IPO ANNUAL PROGRESS REPORT

Nr.7 1972

EDITION : D. BOUWHUIS

ADVISER: H. J. A. VAN BECKUM

THE NUMBERS ABOVE REPRESENT THE RELATIVE LUMINANCE LEVELS ON A PHOTOMATRIX WHEN THE LETTERS "IPO" ARE SUPERIMPOSED ON IT. SEE: "PHOTOMATRIX COMPUTER INTERFACE" CONTRIBUTED BY P. VERHAGEN AND M. MALEWIJNSE ON PAGE 115.

INSTITUUT VOOR PERCEPTIE ONDERZOEK
INSULINDELAAN 2 EINDHOVEN
TELEPHONE 040-756605/040-472485
ORGANISATION

supervisory board i.p.o

Ir. K. Kooij (chairman) - Eindhoven
Drs. J.H. Bannier - 's-Gravenhage
Prof. Dr. J.F. Schouten - Eindhoven
Dr. Ir. K. Teer - Eindhoven
Prof. Dr. J. Wemeisfelder - Eindhoven

scientific board i.p.o.

Prof. Dr. H.B.G. Casimir (chairman) - Heeze
Prof. Dr. M.C. Colenbrander - Utrecht
Prof. Dr. J. Drooglever Fortuijn - Groningen
Prof. Dr. N.H. Frijda - Amsterdam
Prof. Dr. L.B.W. Jongkees - Amsterdam
Prof. Dr. J. Koekebakker - Leiden
Prof. Dr. S.L. Kwec - Eindhoven
Prof. Dr. Ir. H. Mol - Amsterdam
Prof. Dr. A.J.B.N. Reichling - Amsterdam
Prof. Dr. K.J. Ritsma - Groningen
Prof. Dr. P.C. Veenstra - Eindhoven
Prof. Dr. A.J.H. Vendrik - Nijmegen
Prof. Dr. Ir. C.J.D.M. Verhagen - Delft
Dr. Ir. P.L. Walraven - Soesterberg
Prof. Dr. P.J. Willems - Tilburg
Prof. Dr. Ir. A. van Wijngaarden - Amsterdam

director i.p.o.

Prof. Dr. C.A.A.J. Greebe - Eindhoven

research associates

D.J.H. Adambaal
J.J. Andriessen
Dr. H. Bouma
Drs. D.G. Bouwhuis
Drs. B.L. Cardozo
Dr. H. Duifhuis, for one year at Massachusetts Institute of Technology U.S.A.
Dr. J.P.M. Eggermont
Ir. F.L. Engel
J. 't Hart
Drs. A.P.V. van Katwijk
Ir. W.G. Koster
P.F. Leopold
G.J.J. Moonen
H.F. Muller
Dr. F.L. van Nes
Dr. S.G. Nootbooom
I.H. Slis
Ir. J.A.J. Roufs
J.C. Valbracht
Ir. L.L.M. Vogten
J. Vredenbrocht
Ir. L.F. Willems

research staff

M.A. Alewijnse
H.J. Bleileven
P.B.W. Boonstra
Th.M. Bos
J.J.A.M. van Dun
Th.A. de Jong
C.A. Lammers

IPO annual progress report 7 1972
G.H. van Leeuwen
H.J. Meulenbrugge
J.J.M. Neelen
A.C. van Nes
W.H. Noordermeer
A.L.M. van Rens
R. van Schuur
H. de Vries

secretaries
Mrs T. Bennebroek-Tesser
Miss L.J. Clignet
Mrs M.J.J. Hougardy-Verburg
Miss J.W. Tieleman

library
Mrs J.M. Hoogervorst
Mrs J.P.M. Swinkels-Kool

workshop
M.A. van den Ban
C.G. Basten
J.H. Bolkestein
P.A.N. Broekmans
C.Th.P. Godschalx
G.C. Jonathan
H.E.M. Mélotte
J.H. de Vroome
D.J. van der Wees

advisers
Prof. Dr. A. Cohen (University of Utrecht)
Dr. G.J. Fortuin

guest researchers
Mrs J. ten Have-de Labije, 4/9-19/9/1972. Laboratory of Psychophysiology
Jan Swammerdam Institute, Amsterdam.
Miss C.M.J. Hulst (part-time) Occupational Medical Dept. of the Nederlandse Spoorwegen (Netherlands Railways), Utrecht.
L.P.A.S. van Noorden, Netherlands Organisation for the Advancement of Pure Research.
C.J.W. Schiepers, Netherlands Organisation for the Advancement of Pure Research.
Mrs Dr. N.N. Tsavena, 7/3-30/4/1972 Research Institute for Labour Protection and Occupational Diseases, Sofia, Bulgaria.

student assistants
This report or any part thereof may not be reproduced in any form without the written permission of the Institute for Perception Research. Reprints of the separate contributions are available. Illustrations may be reproduced only with explicit mentioning of source; copies will be appreciated.

IV

IPO annual progress report 7 1972
INTRODUCTION

The retirement of Dr. Schouten on August 31st, 1972, should constitute the main theme of this introduction to the seventh issue of the I.P.O. Annual Progress Report. He has led the Institute from its foundation in 1957 until that day. To all those who regularly read this publication it will be clear that what he left to his successor is a well equipped laboratory where a wide variety of subjects ranging from prosodic phenomena in human speech to the ergonomical aspects of oscilloscopy is studied in an ambiance of not only interdisciplinary co-operation but also fertilisation.

It is obvious that various styles are required for doing relevant research in such widely different fields. These can in general not be expected to ensue from the mind of one single person as is, of course, also to be seen from the striking differences in personality displayed by the various people who work on all these subjects in the I.P.O. The one and outstanding exception to this has always been Dr. Schouten who continuously showed an extremely keen and very often creative interest in everything that went on in the Institute.

Personally I am in particular very grateful to him for his guidance during the period of six months preceding the date of his retirement, during which the shift of the various responsibilities which together constitute the directorship of a laboratory was gradually effected.

This year's Annual Progress Report has been constructed along largely the same lines as the previous issues, being intended to serve the same general purpose. The major change which we made is the introduction of a few papers of a review character which should provide the reader with a context in which the more specific contributions to this Report as well as other publications from the Institute can be placed.

C.A.A.J. Greebe.
CONTENTS

organisation ii
introduction v
contents vi

1 auditory perception

B.L. Cardozo
Topics in Audition 1

L.L.M. Vogten
Pure-Tone Masking of a Phase Locked Tone-Burst 5

2 speech

S.G. Nooteboom
A Brief Survey of some Investigations into the Temporal Organisation of Speech. 17

S.G. Nooteboom
The Interaction of some Intra-Syllable and Extra-Syllable Factors Acting on Syllable Nucleus Durations. 30

I.H. Slis
On Coarticulation of Articulatory Effort and Syllable Boundaries. 40

3 visual perception

H. Bouma
Introduction I.P.O. Research on Vision 1972. 45

H. Bouma, F.L. Engel, H.E.M. Mélotte
Technological Devices for the Visually Handicapped: Gap Between Research Effort and Available Aids. 46

Conspicuity and Colour. 55

H. Bouma, A.L.M. van Rens
Reflection Problems of a Passive LX Display Provided with a Flat Mirror. 63

J.J. Andriessen
Preferred Viewing Distance and Size of the TV Picture. 65

J.J. Andriessen
On the Optimal Luminance Transfer Function in Television. 68

4 perceptual and motor skills

W.G. Koster
Introduction. 71
perceptual and motor behaviour
F.L. van Nes, R. van Schuur
Speed and Errors of Well-learned Verbal Response Sets.

medical physics
W.G. Koster, R. van Schuur
On the Origin of Physiological Tremor.
80
W.G. Koster, R. van Schuur
A Note on Parkinsonian Tremor
88
R. van Schuur, Th. Burema
On Controlling the Heart Rate.
92
J. Vredenbregt
EMG Application in Low Back Pain
98
J. Vredenbregt
Electromyophone
104

5 instrumentation
D.J.H. Admiraal
105
M.A. Alewijnse
A Pulse-Width Light Modulator.
107
M.A. Alewijnse, J. Vredenbregt
A Six-Channel Electro-Myographic Measuring System.
112
P.W. Verhagen, M.A. Alewijnse
Photomatrix Computer Interface,
115

6 i.p.o. publications 1972
121
1 AUDITORY PERCEPTION
TOPICS IN AUDITION

B.L. Cardozo

It is intended to present in this article a general view of auditory work carried out in the Institute for Perception Research and to give a brief account of investigations in progress not reported in separate contributions. First, however, a few rather generalizing remarks will be made about auditory research as a whole.

It is our belief that it is valuable to try and combine psychoacoustic and physiological data, even though our grasp of the hydromechanics of the cochlea is quite weak, even though we have only sampled knowledge of the coding in the auditory pathway and even though, in particular, there is up to now virtually no physiological access to the connectivity and cooperation of auditory neurons. This belief does not imply that physiology is to become part of our research programme. But it does form an argument for a good contact with workers in the field of physiology.

