysis of the articulatory parameters from the speech wave.
(3) Transformation of the articulatory information into quasiarticulatory coded tactile patterns.
(4) Presentation of tactile patterns.

The investigation reported here is concerned with the development of the quasiarticulatory coding of tactile stimulus patterns. To yield an approximately geometric mapping of the places of articulation the forearm was selected as the tactile stimulation area. The experiments were executed with the 'System for Electrocutaneous Stimulation' (SEHR-2) presenting current-controlled bipolar pulse train sequences of the basic form shown in Piroth/Tillmann 1984, Fig. 1 [5]. The sixteen channels of the stimulation device (cf. Tillmann/Piroth 1986 [10]) were connected with 16 pairs of round gilded brass electrodes (9 mm in diameter). The smallest distance between the electrodes of a pair was 1 mm.

The electrode arrangement and the order of successive stimulations was defined according to a set of basic criteria for the coding method. First, complete tactile patterns are syllable analogues, i.e. each syllable is in one-to-one correspondence to a complete pattern. Second, a complete pattern is composed of partial patterns representing the consonant and vowel phonemes. In general, vowel patterns move longitudinally along the arm, consonantal patterns circumferentially. (Fig. 1 shows the arrangements of electrodes as well as the stimulation area of the central vowel /ə/.) Third, places of articulation are mapped to the place of tactile stimulation so

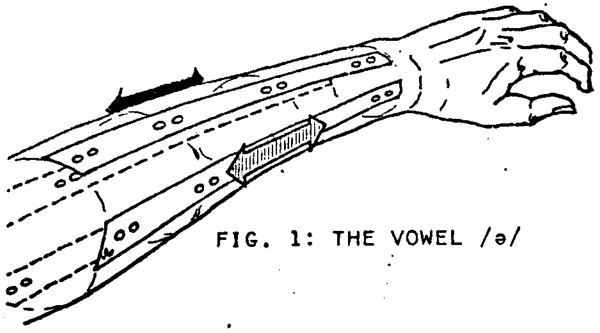

FIG. 1: THE VOWEL /ə/

that front articulations (of vowels and consonants as well) correspond to distal patterns near the wrist and back articulations to proximal patterns near the elbow. High vowels are mapped to the dorsal side. low vowels to the volar side of the forearm. Intermediate places of articulation are coded by stimulating the intermediate tactile areas. Fourth. as the starting point of the circumferentially moving consonant patterns depends on the stimulation area of the preceding or following vowel. an rudimentary form of `coarticulation' is implemented in the coding method. Fifth. in the present experiments the temporal duration of the tactile syllable equivalents correspond to those of extremely slow and very explicitly uttered natural syllables. Nevertheless. patterns are constructed in a way that overall duration can be shortened by omitting pulses from the pulse train sequences without hereby altering the phenomenal `gestalt' of the patterns. Former investigations have shown that vowel patterns are easily identified even by untrained subjects and that consonantal places of articulation -although identification is not as good- are recognized well above chance level without training. The following experiments include the fortis-lenis- and the plosive-fricative distinction into the system of consonantal patterns to yield a construction method for the complete system of obstruents, and they use a learning paradigm to improve the identification results by training.

EXPERIMENT I

According to the basic criteria a system of tactile fricatives and the vowel /ə/ was constructed and combined to form syllables. Tactile /fə:/, /sə:/, /ʃə:/. /çə:/, /və:/, /zə:/, /ʒə:/, /jə:/ and their VC-equivalents were presented in a learning test to reveal whether identification of CV-equivalents can be improved by learning. A succeeding control test was run to show whether a transfer of learned skills enhances the identification of the VC-patterns.

EXPERIMENT II

In the same way a plosive-vowel system was used consisting of tactile /pə:/, /tə:/, /kə:/, and /bə:/, /də:/, /gə:/.

Table 1
Fricative Vowel System

	Number of Taps	Tap-duration (ms)	ITI (ms)	Overall duration (ms)
V	8	5.2	20	201.6
FF	8	5.2	15	161.6
LF1	4	5.2	35	160.8
LF2	8	5.2	5/30	181.6

V: Vowel, FF: Fortisfricative, LF1: Lenisfricative (1 ring: simple pattern, LF2: 2 rings: complex pattern).

