ARTICULATION-BASED TACTILE SPEECH FOR THE DEAF: A COMPLETE SET OF TACTILE SEGMENTAL FEATURES FOR GERMAN

H. G. Piroth and H. G. Tillmann

Institut für Phonetik und Sprachliche Kommunikation
der Universität München, Germany

ABSTRACT

This paper presents the definitions for a synthesis of quasiarticulatory tactile speech stimuli and the method of presentation as developed in a German Research Council project concerned with tactile syllable equivalents.

1. INTRODUCTION

Most of the investigations concerned with tactile speech for the deaf use systems that transmit tactile transforms of the acoustic speech wave to the skin. (For an overview see [5]). As opposed to these approaches, the concept of a speech-to-skin transmission system has been proposed by us that fully relies on articulatory information to code tactile speech equivalents [1,6].

Such a system needs to contain two main components. The first has to extract the articulatory information from the speech wave, and the second has to transform it into tactually well-distinguishable patterns. To investigate the general applicability of such a concept our initial experimental research has been limited to the second component.

In several previous investigations proposals for the coding of the articulatory features to construct a tactile system of vowel and obstruent equivalents have been made and modified dependent on the test results (e.g. [2,3]). In preparation of the test stimuli for an investigation of the recognizability of tactually presented words [4] the need for a complete and consistent set of tactile features to code at least the phonemic distinctions of Standard German arose. Only labialization is not included in this version, since lip movements are easily detectable by lip-reading.

2. APPARATUS AND GENERAL METHOD

For the execution of experimental research SEHR ("System for Electrocutaneous Stimulation") has been developed. SEHR enables the computer-controlled synthesis and presentation of electric pulse trains to the skin.

A PDP-11 is connected with a 16-channel stimulus generation device. SEHR produces current-controlled bipolar impulses without a d.c.- component. The impulse forms of the versions SEHR-2 and SEHR-3 are given in Fig. 1. Using version SEHR-2 the pulse repetition rate can be manually adjusted between 100 and 500 pps. The duration of the rectangular part of the impulse is software-controlled and variable between 0 and 500 µs. Impulse amplitude is digitally adjusted to one of 64 steps between 0 and 5 mA. For SEHR-3, also pulse repetition rate is preset via software. Durations are allowed to vary in steps of 32 µs between 0 and 512 µs and intensities in steps of 0.33 mA from 0 to 4.95 mA.

The channel outputs of SEHR are delivered to the subjects via 16 pairs of circular gold-layered electrodes (9 mm in diameter and 1 mm apart from one another). For the experiments on an articulation-based system for tactile speech presentation electrodes are placed as shown in Fig. 2. The PDP-11 is equiped with an AD-converter and a Schmitt-trigger that receive their inputs from a small box with a potentiometer knob and a button. The software package includes several procedures for intensity threshold determination. Thus, during a calibration procedure at the beginning of each test session Ss can adjust impulse intensities separately for each channel by turning the knob and store the desired values by pressing the button. For test presentation, the package contains several modules that take detailed stimulus descriptions (sequence of channels, numbers of impulses, intervals between successive pulse trains) as input as well as the calibration results (pulse amplitudes and durations) and randomization lists for identification or discrimination tests. Some identification test modules allow answering by pressing keys on the computer keyboard and give confusion matrices of the results as output.

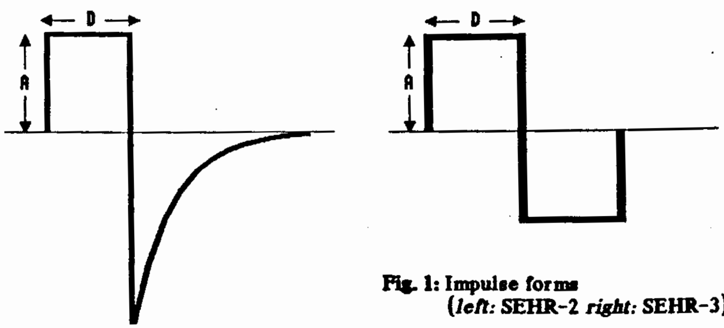


Fig. 1: Impulse forms
(*left:* SEHR-2 *right:* SEHR-3)