In line with this, a Symposium on Hearing Theory was held at the Institute on June 22 and 23, 1972, during which both theoretical and experimental work was presented on such problems as: the hydromechanics of the cochlea, the coding of frequency information and time information in the auditory pathway, inferences from the frequency and time characteristics of masking, and relations between dichotic hearing and the perception of pitch, in particular the pitch of absent fundamentals.

We come now to our research programme. One of the centres of standing interest has been, and still is, auditory frequency selectivity. This has been investigated in the past by measuring the audibility of harmonics in periodic pulse trains as well as in periodic noise (Duifhuis 1970a, 1971 and Duifhuis and Tomesen, 1970). The notion that a masking curve can be measured without interference from beats was described by Cardozo (1967). Recently Vogten (1972) found an elegant method for scanning the masking curve of a continuous sinusoid with a phase locked probe, which is free from beats proper. A thorough insight into the masking process is of basic importance for our understanding of the critical band. Ultimately it should be possible to explain why the critical ratio happens to be of the order of -4 dB, a value which is so basic to the design of communication systems that one is inclined to take for granted that it will also hold in impaired ears - which probably is often not true (cf De Boer, 1961).

It is evident that the critical ratio does not depend only upon the filter characteristic of the cochlea but involves also whatever device is operating at the output of this filter. Modelling of such devices with due regard to physiological data is one way of scanning the many possibilities and of concretizing intuitive notions. (cf Duifhuis, 1972).

Fortunately there are other phenomena than masking that shed some light upon auditory frequency selectivity, be it in a more general sense. The investigations of Van Noorden (1972) into rhythmic fission and fusion should be mentioned here. Rhythmic fusion proper is confined to frequency jumps that corresponded rather well with the critical bandwidth. But the fusion band may be widened by the listener when he is consciously shifting attention to the different frequencies involved. For these shifts a certain amount of time, not infinitely small, is needed. A second vital finding of Van Noorden is that fusion is not perceived between pitches evoked by complex tones, whose band spectrums are sufficiently separated in frequency, even though the pitches of the tones are quite closely together.
These findings may once more warn us for simplifying theories in which the entire process of pitch perception would be reduced to a purely mechanical operation in the cochlea. Rather, they seem to call for a hierarchy of processing stages of stimuli even as simple as pure tones.

The intricacies of pitch perception have been brought out in recent years by the elegant experiments of Houtsma and Goldstein (1972), showing that listeners with good musical ability perceive pitch (in terms of musical intervals) corresponding to a frequency $g$ when presented with components $ng$ and $(n+1)g$, irrespectively of whether these components are presented to the same ear or to opposite ears. Their conclusion that "the pitch of complex tones is mediated via a central processor operating on neural signals derived from those effective stimulus harmonics that are tonotopically resolved" refutes the idea that unresolved components are essential for the perception of absent fundamentals. Experiments of Houtsma and Goldstein type IIA (identification of musical intervals) have been repeated in this laboratory by our guest Letowski and Van Leeuwen.

They asked 8 subjects whether a given musical interval was a full tone up, a semitone up, a semitone down or a full tone down. The fundamental frequency $g$ was chosen in the neighbourhood of 400 Hz, the components $ng$ and $(n+1)g$ were always presented dichotically and $n$ was either 2 or 3. Of course, $n$ was randomized, so that listening to the components would produce correct answers only in a fraction of the trials. 5 subjects (4 of them with musical training) performed nearly perfectly as if they heard the absent fundamental.

Within the experimental error performance was shown to be independent of the sensation level of the tones in the range 15 dB SL to 90 dB SL. At the high levels, a background of white noise was necessary that would mask (crosstalk-)sinusoids up to 35 dB SL. Subsequently the experiment was given a slight variation by introducing different levels into the right and into the left ear. As is shown in Fig. 1 the performance drops rather abruptly to chance level when the difference in sound levels between the ears exceeds about 20 dB.

Our experiments measuring the interaural difference in the arrival time of clicks eliciting lateralization have been repeated in order to find a better estimate of the internal (neural) jitter in the auditory system. In this series, the pulse rate was taken to be 200 Hz in stead of 50 Hz as was the case in the earlier experiments (cf Neelen & Van Leeuwen, 1971). The general result is a confirmation of the earlier conclusion that the binaural lateralization system is characterized by an internal jitter of the order of 25/μs. The accuracy of the later measurements, unfortunately is not better than those already described. No attempt has been made to asses this value for separate frequency regions in order to see whether jitter in the lateralization system is a function of frequency like was found in the pitch extracting system (cf Cardozo, Rijckaert & Neelen, 1969). The measurements are quite laborious and a considerable improvement in the technique would be needed.

Of a less theoretical nature is our interest in audiometry. Our efforts here are directed towards two goals. The first is to try and introduce techniques and strategies which have been developed in the laboratory towards finding thresholds or, more generally, measuring perception performance, into audiometrical practice. In the second place it is felt that there is a need for a better understanding of processes that determine intelligibility of speech. Although we are aware of the extreme complexity of the problem, we think that even certain basic factors have not been explored sufficiently. For one thing, the differences in perceived loudness between various speech segments are not too well known. But the experiments of Van Nierop
The heavy lines (open circles) refer to 90 dB SL at right ear; at a difference of 55 dB, the left ear stimulus would disappear in a masking noise, applied equally to both ears. For the thin lines, these figures are 40 dB SL and 30 dB respectively.

(cf Cardozo & Van Nierop, 1970, and Slis & Van Nierop, 1970) have shown that sound intensity of one speech segment may well affect the perception of a word. This problem may also be phrased in the following way: what is the effect of the consonant-to-vowel power ratio upon the intelligibility of speech? This has led to measurements of the dynamic range in speech as a function of vocal level. And this has in turn inspired de Jong to measure the static pressure in the mouth during occlusion with /p/ and /t/ in combination with the sound pressure produced during the release of the plosives and also during the adjacent vowel (Cardozo & de Jong, 1970, de Jong 1971). These measurements have been repeated and extended recently. The overall result was a confirmation (of the previous finding) that the sound level of the plosives is directly proportional to the intraoral pressure before release, but that the sound level of vowels is directly proportional to (an estimate of) the subglottal pressure in the power 1.6 approximately.
references


Vogten, L.L.M. (1972) Pure Tone Masking of a Phase Locked Tone-Burst, this issue.
PURE-TONE MASKING OF A PHASE-LOCKED TONE BURST

L.L.M. Vogten

introduction

In literature a large number of auditory masking experiments have been reported, in which the threshold shift of a masked signal (the probe) is investigated owing to the presence of a random or pseudo random masker, e.g. a band of noise. Substantially fewer results have been published about simultaneous masking with narrow-band deterministic stimuli, particularly with a stationary sinusoidal masker. In general, the use of narrow-band stimuli results in more complicated masking patterns, owing to the detection of interferences or beats, combination tones and harmonics (e.g. Egan 1950, Zwicker 1967, Ehmer 1959 and Greenwood 1971). If, however, we take an entirely deterministic stimulus, it will be possible to give an exact physical description of the interaction between probe and masker, also for unequal frequencies. Mainly within the frequency region where beats affect the masking results, the interaction and cross correlation between masker and probe is determined by the difference in frequency as well as by the mutual phase relation. Fixing this phase relation, even when probe-carrier and masker frequency are not exactly equal, allows of determining which part of the threshold shift must be attributed to changes in the stimulus.

In the present masking experiments we use, for system-theoretical reasons, a stationary sinusoidal masker. The masking of a probe with a fixed carrier frequency $f_p$ will be studied as a function of the masker frequency $f_m$ and of the masker level $L_m$, especially in the region where beats occur.

Before presenting the results concerning a probe with a smoothed envelope, we shall calculate expressions of several basic physical parameters, like envelope, instantaneous frequency and energy, for a stimulus with a rectangular probe. For the present we shall not go into detailed measurements on the detection of combination tones and harmonics.

literature

Experiments with a stationary sinusoidal masker are described by Egan (1950), Ehmer (1959), Small (1959), Green (1969), Greenwood (1971) and McFadden (1972). Table I shows the most important data. All authors use as the probe a tone burst, with a duration varying from 10 up to 500 ms, except Small, who takes a stationary sinusoid. Only McFadden and Green mention a fixed phase relation between probe and masker (viz. zero), however, only for the case of equal frequencies. Simple comparison of the results is not possible, owing to the different probe durations and to the different choice of the variables plotted in the figures. Nevertheless, some remarks can be made:

(a) Details of the masking curves show large discrepancies. In the curves of Egan and McFadden relative minima for $f_m = f_p$ are clearly discernible, while such minima are absent in Ehmer's curves. The other authors (Green, Small) present too few data points or avoid $f_m = f_p$ at all (Greenwood). It is not clear whether these differences are caused by differences in intensity, probe duration, instruction of the subject or by an unfixed phase relation.