GENERAL METHOD

Stimuli.
Pulse trains (`taps') of three pulses having the form described above with a constant pulse width of 200 µs. a variable amplitude, a constant interpulse-onset interval of 2.5 ms and an overall duration of 5.2 ms were used as basic stimuli. They were arranged to sequences in which the places of stimulation are changed according to the basic criteria cited above. So. the tactile syllable equivalents consisting of fricative patterns and a /ə/-pattern were constructed. Number of taps, inter-tap intervals (ITI), tap-duration and overall duration of the patterns are given by Tab. 1 for each type of stimulus. The local shifts along the first or first and second electrode rings (i.e. the distal ones) in the fricative patterns /f/ and /v/ are shown in Fig. 2, of /s/ and /z/ in Fig. 3.

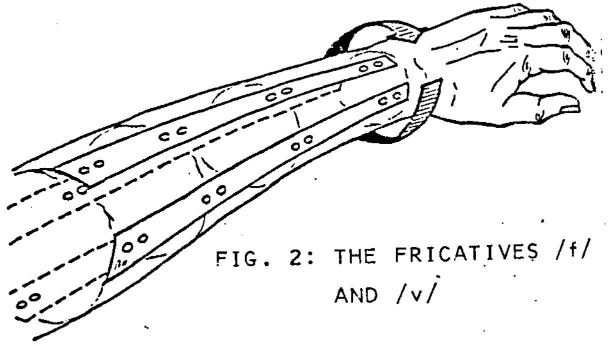
FIG. 2: THE FRICATIVES /f/ AND /v/

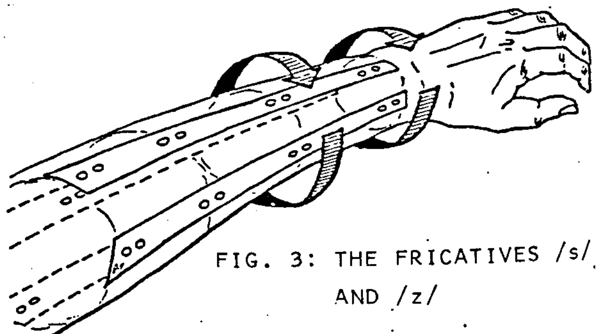
FIG. 3: THE FRICATIVES /s/ AND /z/

/ʃ/ and /ʒ/ resemble /f/ in being simple patterns consisting of one circumferential tap sequence only, but they are constructed as surrounding the second electrode ring instead of the first. /ç/ and /j/ like /s/ and /z/ are complex patterns (two rings) which are presented quasi-simultaneously (i.e. with the proximal ring having a delay of 5 ms) at the second and third electrode rings. As shown in Tab. 1, the fortis-lenis difference is encoded in the inter-tap interval of the sequences: fast moving patterns are fortis. slowly moving ones [6] lenis. (Preliminary investigations had shown that a difference in ITI of 20 ms is perceivable in similar patterns presented without a context.) To keep overall duration constant. the number of taps was halved in the case of /v/ and /ʒ/. Since /z/ and /j/ are presented by stimulation of 8 loci this was not possible. So /z/ and /j/ are 20 ms longer in overall duration. The neutral vowel /ə/ is transformed into an 8-tap pattern moving along the mid pairs of radial and ulnar electrode rows (Fig. 1).
Plosives are built as patterns sweeping between neighbouring electrode pairs as shown in Fig. 4 which represents /p/ and /b/. Analoguously, /t/ and /d/ surround the second ring, /k/ and /g/ the third.

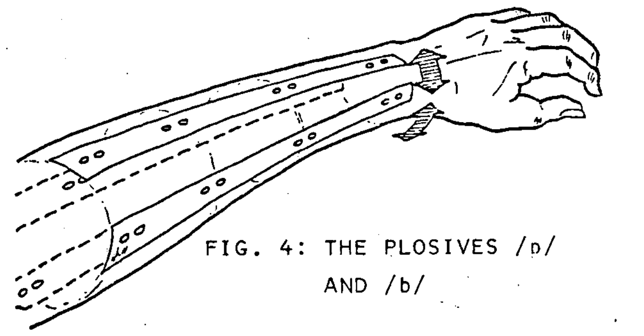
FIG. 4: THE PLOSIVES /p/ AND /b/

Regarding the basic criteria information of the starting point of the vowel pattern is preserved by starting the circumferent pattern at the same electrode row where the following vowel pattern starts. The velocity of the pattern (5 ms/ITI) is clearly below the threshold for discrimination of successive taps and is not used to carry the fortis-lenis information. Instead, this feature is encoded in the neighbouring half of the vowel-pattern as listed in Tab. 2.