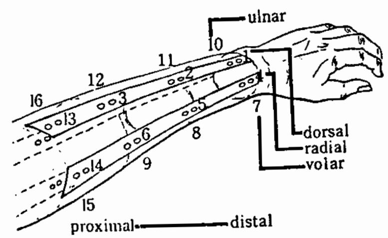


Fig. 2: Electrode arrangement for tactile stimulation

3. THE CODING METHOD

The forearm has been chosen as tactile stimulation area, since it can be interpreted as a mapping of the local relations within the vocal tract. The stimulus patterns are dynamic sequences of pulse trains consisting of three impulses with a duration of 256 µs for each rectangular part and an amplitude corresponding to the subjective mid value between absolute and annoyance thresholds.

Under these conditions, quasiarticulatory patterns are defined on the syllable level: syllables equal in narrow phonetic transcription are represented by equal tactile patterns consisting of subpatterns representing the segmental structure of the syllables. The basic distinction between vowels and consonants is coded by the distinction between longitudinal and circumferent subpatterns (i.e. subpatterns moving along the arm or subpatterns surrounding it).

The coding features of vowels are height (from dorsal to volar) and front/back (distal/proximal). According to this phonetic feature transformation, /i/ is a subpattern oscillating between the two distal electrode pairs on the dorsal side, /u/ between the two proximal ones of the same side, a front /a/ is a subpattern oscillating between the two distal electrodes on the volar side. For a complete coding of vowel height more than two distinctions have to be made: mid vowel height is transformed to a subpattern oscillating both at the ulnar and radial sides of the arm. To create a fourth level, /ε/ moves along the radial and ulnar as well as the volar side of the arm. (For details see Tab. 1.)

For consonants, the front/back distinction is applicable as well, /f/ surrounds the arm at the distal electrode ring, /h/ at the proximal one, others are in between according to their places of articulation. Since more places have to be coded than electrode rings are available, some fricatives are represented as double rings contrarotating at neighbouring electrode rings. The fortis/lenis feature is coded as a difference in inter pulse train intervals. Fortis fricatives have twice as many pulse trains as lenis fricatives and a shorter inter pulse train interval to preserve equal overall durations. Plosive subpatterns, as opposed to fricatives do not form a complete ring, but after start-

Tab. 1: Sequences of Electrode Pairs Stimulated for Subpatterns (Consonants Initial before /a/)

No. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Vowels (IPTI=20)

/i/	1 2 1 2 1 2 1 2
/a/	8 9 8 9 8 9 8 9
/u/	3 13 3 13 3 13 3 13
/e/	4 5 11 10 10 11 5 4
/o/	6 14 16 12 12 16 14 6
/ə/	5 6 12 11 11 12 6 5
/ɛ/	4 7 5 8 11 8 10 7

Fricatives (IPTIfortis=15, IPTIlenis=35)

/f/	1 1 4 4 7 7 10 10
/v/	1 4 7 10
/ʃ/	2 2 5 5 8 8 11 11
/ʒ/	2 5 8 11
/x/	3 3 6 6 9 9 12 12
/h/	13 13 14 14 15 15 16 16

Fricatives (IPTIfortis=7, IPTIlenis=15)

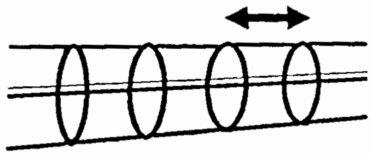
/s/	1 1 8 8 4 4 5 5 7 7 2 2 10 10 11 11
/z/	1 8 4 5 7 2 10 11
/ç/	2 2 9 9 5 5 6 6 8 8 3 3 11 11 12 12
/j/	2 9 5 6 8 3 11 12

Plosives (IPTIfortis=7, IPTIlenis=20)

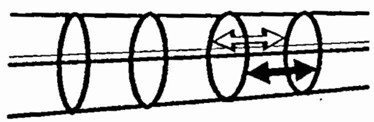
/p/	1 4 4 4 4 4
/b/	1 4 4 4 4 4
/t/	2 5 5 5 5 5
/d/	2 5 5 5 5 5
/k/	3 6 6 6 6 6
/g/	3 6 6 6 6 6
/ʔ/	13 14 14 14 14 14