(b) As far as masking curves are given for more than one intensity, the relation between maximum masking and the masker level is not a linear one.
<table>
<thead>
<tr>
<th>Masker</th>
<th>Stationary Sine Wave</th>
<th>Probe</th>
<th>Method of Measurement</th>
<th>Phase Locked</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level $L_m$</td>
<td>Frequency $f_m$</td>
<td>Type</td>
<td>Level $L_p$</td>
<td>Frequency $f_p$</td>
<td></td>
</tr>
<tr>
<td>Egan 1950</td>
<td>40, 60, 80 dBSPL, 0 15 dBSPL</td>
<td>0.4 kHz</td>
<td>Tone Burst</td>
<td>0.7 Hz, $T = 500$ ms, $\tau = 100$ ms</td>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Ehmer 1959</td>
<td>20, 40, 60, 80, 100 dBSL</td>
<td>0.25, 0.5, 1, 2, 4, 8 kHz</td>
<td>Tone Burst</td>
<td>2.5 Hz, $T = 200$ ms, &quot;Without Transient's&quot;</td>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Small 1959</td>
<td>Dependent Variable, Independent Variable</td>
<td>Stationary Sine Wave</td>
<td>15, 30, 60 dBSL</td>
<td>0.4, 0.8, 1.6, 3.2, 6.4 kHz</td>
<td></td>
</tr>
<tr>
<td>Green 1969</td>
<td>70 dBSPL</td>
<td>Independent Variable</td>
<td>Tone Burst</td>
<td>$T = 9.5$ ms, &quot;Lightly Filtered&quot;</td>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Greenwood 1971</td>
<td>40, 45, 50...85, 90 dBSPL</td>
<td>0.3, 0.6, 1, 2 kHz</td>
<td>Tone Burst</td>
<td>3 Hz, $T = 165$ ms, $\tau = 30$ ms</td>
<td>Dependent Variable</td>
</tr>
<tr>
<td>McFadden 1972</td>
<td>70 dBSPL</td>
<td>0.4 kHz</td>
<td>Tone Burst</td>
<td>$T = 100$ ms, $\tau = 10$ ms</td>
<td>Dependent Variable</td>
</tr>
</tbody>
</table>

Table I.

Review of some masking experiments with a stationary sinusoidal masker. $T$ is the effective duration and $\tau$ is the rise/decay time of the probe. For more details the reader is referred to the original papers.
(c) The masking curves with the masker level kept constant are almost symmetrical on a logarithmic frequency scale up to 40 dB SL masker level (Greenwood, Ehmer). Beyond 40 dB SL the curves are asymmetrical; the slope of the steepest side remains almost constant while the other slope varies widely.

(d) For medium levels, notches in the masking curves for \( f_m < f_p \) can be attributed to the detection of combination tones (Greenwood, Small), and for high levels to the detection of harmonics (Egan, Ehmer, Greenwood).

**envelope, instantaneous frequency and energy**

The total energy content of the stimulus (probe + masker) depends largely on the difference in frequency and phase. In this section we shall elaborate this dependence. Although a smoothed probe is used in the experiments, in this section the expressions are, for simplicity, given for a rectangular tone burst, see Fig. 1.

![Fig. 1. The stimulus as used in the calculations. Masker (upper curve): stationary sine wave with amplitude \( M \) and frequency \( f_m \). The probe (lower curve) has an amplitude \( P \) and a carrier frequency \( f_p = 1\) kHz. The probe duration \( T \) is an integral number of probe carrier \( f_p \) periods. \( T_o \) is an integral number of masker periods and depends on \( f_m \).](image)

The **probe** with a duration of \( T \) s is given by

\[
P \sin \omega_o t \quad \text{for } 0 \leq t \leq T
\]

\[
0 \quad \text{for } T < t < T_o.
\]

with \( \omega_o = 2\pi f_p \) the carrier frequency, \( T_o \) the repetition time of the (quasi) periodically presented probe (about 500 ms), and \( T = k/f_p \) with \( k \) an integer.

For the **masker** we may write

\[
M \sin (\Omega t + \phi) = M \sin (\omega_o t + \Delta \omega t + \phi),
\]

in which \( \Delta \omega = 2\pi \Delta f \) is the difference between the masker frequency \( \Omega = 2\pi f_m \) and the probe frequency \( \omega_o \), and \( \phi \) is the phase shift for \( t = 0 \), the time at which the probe starts (see Fig. 1).

After some calculations for the stimulus in the interval \( 0 \leq t \leq T \), addition of (1) and (2) yields:
\[ MV_1 + p^2 + 2p \cos (\Delta \omega t + \psi) \sin (\omega t + \psi(t)), \]  

with \( p = P/M \) and \( \psi(t) = \arctan \frac{\sin(\Delta \omega t + \psi)}{p \cos(\Delta \omega t + \psi)} \) from which it can be seen that the stimulus is modulated in both amplitude and frequency (phase) since both are time dependent.

For \( 0 \leq t \leq T \) the instantaneous amplitude or stimulus envelope \( A(t) \) is given by

\[ A(t) = MV_1 + p^2 + 2p \cos (\Delta \omega t + \psi) \]  

and the instantaneous stimulus frequency \( \omega_{\text{mom}} \), defined as \( \frac{d}{dt} (\omega t + \psi(t)), \) is given by

\[ \omega_{\text{mom}}(t) = \omega_0 + \Delta \omega \frac{1 + p \cos(\Delta \omega t + \psi)}{1 + p^2 + 2p \cos(\Delta \omega t + \psi)} \]  

With \( p \ll 1 \) (if \( M \) is not too low, \( p \) is smaller than 0.05), (4) and (5) may, to a good approximation, be replaced by

\[ A(t) = M + pM \cos(\Delta \omega t + \psi) \]  

\[ \omega_{\text{mom}}(t) = \omega_0 + p \Delta \omega \cos(\Delta \omega t + \psi). \]  

The probe energy is

\[ E_p = \int_0^T (P \sin \omega t)^2 dt \]  

or

\[ E_p = \frac{1}{2} P^2 T (1 - \sin \frac{2 \omega_0 T}{2 \omega_0 T}). \]  

Since the probe contains an integral number of carrier periods \((T = k/f_p)\), (6) becomes

\[ E_p = \frac{1}{2} P^2 T. \]  

In the same way we obtain for the masker energy

\[ E_m = \int_0^T (M \sin (\omega t + \Delta \omega t + \psi))^2 dt \]  

with \( T_0 = \ell/f_m \) and \( \ell \) an integer:

\[ E_m = \frac{1}{2} M^2 T_0. \]  

The total stimulus energy \( E_s \) in the interval \( 0 \leq t \leq T_0 \) is

\[ E_s = \int_0^T (P \sin \omega t + M \sin (\Delta \omega t + \omega t + \psi))^2 dt \]  

or

\[ E_s = E_m + E_p + \frac{2MP \omega_0}{\Delta \omega (2 \omega_0 + \Delta \omega)} (\sin(\Delta \omega T + \phi) - \sin \phi). \]

Hence, the energy difference between the stimulus with and without probe is given by

\[ \Delta E = E_s - E_m = \frac{1}{2} P^2 T + MPT \frac{2 \omega_0}{2 \omega_0 + \Delta \omega} \sin \frac{\Delta \omega T}{\Delta \omega T} \cos(\Delta \omega T + \phi), \]  

being the sum of \( E_p \) and the cross correlation of probe and masker.

Special cases:

(a) equal frequencies:

\[ \Delta E = \frac{1}{2} P^2 T + MPT \cos \phi \]  

(b) unequal frequencies,

\[ \phi = 0 : \]

\[ \Delta E = \frac{1}{2} P^2 T + MPT \frac{2 \omega_0}{2 \omega_0 + \Delta \omega} \cos \frac{\Delta \omega T}{\Delta \omega T} \]  

\[ \phi = \pi/2 : \]

\[ \Delta E = \frac{1}{2} P^2 T + MPT \frac{2 \omega_0}{2 \omega_0 + \Delta \omega} \cos \frac{\Delta \omega T - \pi}{\Delta \omega T} \]  

Formula (9) shows how the total energy difference in the stimulus is mainly determined by the cross correlation term, especially in the region where probe carrier and masker frequency are close to each other (\( \Delta \omega \) not too large).

With \( \phi = 0 \), for example, the second term in (9b) is zero when \( \sin \Delta \omega T = 0 \). Thus at all masker frequencies where \( \Delta f = n/2T \), with \( n \) a non-zero integer, the total energy difference will drop drastically.
Example: for $T = 50 \text{ ms}$, $M = 2.5 \text{ mV}$ and $P = 0.1 \text{ mV}$ it follows that $\frac{1}{2}P^2T = 0.25 \cdot 0^{-9} \text{ Ws}$ and $MPT = 12.5 \cdot 10^{-9} \text{ Ws}$. So with $\phi = 0$ the energy difference $AE$ for $\Delta f = 10 \text{ Hz}$ (the first zero of sinc $\Delta \omega T$) is $10 \log \frac{0.25 + 12.5}{0.25} = 17 \text{ dB}$ lower than at $\Delta f = 0$.