Subjects and Procedure.
Four unexperienced Ss participated in Exps. I and II. All Ss were tested singly and were informed about the details of the learning test series and the coding method.
1. Intensity adjustment. Each test session started with a calibration procedure. Ss were asked to adjust subjective intensity to a mid value between absolute threshold and unpleasentness for each place of stimulation. The resulting values were taken as impulse amplitude values in the immediately following test runs.
2. Presentation of the patterns. The inventory of the 8 (or 6) CV-patterns was presented five times in a systematic order to the S as following: for each pattern a phonological transcription was presented via terminal followed by the tactile pattern corresponding to this syllable. After an interval of 4 s the transcription of the next syllable was presented.
3. Feedback tests. For a single FB-test the 8 CV-patterns were presented 6 times in Exp. 1 and the 6 CV-patterns in Exp. 2 were presented 8 times in completely randomized order to yield 48 presentations. By pressing a key on the computer keyboard the S started the presentation of a pattern. After an interval of 1s the S had to name the just presented syllable via keyboard. Then the transcription of the syllable that was presented was given to inform the S whether his answer was correct or not. After the presentation of 5 equal test runs (i.e. 30 or 40 repetitions of each pattern) the test session was finished. The Ss underwent 5 equal test sessions with a pause of 1 or 2 days between each two successive sessions. Finally, in a sixth (control-) session structured in the same way the whole inventory was presented in VC-ordering. Two of the four Ss first underwent Exp. I, the remaining two Ss first Exp. II.

Table 2
Plosive vowel System

	Number of Taps	Tap-duration (ms)	ITI (ms)	Overall duration (ms)
FP	16	5.2	5	163.2
LP	16	5.2	5	163.2
FP+V	16	5.2	8x15 +8x25	403.2
LP+V	12	5.2	4x35 +8x25	402.4

V: Vowel, FP: Fortisplosive (burst).
LP: Lenisplosive (burst).

RESULTS AND DISCUSSION

Fig. 5a presents the average identification rate for all subjects in Exp. I. Fig. 5b gives the computed results that show the recognition of the fortis-lenis feature. (Identification of a fortis-pattern was assumed to be correct when the S after presentation of a fortis-pattern answers with a fortis consonant.)

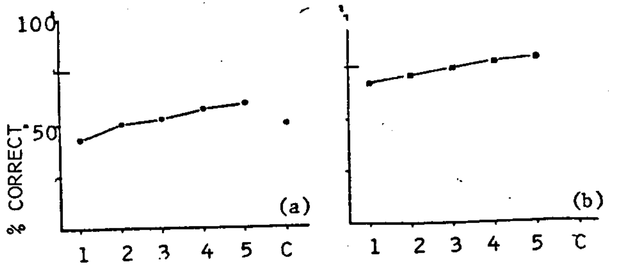

Fig. 5: Results of Exp. I (a) Syllable
(b) Fortis-Lenis Identification

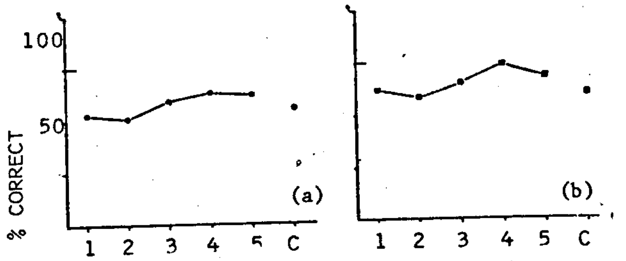

Fig. 6: Results of Exp. II (a) Syllable
(b) Fortis-Lenis Identification

Figs. 6a and b present the results of Exp. II in the same way. To evaluate the effects after a transformation of the dependent variable by

$$y = \text{arc sin } (x/100)^{1/2}$$

a one-factrorial univariate analysis of variance was calculated, first as a trend analysis of sessions 1 to 5 by the method of orthogonal polynomials, then the analysis was expanded to 6 sessions to determine the relevant a priori contrasts. Since the analysis was repeated with fortis/lenis results the level of significance $\alpha = 0.05$ was lowered to α^* by $\alpha^* = 1-(1-\alpha)^{1/2}$. A highly significant variation over the 5 CV-sessions was found in all cases yielding a linear trend in Exp. I and a cubic one in Exp. II (Tab. 3). For the analysis of contrasts the comparisons of 1st and 5th, 1st and 6th and 5th and 6th session were chosen. According to Tab. 3 the contrasts between the 1st and the 5th CV-session are always significant and show a consistent learning effect. But the results indicate that there is no transfer of learning from the CV-series to the VC-session: in all cases the contrast between the first CV-session and the VC-session is not significant. With a simple exception, the results of the 5th CV-session and the VC-session differ significantly. This may be due to the fact that only one control-session was run. A series of experiments is in preparation to show whether a transfer of learning is possible if the S has to manage more trials in the altered condition.