Duration of a single Pulse Train: 5.2 µs
Interval Between Pulse Trains: IPTI [ms]

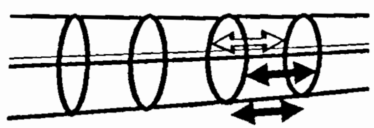
Example: /i/



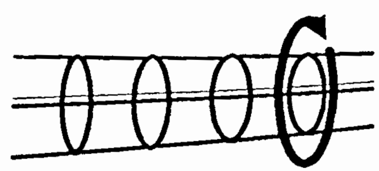
Example: /e/



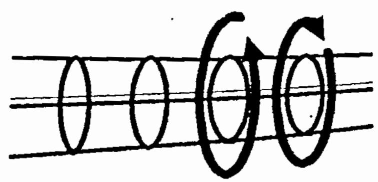
Example: /ɛ/



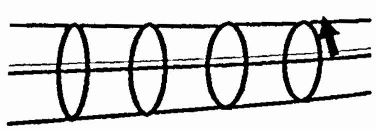
Example: /f/ or /v/



Example: /s/ or /z/



Example: /p/ or /b/



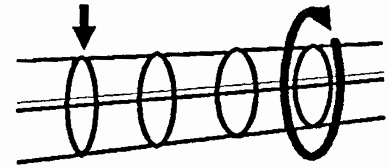
Nasals (IPTI=8)

/m/	1 1 13 13 4 4 13 13 7 7 13 13 10 10
/n/	2 2 13 13 5 5 13 13 8 8 13 13 11 11
/ŋ/	3 3 13 13 6 6 13 13 9 9 13 13 12 12

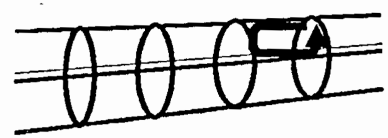
Liquids (IPTI=15)

/r/	1 2 5 4 1 2 5 4
/l/	2 3 6 5 2 3 6 5
/ʀ/	3 13 14 6 3 13 14 6

Example: /m/



Example: /r/



ing at one point of the electrode ring corresponding to their place of articulation, the subsequent pulse trains are statically delivered to the place neighbouring the starting point to simulate their noncontinuous production. Fortis plosives move fast and have a pause added afterwards that resembles the occlusion phase. Lenis plosives move more slowly and lack the pause to keep overall durations of plosives constant.
Nasals are coded like the corresponding lenis fricatives, but with intermediate pulse trains at the dorsal proximal electrode pair to imitate velum behaviour. Liquids are circular movements using the dorsal and radial side electrodes of two neighbouring rings according to their places as described below. To include an analogue of coarticulation consonantal subpatterns start where the preceeding vowel stops or stop where the subsequent vowel starts. Thus, the exact electrode pairs involved in a tactile consonant equivalent change with the context vowel giving additional hints for the recognition of vowels with the effect that the only patterns that are exactly identical are recurring syllable equivalents, not segments.
In this system overall durations of consonants and vowels are constant (except for some subthreshold variations) for experimental purposes, but durational variations of careful and explicit speech could be implemented without changing the main characteristics of the percept.

4. REFERENCES

[1] PIROTH, H.G.(1985), "Elektrokutane Silbenerkennung mit quasi-artikulatorisch kodierten komplexen zeitlich-räumlich strukturierten Reizmustern", *Forschungsberichte d. Instituts f. Phonetik u. Sprachl. Komm. Univ. München 22.*

[2] PIROTH, H.G.(1987), "Incorporation of the fortis-lenis feature in a quasiarticulatory system of tactile speech synthesis by adding temporal variations", *Proc. 11th ICPhS, Vol. 1*, 369-372.

[3] PIROTH, H.G. et al.(1989), "Electrotactile encoding and recognition of a 16-obstruent system", *J.Acoust.Soc.Am. 86, Suppl.1*, S95.

[4] RENZELBERG, G.(1990), "Taktile Analoge", *Phil. Diss.*, Munich.

[5] SHERRICK, C.E.(1984), "Basic and applied research on tactile aids for deaf people", *J.Acoust.Soc.Am. 75*, 1325-1342.

[6] TILLMANN, H.G.(1970), "*Technische Kommunikationshilfen für Gehörlose*", Berlin: Marhold.