Therefore, interpreting masking results, it has to be realized that the alterations in energy difference of the stimulus as a function of $\Delta \omega$ and $\phi$, play an important role in the detection of the probe.

In the section on results we shall return to this point.

**method**

**STIMULUS**

The probe we used in the experiments is smooth shaped as shown in Fig. 2a. For an effective probe duration of 10 ms and 50 ms the measured amplitude spectrum is plotted in Fig. 2b.

**Fig. 2a.** The smooth-shaped probe as used in the experiments.

**Fig. 2b.** The measured (relative) amplitude spectra for probe durations of 10 ms and 50 ms.

The stimulus is the sum of the quasi periodically repeated probe and a stationary sinusoidal masker. During the experiments the amplitude $M$, the frequency $f_m$ and the phase shift $\phi$ are variable. The probe is only variable in amplitude $P$ and in effective duration $T$. Probe carrier frequency $f_p$, rise and decay time $\tau$ and the repetition frequency are kept constant at 1 kHz, 3 ms and about 2 Hz, respectively.
APPARATUS

A simplified block diagram of the equipment is given in Fig. 3.

Fig. 3 Block diagram of the equipment.

Oscillator A generates the sinusoidal masker of variable frequency $f_m$. From this voltage short impulses are derived in the circuits PHASE and SYNC, having the same frequency as the masker and a phase shift that is adjustable from 0 to $2\pi$. Together with a 2 Hz square voltage, providing partly the repetition frequency of the probe, these impulses produce a rectangular voltage whose leading edge coincides with one of the impulses. Hence, the time interval between two probe presentations is restricted to an integral number of masker periods.

The leading edge of the rectangular voltage starts both the 1 kHz generator C for the probe carrier and the envelope generator (PWM-gate, Admiraal 1971). Multiplication of carrier and envelope (Vario-S-gate, Admiraal 1971) provides the tone burst, serving as the probe.

In this way the phase relation between probe and masker is identical for every particular probe presentation.

PROCEDURE

The subject, seated in a sound-insulated booth (IAC) is listening to the stimulus diotically by Sennheiser HD 414 headphones.

During one trial he adjusts either (a) the masker frequency or the probe level at a constant masker level, or (b) the masker frequency or level at a constant probe level, so that the probe becomes just inaudible. The frequency is continuously variable and the levels can be adjusted in steps of 1 dB.

The subject is instructed to use as criterion of detection: anything with a repetition frequency of 2 Hz; he is not obliged to identify the probe in terms of a 1 kHz tone burst. In the regions where the curves are rather flat the measurements are carried out at a fixed masker frequency by adjusting the amplitude of the masker or the probe. In the steeper parts the masker frequency is adjusted at fixed masker and probe levels. Each data point is the result of 5 adjustments. The 2$\sigma$-interval is estimated by taking the product of the range (maximum minus minimum) and $2\Delta n$ with $n = 5$ and $A_5 = 0.43$. 
results

The masking or threshold shift of a probe with a fixed carrier frequency is a function of two main variables: the masker level \( L_m \) and the masker frequency \( f_m \). So we can distinguish three methods of measuring the "masking surface" shown in Fig. 10.

(a) measurement of the iso-\( L_p \) curves. The subject adjusts the masker frequency or the masker level so that the probe of a fixed level is just inaudible.

(b) measurement of the iso-\( L_m \) curves. Here the subject varies either the masker frequency or the probe level to make the probe inaudible.

(c) measurement of the iso-\( f_m \) curves. At a fixed masker frequency the subject varies the masker or probe level to mask the probe.

Neither the iso-\( L_m \) curves nor the iso-\( L_p \) curves are "masking curves" in the conventional sense, because the masker frequency is variable and the probe carrier frequency is not. Iso-\( f_m \) curves have been determined for the case in which \( f_m \) is equal to the probe carrier frequency \( f_p \) and will be discussed afterwards. Although the data are based upon only one subject (LV), some preliminary experiments with four other subjects showed the same trend.

It should be noted that until now only the frequency region where \( f_m \) is close to \( f_p \) has been investigated in detail. So no conclusions can be drawn about detection of combination tones or harmonics.

masking for zero phase shift

In Figs. 4 and 5 the iso-\( L_m \) and iso-\( L_p \) curves are shown for a probe duration of 10 ms. The probe starts at zero phase shift of the masker. It is striking that for increasing intensity the curves shift towards lower frequencies. One would expect that the extremum occurs at a masker frequency \( f_m \) equal to the probe carrier frequency \( f_p \).

Here, however, it appears that at low levels (10 or 20 dB SL) maximum masking occurs when \( f_m \) is 50 to 80 Hz above \( f_p \), whereas at levels above 60 dB SL the masking is maximum at a masker frequency of 100 to 200 Hz below \( f_p \).

Only at masker levels below 40 dB SL the iso-\( L_m \) curves are almost symmetrical, although round a frequency that is shifted with respect to \( f_p \), see Fig. 4.

At higher masker levels asymmetry appears; consequently the iso-\( L_p \) curves of Fig. 5 are also asymmetrical even at a low probe level, except in a small region round the extremum.

The flanks on the high frequency side in Fig. 4 have a slope of roughly 80 dB/oct and in Fig. 5 of 190 dB/oct. On the low frequency side the slope varies from 20 to 40 dB/oct in Fig. 4 and from 10 to 100 dB/oct in Fig. 5.

Results of measurements with other probe durations are presented in Figs. 6 and 7 for \( T = 50 \) ms and in Fig. 8 for \( T = 200 \) ms. For the latter case the complementary group of iso-\( L_m \) curves has not yet been measured.

Very striking is the fact that for increasing probe duration, and thus for decreasing spectral width, the width of the curve increases. Comparison of Figs. 5, 7 and 8 clearly shows a gradually increasing frequency region between the flanks of the iso-\( L_p \) curves, the extremes becoming less "sharp".

masking for a phase shift of \( \pi/2 \)

Above, we have presented the formal expressions of the stimulus energy difference \( \Delta E \) for a rectangular probe. On account of these results, (9b) and (9c), one would expect very pronounced peaks in the curves for not too large frequency differences, also when using a smoothed probe. Therefore we carried out more detailed measurements of the iso-\( L_p \) curve for a probe of 20 dB SL and 50 ms duration at \( \phi = 0 \) and \( \phi = \pi/2 \), see Fig. 9. In fact, the \( \phi = 0 \) curve of Fig. 9 is a detail of the 20 dB SL iso-\( L_p \) curve of Fig. 7.
Fig. 4. Masking of a 10 ms probe (1 kHz carrier) as a function of the masker frequency $f_m$, for several masker levels $L_m$ (iso-$L_m$ curves). The estimated 95% confidence interval is indicated by the horizontal bars.

Fig. 5. Masker level necessary to mask a 10 ms probe as a function of the masker frequency for several probe levels $L_p$ (iso-$L_p$ curves). Bars indicate the 95% confidence interval.

Fig. 6. Iso-$L_m$ curves for a probe duration of 50 ms.

Fig. 7. Iso-$L_p$ curves for a probe duration of 50 ms.
Fig. 8. Iso-$L_p$ curves for a probe duration of 200 ms.

Fig. 9. A detail of the Iso-$L_p$ curve of Fig. 7 with $L_p = 20$ dBSL. Estimated 95% confidence intervals are indicated by vertical bars.

Fig. 10. Interrelationship between iso-$L_m$, iso-$L_p$ and iso-$f_m$ curves. Plotting the probe threshold shift as a function of both masker frequency $f_m$ and masker level $L_m$ results in the "masking surface" as sketched.

Fig. 11. The probe threshold shift as a function of the masker level for equal probe and masker frequencies (iso-$f_m$ curve). The solid line corresponds to a just noticeable amplitude increment of 4% or 1/3 dB. Theoretical probe threshold for $\phi = \pi/2$ are given by the dotted line. Vertical bars indicate the 95% confidence interval.
For \( \phi = 0 \) no clear dips occur at 1010 Hz and 990 Hz. The difference with respect to the level at 1 kHz is only 4 dB instead of the 17 dB calculated in the section on energy, for a rectangular probe. Moreover, the difference between the \( \phi = 0 \) curve and the \( \phi = \pi/2 \) curve is much smaller than would follow from (9b) and (9c). Except for \( f_m = f_p \) it is even insignificant.

### Masking at \( f_m = f_p \)

In this section we shall go further into the masking as a function of the masker level \( L_m \) at equal frequencies, \( f_m = 1 \) kHz.