Table 3
Results of Exps. I and II

Trend analysis by orthogonal polynomials (sessions 1-5):

Exp. I Syllables	F(4.95)= 4.421	p<0.005**
Linear trend	F(1.95)=17.076	p<0.005**
Exp. I Features	F(4.95)= 6.187	p<0.005**
Linear trend	F(1.95)=23.982	p<0.005**
Exp. II Syllables	F(4.95)= 9.869	p<0.005**
Cubic trend	F(1,95)=10.288	p<0.005**
Exp. II Features	F(4.95)=10.366	p<0.005**
Cubic trend	F(1,95)=18.379	p<0.005**

A priori contrasts (sessions 1-6):

Exp. I Syllables				
1 vs. 6:	t=-0.725	df=114	p=0.470	n.s.
1 vs. 5:	t=-3.829	df=114	p=0.000	**
5 vs. 6:	t= 3.104	df=114	p=0.002	**
Exp. I Features				
1 vs. 6:	t=-0.762	df=114	p=0.447	n.s.
1 vs. 5:	t=-4.102	df=114	p=0.000	**
5 vs. 6:	t= 3.440	df=114	p=0.001	**
Exp. II Syllables				
1 vs. 6:	t=-1.552	df=114	p=0.124	n.s.
1 vs. 5:	t=-3.253	df=114	p=0.002	**
5 vs. 6:	t= 1.701	df=114	p=0.092	n.s.
Exp. II Features				
1 vs. 6:	t=-0.013	df=114	p=0.989	n.s.
1 vs. 5:	t=-2.418	df=114	p=0.017	*
5 vs. 6:	t= 2.404	df=114	p=0.018	*

Reduced level of significance:
p<0.00501 ** p<0.02532 *

REFERENCES

[1] C.A. Fowler, "An Event Approach to the Study of Speech Perception from a Direct-Realistic Perspective", J. Phon. 14, 1986, 3-28.

[2] R.H. Gault, "An Experiment on Recognition of Speech by Touch", J. Wash. Acad. Sci. 15, 1925, 14.

[3] M.H. Goldstein, R.E. Stark, "Modification of Vocalizations of Preschool Deaf Children by Vibrotactile and Visual Display", J. Acoust. Soc. Am. 59, 1976, 1477-1481.

[4] R. Lindner, "Physiologische Grundlagen zum elektrischen Sprachetasten und ihre Anwendung auf den Taubstummenunterricht". Zeitschr. f. Sinnesphysiologie 67, 1937.

[5] H.G. Piroth, H.G. Tillmann, "On the Possibility of Tactile Categorical Perception", M.P.R. v.d. Broecke, A. Cohen, Proc. 10th ICPhS, Dordrecht 1984, 764-768.

[6] H.G. Piroth, "Elektrokutane Silbenerkennung mit quasiartikulatorisch kodierten komplexen zeitlich-räumlich strukturierten Reizmustern", Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 22, München, 1985.

[7] H.G. Piroth, "Electrocutaneous Syllable Recognition Using Quasiarticulatory Coding of Stimulus Patterns" (Abstr.), J. Acoust. Soc. Am. 79, 1986, S73.

[8] D.W. Sparks et al., "Investigating the MESA (Multipoint Electrotactile Speech Aid): The Transmission of Segmental Features of Speech". J. Acoust. Soc. Am. 63, 1978, 246-257.

[9] H.G. Tillmann, "Phonetik und Phonologie sowie die natur- und geisteswissenschaftliche Erforschung der gesprochenen Sprache", FIKPM 19, 1984, 9-32.

[10] H.G. Tillmann, H.G. Piroth, "An Order Effect in the Discriminability of Pulse Train Sequences" (Abstr.), J. Acoust. Soc. Am. 79, 1986, S73.

ACKNOWLEDGEMENTS

This investigation was supported by the German Research Council (DFG-grant Ti 69/25).
Statistical advice was given by Dr. Alexander Yassouridis, Department of Biostatistics, Max-Planck-Institute for Psychiatry, Munich F.R.G.