For a probe duration of 50 ms the measured iso-\( f_m \) curves are plotted in Fig. 11 for \( \phi = 0 \) and \( \phi = \pi/2 \). Proportionality to \( L_m \) holds for \( \phi = 0 \) between 30 and 70 dB SL masker level and the data can be fitted by a straight line with slope 1. The iso-\( f_m \) curve for \( \phi = 0 \) also represents (implicitly) the just noticeable relative amplitude difference \( P/M \) as a function of \( L_m \). Between 30 and 70 dB SL this relative difference is independent of the masker level and has a value of 4% or 1/3 dB, shown as the solid line in Fig. 11. Deviations from the 4% line occur for low as well for high masker levels. They will be treated in the discussion.

From the known probe thresholds for \( \phi = 0 \), the thresholds for \( \phi = \pi/2 \) can be predicted if we assume that the \( \pi/2 \) thresholds are determined by the same amplitude or energy increment of the stimulus as necessary for probe detection at \( \phi = 0 \). For \( \Delta \omega = 0 \) the amplitude difference \( \Delta A \) between the stimulus with and without probe is given by (4) as

\[
\Delta A = (1 + p^2 + 2p \cos \phi - 1)M \quad \text{with} \quad p = P/M.
\]  
(10)

The difference in energy \( \Delta E \) is

\[
\Delta E = (jp^2 + p \cos \phi) M^2 T
\]  
(9a)

Let \( p_o \) be the just audible probe-to-masker ratio for \( \phi = 0 \). Then from (9a) as well as from (10) follows \( jp^2 + p \cos \phi = jp_o^2 + p_o \), so the relative threshold \( p \) as a function of \( \phi \) is given by

\[
p = \sqrt{\cos^2 \phi + p_o^2 + 2p_o^2} - \cos \phi,
\]  
(11)

in which the plus signs hold for \( -\phi \leq \phi \leq \phi \) (increment of amplitude and energy) and the minus signs else (decrement).

The transition between the two cases takes place at the "boundary angle" in the second or third quadrant, given by

\[
\cos \phi_g = -\sqrt{2p_o - p_o^2} \quad \text{or} \quad \sin \phi_g = \pm (1-p_o)
\]  
(12)

From (11) follows the threshold difference between \( \phi = \pi/2 \) and \( \phi = 0 \) as

\[
10 \log \left(1 + 2/p_o\right) \quad \text{dB}
\]  
(13)

with \( p_o \) the just noticeable relative amplitude increment.

Plotting the \( \pi/2 \) thresholds calculated with the aid of (13) then leads to the theoretical \( \pi/2 \) curve as given by the dotted line in Fig. 11. Only at low masker levels the correspondence with the measured \( \pi/2 \) thresholds is good. Beyond \( L_m = 30 \) dB SL the measured values are considerably lower than one would expect on account of energy changes in the stimulus. In the discussion we shall go further into this discrepancy.

### Discussion

All experiments reported upon above have been carried out with a probe of fixed frequency. The classical term "masking curve", however, is used when masking, caused by a fixed-frequency masker, is plotted as a function of the probe frequency.
From now on the reader should be aware of the fact that the probe frequency is fixed and the masker frequency variable. For not too large frequency differences our iso-
$L_m$ curves can be transformed into the classical masking curves. The wording used is applicable to both situations.

Qualitatively the data from literature on masking at large frequency differences between masker and probe are corroborated. For small frequency differences, however, we find that, generally, masking is maximum when the probe carrier frequency is not equal to the masker frequency. This phenomenon can be summarised as follows.

I At low levels the masking is maximum for a probe frequency definitely below the masker frequency. A deviation of 50 to 80 Hz has been measured in the $f_m = 1$ kHz region. As the level is raised, the deviation becomes smaller, that is:

II There is an intermediate intensity at which the maximum masking occurs when probe and masker frequencies are equal.

III Increasing the intensity further, we find maximum masking for a probe frequency definitely above the masker frequency. Here deviations of 100 to 200 Hz have been found.

While most investigators of masking take it for granted that maximum masking occurs when probe and masker frequency coincide, one can find a few indications of case III in the literature. One such indication is found in Zwislocki's (1968) paper, where, however, central masking is described. He found that a loud tone in one ear has a maximum masking effect when the probe in the contralateral ear has a frequency which is slightly above that of a loud tone.

A second indication stems from Plomp (1971) who performed experiments on forward masking. He found a deviation similar to case III.

Scrupulously examining the literature, we find some further indications for a top shift in the data of Small (1959: Figs. 2 and 4) and Greenwood (1971: Figs. 3, 4 and 5). In their texts, however, this phenomenon goes unmentioned.

It seems interesting to try to find a relation between known asymmetries in the cochlear excitation pattern and the phenomena mentioned above.

Two types of asymmetry are well known.

First we know that the excitation pattern is organized along an approximately logarithmic frequency scale (the Bark scale, Zwicker 1967). At low levels the excitation pattern of the probe is steeper on the high frequency side, owing to the spectral width of the probe, whereas the masker excitation is symmetrical. If it is assumed that the probe is detected when the excitation by addition of the probe exceeds a critical amount, this detection will be the most difficult at a probe frequency slightly above the masker frequency. At high levels the excitation pattern of the masker is less steep on the high frequency side, whereas the probe excitation then becomes more symmetrical. So in both cases one would expect maximum masking when the probe frequency lies slightly above the masker frequency. This is what we find in case III (high levels), but this trivial argument cannot explain the results quantitatively. With a 200 ms probe, for instance, a shift of 100 to 200 Hz is clearly at variance with any reasonable spectral width of the probe.

Secondly, there is the top shift in amplitude of the cochlear microphonics towards higher frequencies when the intensity of a pure tone of constant frequency is increased (Honrubia 1968). Now, if probe and masker do not excite simultaneously one and the same cochlea, one might explain case III qualitatively from this top shift. However, for high intensities the probe excitation would shift in the same direction when masker and probe excite one cochlea simultaneously. So this would result in no shift at all. Summing up, these explanations based upon some asymmetry of the cochlear excitation pattern do not fit quantitatively the observed deviations.
At low levels they even point in the wrong direction for the observed phenomenon I.

Varying the probe duration, we find that the masking area broadens for increasing duration and thus for decreasing spectral width. It is not clear what role the spectral composition of the probe plays in its off-frequency detection and how the probe duration affects the shape of the masking curves.

The masking as a function of the masker level at equal masker and probe carrier frequencies shows the same trend as found by Leshowitz (1971: Fig. 2) for a 10 ms rectangular probe. For medium levels and $\phi = 0$ the probe threshold is proportional to the masker level.

The deviations at low sensation levels ("negative masking") can be attributed to a physical change in the stimulus. Addition of the masker then leads to a decreasing probe threshold because of the increasing energy difference $\Delta E$ in (9).

For high levels the detection of the probe is apparently facilitated, probably by the dominating spectral side lobes.

A phase shift of $\pi/2$ between probe and masker results in a probe threshold shift that only for a masker level below 30 dB SL corresponds to the expected value. For medium and high levels the measured threshold for $\phi = \pi/2$ is much lower than calculated from the in-phase situation on the basis of energy considerations.

Leshowitz (1971) showed that in cases where off-frequency detection is obstructed by filtering of the probe, the proportionality region is considerably extended. So especially for a $\pi/2$ phase shift the detection of side lobes seems to have an important influence on the probe threshold.

summary

Masking of a 1 kHz tone burst by a stationary sine wave as a function of the masker frequency for various sensation levels is investigated. The probe is phase locked to the masker, also for unequal frequencies. Two methods of measurement are used: (a) direct determination of the probe threshold shift at a constant masker level, and (b) measurement of the masker level necessary to mask a probe of a constant level.

At low levels masking appears to be maximum for masker frequencies up to 8% above the probe carrier frequency. For increasing levels, the masker frequency at which masking is maximum decreases systematically.

references


A SURVEY OF SOME INVESTIGATIONS INTO THE TEMPORAL ORGANISATION OF SPEECH

S.G. Nooteboom

Introduction

In our Institute experimental phonetics was introduced as a research activity in 1959. At that time the major effort in many speech laboratories was directed towards the spectral organisation of speech. In contrast to this it was deliberately decided that in the IPO one of the main lines of phonetics research should be the study of the temporal organisation of speech. In the research done since that time a strong linguistic bias has always been present in that the phonetic investigations were inspired by a desire to make explicit the relation between the linguistic code and the speech event, both in the production and the perception of speech. The explicitation of this relation may guide us in investigating how, on the one hand, the production and perception of speech are controlled by the mental structures of language, and how, on the other hand, the organisation of language structure is constrained by the production and perception mechanisms.

At present we are still very far away from formulating any extensive theory of the mutual interaction between language structure and the production and perception mechanisms. In fact, it must be admitted that this domain of research is still largely in the data gathering stage, although in the past twenty years a great deal of knowledge has been gained in a number of laboratories all over the world. The lack of a coherent theory leaves us with a number of rather isolated interesting findings, which may perhaps ultimately help in formulating such a theory, but which as yet cannot be related to each other in a meaningful way. This also has the effect that experimental data do not rapidly lose their actuality. Data obtained quite a long time ago keep staring at us, asking for an explanation. We also have to be careful not to forget them as long as they have not lost their potential capability of inspiring new and interesting hypotheses. It is for this reason that an attempt will be made here to give a survey of the main findings concerning the temporal organisation of speech, as obtained in our Institute. From this survey the work on intonation, constituting a major part of the speech research in the IPO, will be omitted. Even so it will not be possible to give, within the frame of this rather brief survey, the results of each particular investigation the attention they deserve. The main purpose of this paper is to be a reminder. For those who wish to know more particulars a rather extensive list of references is provided. The interpretations and speculations given are my own responsibility.

Subjective analysis and simple synthesis

In the early stages of phonetics research in the IPO much use has been made of electronic gating devices (Cohen and 't Hart 1964, Admiraal 1971). One of the purposes was to analyse perceptually complexes of speech sounds with a view to determining constituent parts. This involved establishing segments of speech which are more or less perceptually homogeneous, finding perceptual boundaries, and assessing the role of the amplitude envelope in speech perception. These phenomena were studied by varying the width, position and amplitude control function of the gate, and judging the perceptual results. From these investigations hypotheses were derived concerning relevant aspects of the time structure of simple sequences of speech sounds. The time structure of speech appeared to be extremely important for perception.
This was demonstrated convincingly with a simple form of speech synthesis. It turned out to be possible to synthesise understandable speech with a reasonable quality from spectrally homogeneous segments, if the durations and amplitude envelopes were carefully controlled. This naturally does not show that formant transitions are not important for speech perception, but they do demonstrate that transitions are dispensable if the temporal composition is optimally shaped (Cohen, Schouten and 't Hart 1962, Schouten, Cohen and 't Hart 1962). (Fig. 1).

perception of synthetically generated isolated vowels

In order to test whether users of a particular language have a "built-in pattern", which enables them to identify certain acoustic signals as sounds of their language and to reject others, a number of perceptual experiments with synthetically generated isolated vowel-like signals was performed. In one of these experiments the subjects had the task to adjust both the duration of that part of the signal that had uniform intensity and the decay time. The results showed that the preferred durational characteristics of a vowel depend on the linguistically given vowel quantity category to which this vowel belongs (Cohen, Slis and 't Hart 1963).

the serial ordering of linguistic units in the brain

Errors of speech give as it were a window on the linguistic organisation of messages in the brain, and bear on the temporal organisation of speech in a quite different way from the earlier mentioned experiments. In such errors as "to collerate quite a lot" (instead of "to tolerate quite a lot"), we see that phoneme-like units change positions like the letters of a compositor in printing errors. Some studies of regularities in errors of speech (Cohen 1966, 1968, Nooteboom 1967, 1969, 1972 b) give empirical support for the psychological reality of some aspects of linguistic descriptions, and provide new data as to the strategy of language users in handling linguistic information in the programming of speech.

![Fig. 1. Time pattern required for building the synthetic word: phonetics (English). Amplitudes are normalized.](image)

![Fig. 2. Relative frequency distribution of spontaneous phonemic speech errors in Dutch as a function of the number of syllables between origin and target.](image)
Units frequently involved in speech errors are phonemes, consonant clusters, syllables, morphemes, words. The distributional constraints of the language are very rarely violated when such units shift from one position to another. The chance of two units interfering with one another in a speech error (by anticipation, perseveration or transposition) increases when the two units become more similar in phonetic form, stress level, distributional properties and/or meaning, (Fig. 2).

In speech errors anticipations are far more frequent than perseverations, whereas transpositions are rarest. The abundance of anticipations shows that a speaker's attention is more directed to what is yet to come than to what has been said. The amount of language material intermediate between two units involved in a speech error generally does not exceed seven syllables. Of particular interest is that units shifting from one position to another in a speech error do not take with them the coarticulation features of their original position, and prosodic features such as stress, intonation and duration. For example in such lapses as the earlier quoted "to colerate quite a lot" the lip rounding in the /k/ of "quite" was found not to be present when this /k/ is anticipated and intrudes in the position of the /t/ of "tolerate". There the /k/ adapts itself to the coarticulation requirements of the new position. This shows that the mental level of speech programming where phonological errors occur in some way is separate from and presumably precedes a lower level where coarticulation is programmed. In another slip, "how things bad were" (instead of "how bad things were"), the slip is quoted from Boomer and Laver (1968), we see that two words are interchanged, leaving the stress pattern, intonation pattern and durational pattern of the sentence intact. (See also Fromkin 1971). We have to assume that prosodic features are in some way programmed separately from segmental units (although stress levels do affect the chance of occurrence of segmental errors). In this respect it may be interesting to note that phonological quantity, often regarded as a prosodic feature (e.g. Lehiste 1970), behaves as a segmental feature in errors of speech. Long and short vowels may replace each other, taking the quantity feature with them.

The special role of the vowel onset in the temporal organisation of speech

It has been shown by Huggins (1964) that the intelligibility of interrupted speech is minimal when the rate of interruption is such that the syllable structure is most affected. However, Huggins' interruptions were not related to each individual syllable, but periodical. If indeed the syllable has any special significance to intelligibility, it would be worth while to know which part of the syllable and what properties of sequences of syllables are most relevant to intelligibility. Therefore, an experiment was designed (Van Katwijk and 't Hart 1967 a, and b, 1970), in which the intelligibility of speech with syllable-tied interruptions was assessed by shadowing. In each syllable either 110 ms were acoustically suppressed or 110 ms were left audible, the remainder being suppressed. The position of the audible or suppressed interval with respect to the vowel onsets was varied. There were two main divisions of the stimulus material, one being the presence or absence of the CV-junctions in the audible parts of the syllables, the other the fixed time relation between the vowel onset and the beginning of the audible part of the same syllable versus the situation in which this time relation was disturbed.

It was found that if CV-junctions are suppressed, intelligibility is low, and if the pattern of onsets of audible parts of syllables is the same as that of vowel onsets in the original speech, intelligibility is better than if this pattern is disturbed.
These results indicate that the perceptual cues relevant to the decoding of speech are not evenly distributed along the time axis, but rather show peaks of concentration round the CV-junctions.

Furthermore, it seems that the CV-junctions play a special role in the temporal organisation of speech. The distribution of vowel onsets along the time axis apparently constitutes some sort of "rhythmical" patterns, having perceptual relevance. This is confirmed by a number of other experiments in which the role of the vowel onset as constituting a "syllable beat" (Allen 1967) is investigated by having subjects synchronise finger taps with audible syllables, audible clicks with audible syllables (Van Katwijk and Van der Burg 1968), a routine scansion of a Dutch poem with audible, equidistant clicks, equidistant audible clicks with audible metrically spoken poem, and on finger taps with audible metrically spoken poem (Eggermont 1969). In all these cases the point of synchronisation was the vowel onset.

We may speculate that the temporal patterns in the sequences of vowel onsets constrain the range of possible interpretations of the speech waveform, and thus constitute valuable aids in the perceptual decoding of speech.

some details in the temporal organisation of speech

One of the linguistic distinctions in the production and perception of which temporal details appear to be very important is the voiced-voiceless distinction. At first sight this distinction seems to be a very simple one. In articulation it seems to be made by the presence or absence of periodicity in the lower parts of the spectrum, and in perception by the corresponding perceptual correlates. This view appears far too simple. It has been shown that in production and perception there is more to this distinction than the correlates of vocal fold vibration alone (Slis 1966, 1967, Slis and Cohen 1967, Slis 1970 b). In measurements of the articulation and the acoustic signal a number of attributes of the voiced-voiceless distinction were found, of which, in perceptual experiments with mutilated real speech and synthetically generated speech materials the following was shown to be relevant perceptual cues:
(1) acoustical duration of the consonantal segment
(2) acoustical duration of the preceding vowel
(3) duration and spectral extensiveness of the vowel formant transitions
(4) presence or absence of vocal vibrations during the consonantal closure
(5) acoustical duration of the noise burst of plosives
(6) sound pressure of the noise burst of plosives and friction noise of fricatives
(7) peak value of the fundamental frequency of the surrounding vowels and the contour of the fundamental frequency in the following vowel.

Fig. 4. Steps involved in changing an unvoiced into a voiced consonant. Sonagrams of the synthetic word "petal" gradually transformed into "pedal".

These perceptual cues are given in arbitrary order. One may notice that 4 of these cues are durational. Whereas in word initial position the presence or absence of the voice bar seems to be more important than the other cues, this appears not to be the case in intervocalic position. There is no problem in inducing "voiced" judgements to synthetic stimuli without a voice bar. In intervocalic position the voiced-voiceless distinction seems to be carried mainly by the durational and intensity cues. The durational cues do not seem to be individually controlled in production, but may, at least partly, be explained by the physiological mechanisms involved in the implementation of the linguistic voiced-voiceless opposition. Thus on the control level of speech this distinction seems to be less complex than in production and perception.

One of the differences between voiced and voiceless plosives is the greater articulatory effort in voiceless as compared to voiced plosives. This greater effort, as shown by higher electromyographic activity, is also present in stressed versus unstressed syllables, in identical plosives after short vowels versus long vowels, and, at least for the Dutch /a/-/-/a/ pair, in identical plosives preceding long vowels versus short ones (Slis 1968, 1970 b, 1971, this issue). In all these cases the stronger electromyographic activity appears to be correlated with an advancement in time of the articulatory gesture, which does not affect the total duration of the word of which it is part.
We thus may assume that the durational correlates of differences in articulatory effort are not controlled individually but are rather physiologically conditioned side-effects of stronger neural commands, needed for other purposes. In measurements of the time structure (Slis 1968) and electromyographic activity in the orbicularis oris muscle (Slis 1970 a) in isolated nonsense words, it was shown that both articulatory durations and muscle activity in the realisation of a consonantal closure may be affected by properties of a consonant belonging to a different syllable. Such influence may take place over a vowel articulation. This suggests that the temporal integration of articulatory commands may extend over stretches of speech greater than a syllable.

In Dutch, and many other languages, an important linguistic factor in the temporal organisation of speech is phonological vowel quantity. In Dutch not all vowels, however, take part in phonological quantity oppositions. Data obtained in articulatory measurements on vowel durations in three syllable nonsense words (Slis and Nooteboom 1969, Nooteboom and Slis 1970 b, Nooteboom 1972 b) can be convincingly explained by assuming two quantity categories, one containing the genuine phonologically short vowels plus the monophthongs having no long counterparts, the other containing the genuine long vowels plus the diphongs (Three of the generally short monophthongs have long allophones in certain distributional positions).

A difficulty in studying quantity oppositions in the actual speech event is that the measurable durations are affected by many other effects as well, such as stress, position, consonantal environment, etc. Most of these factors can be kept constant in the experimental situation, but inherent properties of the vowels, such as rounding and vowel height cannot. This leads to durational differences within one quantity category, also in positions which are made as similar as possible. For a complete account of the temporal details in speech such physiologically conditioned durational differences should be explained by quantitative models of speech production. In this respect the effect of vowel height has perhaps been studied best. This effect has been attributed to the sluggishness of the lower jaw in speech, which has to open further for low vowels than for high vowels (Lindblom 1967, Lindblom and Sundberg 1971), although the active control of tongue height also appears to be used in differences in vowel height, and subjects may differ as to their strategy (Ladefoged, DeClerck, Lindau and Papcun 1972). In bilabial environment the active control of lip opening may also be involved. In some simple measurements, in which the durational effect of vowel height in "pipe speech" (produced with a pencil clenched between the teeth) and normal speech (both in bilabial environment) was compared, it was found that this durational effect does not differ significantly in the two situations. This implies that the sluggishness which is responsible for the durational difference is probably not of a mechanical nature. Evidently the lower jaw is a far heavier mechanical structure than the lips. The sluggishness should rather be attributed to neuromuscular processes (Nooteboom and Slis 1970 a).

Another implication of the study of pipe speech is that the same vowel quality may be brought forth with rather different command patterns for the control of the vocal tract. These patterns cannot be explained by peripheral neural servo-systems such as the gamma loop, as suggested by MacNeilage (1970). The compensatory articulation, which has been shown not to be dependent on time consuming feedback mechanisms (Lindblom 1971), presupposes extensive reorganisation on a rather high level of
programming. It seems to be the case that the "targets" in vowel production are of a perceptual, auditory nature, and that these auditory targets are translated into articulatory movement patterns, given the current state of the vocal tract (Nooteboom 1971, Lindblom 1971, Ladefoged et al. 1972).

**articulatory and perceptual accuracy in the time domain**

The standard deviations of such articulatory segments as closure durations and vowel durations as measured from consonantal release to consonantal closure, in repetitions of the same word in one test series, generally lie between 5 and 10 ms, in some cases they are below 5 ms (Nooteboom 1972 b). This may give an impression of how accurately articulatory timing may be controlled.

To see how accuracy in articulatory timing relates to accuracy in the internal representation of speech forms, a simple perceptual experiment was performed (Nooteboom 1972 a, b, 1973). Three phonetically non-naive subjects were asked to repeat 20 times as accurately as possible the preferred setting of a stressed vowel in a synthesized word. The standard deviations found ranged from 1.7 ms (for a short vowel /a/, mean duration 70 ms, one subject) to 9 ms (for a long vowel /a:/, mean duration 104 ms, another subject). This shows that the internal representation of a vowel duration may be at least as accurate as the articulatory timing, and more accurate than one would expect on the basis of difference limens for the duration of simple auditory stimuli as tones or noise bursts, which are in the order of 10%. In some cases the internal representation of vowel durations is definitely more accurate than the spectrographic measurements of it.

**prosodic patterning in vowel durations**

The timing of opening and closing movements of the mouth in speech, as reflected in measurable vowel durations, shows wide variations due to prosodic patterns. The analysis of the durational structure of simple nonsense words with varying stress placement, varying number of syllables and both long and short vowels, and no differences in the phonemic make-up of syllables, has given rise to a quantitative description of prosodic patterns for vowel durations in isolated words (Nooteboom 1972 a, b). The main characteristics of these patterns are that the stressed vowel is shortened as a function of the number of syllables following in the word, the shortening following approximately a power function.

The stressed vowels are only slightly shortened by the number of syllables preceding in the word. Unstressed vowels are affected by a moderate lengthening in initial syllables, strong shortening in medial syllables, and lengthening in final syllables (also in embedded words). Furthermore, an unstressed vowel in initial syllable is shortened if the next vowel is stressed.

To test the perceptual reality of such prosodic temporal patterns some perceptual experiments were performed (Nooteboom 1972 a, b, 1973). In these experiments a method of internal matching was used, the main characteristic of which is that subjects are asked to give preferred durations to vowels in different conditions. The vowels belonged to synthetically generated common Dutch words with differing stress patterns, and differing number of syllables. It was shown that the preferred durations could be fairly accurately predicted from the quantitative description of prosodic patterns based on the earlier mentioned articulatory measurements of nonsense words. The articulatory and perceptual experiments on prosodic durations together show that there exist generalized prosodic temporal patterns, governing the control of articulatory timing on the
one hand, and restricting the set of perceptually acceptable speech forms on the other.

![Diagram of synthetic words](image)

Fig. 5. Durational build-up of four synthetic words as adjusted by three subjects.

The prosodic factors in temporal patterning interact with each other (for example the number of syllables in the word with the number of syllables in the foot) and with intra-syllable factors (such as the effect of the following consonant in vowel duration) in a quantifiable way. The study and quantitative description of such interactions may help in finding a set of rules describing the temporal patterns of speech (Nooteboom, this issue).

**forward masking and the temporal organisation of speech**

In some psychoacoustic measurements (Cardozo and Van Nierop 1970) the forward masking effect exerted by a preceding stimulus on a noise burst has been studied. In these experiments the spectral composition of the noise burst was varied (white or band filtered noise), and different kinds of sound were used as the preceding masker, viz. white noise, band filtered noise, formant filtered noise with /a/, /i/ and /u/ colour, and synthetic vowels /a/, /i/, /u/. The masking effect was studied as a function of the interval of time between back slopes of masker and maskee, and as a function of the loudness of the maskee.

The masking effect decreases when the masker goes through the stages of white noise, via coloured noise, to a synthetic "whispered" and finally to a synthetic "voiced" vowel, but is greater when masker and maskee have approximately the same spectral composition than when this is not the case. The masking effect is greatest immediately after the masker, and may be 15 to 20 dB. It expires in about 150 ms.

In a follow-up experiment (Slis and Van Nierop 1970) the forward masking effect of vowels on the noise burst of voiceless final plosives in real speech was studied, by attenuating the noise burst, and by advancing the noise burst in time, shifting it closer to the vowel.

![Schematic illustration of masking](image)

The subjects were asked to identify the consonant. The results show that shortening of the silent interval leads to "fricative" perceptions and the combination of the two may lead to "no consonant" judgements. The consonant had to be attenuated 15 dB before the results showed marked masking effects, and the silent interval had to be decreased 60 ms before masking clearly began to manifest itself.
In interpreting these results it must be kept in mind, though, that in the test materials chosen, both the loudness of the noise burst (30-45 dB above threshold) and the duration of the silent interval (90 to 130 ms) were considerably greater than normal in connected speech, because they were spoken optimally in final position. Furthermore in normal listening conditions reverberance may further enhance masking.

In view of this it does not seem unreasonable to assume that certain factors in the temporal organisation of speech are more or less trimmed to temporal constraints of the hearing system. This should be kept in mind when studying the perception of fast speech, for example by speeding up speech by machine. It may also throw some light on the origin of some universal temporal regularities. Especially with respect to the distinction between voiced and voiceless plosives we should not eliminate the possibility that the durational difference in oral closure is not solely due to physiological mechanisms in speech production, as suggested earlier in this paper, but may find its origin in temporal constraints of the hearing system. These constraints do not necessarily work in an absolute sense. It may be that the internal criteria for accepting a speech sound as a fricative or as a voiced plosive are partly controlled by an overall idea of the speaking rate the listener can make himself. It may be important, though, as a limiting factor, or in some special situations.

**rhythmic fusion and fission**

It is by now well-known that, if we listen to a speech utterance and a spurious noise burst is inserted somewhere in the speech continuum, it is very difficult to locate this noise burst perceptually, as a part of the auditory stimulus pattern. This observation may give rise to the question how it is that we have no difficulty in interpreting speech as a definite sequence of speech sounds, which are temporally integrated in meaningful patterns in perception. Slis of our Institute prepared a demonstration tape (not reported upon elsewhere) which may have some interest in this respect. The words of a short sentence were all spoken optimally in isolation by the same speaker, then cut out from the tape and so spliced together in the correct order for a meaningful sentence are not understandable at all. It is as if the words were spoken by different speakers, in different corners of the room, and the order of the words cannot be determined. This effect, and the difficulty in perception, disappear, however, when pauses of about 100 ms are introduced between the words. The version without pauses apparently lacks the properties which would enable a listener to fuse the subsequent words in one perceptual pattern, suitable for decoding. The main difficulty seems to be caused by the extensive jumps in voice pitch at the beginnings and endings of the words.

With stimuli which are much less complicated than real speech Van Noorden (1971) made an attempt to tackle the problem of rhythmic fusion (and its opposite, rhythmic fission) in some psychoacoustic experiments with tone sequences. He used tone sequences of the form ABA ABA..., A and B being pure tones of equal duration and amplitude, and different pitch. Such tone sequences may be heard as recurring patterns of three tones (in fusion state) or as separate continuous tone sequences, one running twice as fast as the other (in the fission state). \( f_B \) was chosen 1000 Hz, \( f_A \) ranged from one octave below to one octave above \( f_B \). The interval duration between the onsets of two successive tones was varied from 50 to 150 ms. In the results three different ranges of \( f_A - f_B \) were distinguished. In one range, with great differences, there was no fusion at all, in another, with small differences, there was always fusion, and in a third, intermediate range, it depended on the psychological set of the subject whether he heard fusion or fission.
The interval duration had only a significant effect on the intermediate range, the fusion range increasing when the interval duration became longer than 100 ms. This may have to do with the subject getting more time to concentrate on the perceptual pattern, when interval durations increase.

More recently similar experiments (Van Noorden 1972) both diotically (both tone sequences on both ears) and dichotically (one tone sequence on each ear) were carried out. There appeared to be a great difference between the two situations. In the dichotic situation integration into perceptual patterns was found to be much less than in the diotic situation. Furthermore, the subjective nature of the perceptual patterning in the dichotic situation is so different from the rhythmic fusion in the diotic situation that one may doubt whether the term "rhythmic fusion" should be used for both situations.

The implication of the differences between diotic and dichotic situation is that perceptual integration of tone sequences into rhythmic patterns is, at least to a certain extent, dependent on peripheral properties of the hearing system.

Tone sequences are one thing, speech a quite different thing. We cannot be priori certain that Van Noorden's results can be applied to the perceptual integration of speech into meaningful patterns. It seems tempting, though, to compare Van Noorden's results with dichotic listening to the results of Huggins (1964), who showed that switching speech from one ear to another at a rate which interfered most with the syllable rate of speech, decreased the intelligibility of speech. Apparently, alternating between ears interferes with the mechanisms for perceptual integration. This supports the idea that non-central properties of the hearing system are involved in the perceptual integration, and also exert constraints on the temporal organisation of speech. The role of the physiological properties of the hearing system in the perceptual integration of speech has yet to be investigated.

some speculations

The mental structures of language control the temporal organisation of speech in a number of different ways. In the phonology of the language the quantity feature leads to durational differences which are programmed as such, and which constitute major cues to the recognition of long versus short phonemes. The prosodic temporal patterns of the language modulate articulatory timing in a well-defined way, so that the resulting durational regularities provide the listener with cues for the perception of word stresses and word boundaries. The implementation of distinctive features in the production of speech leads to temporal regularities, which often appear not to be programmed as such, and are not part of the abstract phonology of the language, but nevertheless can function as perceptual cues to the recognition of higher order linguistic units. An example is provided by the durational attributes of the voiced-voiceless distinction.

All these factors, and many others not mentioned here, together very strictly control the temporal organisation of speech and thus put heavy constraints on the class of acceptable interpretations to speech wave forms. This may well be the most important function of the temporal patterns of speech. It is only when we accept that a listener has a rather limited amount of freedom in generating possible interpretations to incoming speech signals, that we can understand that small durational differences such as those due to the voiced-voiceless distinction can function as perceptual cues, in spite of the enormous variability in realisations of identical phonemes. The overall temporal patterns serve as a background, as it were, against which temporal details may function properly.
It is not only so that the mental structures of language control the production and perception of speech, but the language structures themselves appear to be affected by the temporal constraints of the production and perception mechanisms. The possibilities and limitations of the production and perception systems provide boundary conditions for the organisation of language. The accuracy in the control of articulatory timing and in the internal representation of temporal aspects of speech forms may be considered a precondition for languages to use both, and at the same time, phonological quantity oppositions and prosodic temporal patterns. It also serves as a precondition for small durational regularities, due to temporal constraints of the production system, to function as perceptual cues. On the other hand the temporal constraints of the hearing system limit the amount of temporal details which may function in speech, and thus limit the raw materials from which language may extract the abstract patterns of phonology, both segmental and prosodic.

The interaction between the possibilities and limitations of the production and perception systems on the one hand and the organisation of language on the other has to be taken into account by any serious theory of abstract phonology and by any serious theory of speech performance. This interaction may also provide explanations for historical changes in the phonology of languages. Again an example is provided by the voiced-voiceless distinction. The difference in vowel duration before voiced and voiceless consonants in many languages is small and appears to be physiologically conditioned. In English, however, this difference is great and is part of the phonology of the language. The origin of this phonological feature of English can be explained by assuming that the small, physiologically conditioned difference, of which we know that it functions as a secondary perceptual cue, and thus must be part of the internal representation of speech forms, has become exaggerated and incorporated in the phonology of the language. The distinction between a mental representation of speech forms, which is rather close to the acoustic wave form, and which reflects rather narrowly the temporal constraints of the production and perception mechanisms, and a more abstract phonological representation of speech forms, in terms of a sequence of non-overlapping phonemes, appears to be very useful in explaining naturalness conditions on phonological systems and historical changes.

references


Slis, I.H. (1972 b) this issue.


THE INTERACTION OF SOME INTRA-SYLLABLE AND EXTRA-SYLLABLE FACTORS ACTING ON SYLLABLE NUCLEUS DURATIONS

S.G. Nooteboom

Until very recently most statements to be found in the literature on the durational behaviour of syllable nuclei concerned single effects. Such statements run for example as follows: the durations of syllable nuclei in stressed monosyllables are somewhat shorter before voiceless plosives than before fricatives. Or: the durations of syllable nuclei in stressed syllables tend to be longer than those in unstressed syllables. Or: the durations of syllable nuclei (and other acoustical segments) decrease with increasing word length.

Such statements do not reveal the intricate interactions which may exist between different factors acting on syllable nucleus durations in words and word groups. This has been recognized for example by Klatt (1971), who stated "Syllable lengthening could not be established until most of the other factors involved in determining segmental length were identified and quantified, which is why I believe that the regularities in this pause system have not been noted previously". And: "The generative rule system that has been described is sufficiently involved that it is not possible to predict rule effects and interactions without explicitly programming the entire rule system".

The strong interaction between many factors acting on syllable nucleus durations has been a major obstacle to research on the temporal organization of speech. Most temporal factors have been studied in isolation, but the quantitative results of such studies do not show how the various factors interact quantitatively with one another in connected speech. The solution to this predicament advanced by Klatt is the construction of an entire rule system, which can be tested and changed until its predictions are satisfactory, but Klatt does, in fact, not show how, heuristically, we are to proceed in order to obtain reasonable hypotheses for the rule system.

In this paper I will propose that one possible way of doing this is to start from the study of very simple nonsense words, in which single factors are studied in optimal conditions, and then, combining these factors in nonsense words which are gradually becoming more complex and closer to real words in the language, see how these factors interact quantitatively. The same approach may also extend over word groups.

I will first give an example of the effect of some intra-syllable factors on syllable nucleus durations in monosyllables, secondly, an example of some extra-syllable factors. Thirdly, I will present some data on the internal, auditory representation of syllable nucleus durations, in order to ensure that the kind of phenomena we are studying is perceptually real. Fourthly, I will present some data on the interaction between different factors. Some of the results presented in this paper have been published earlier and discussed more extensively in Nooteboom (1972, 1973).

some intra-syllable factors in syllable nucleus durations

In a series of measurements on monosyllables of the form /mVC/, each syllable type being spoken 6 times within a randomly ordered list by one speaker, data were obtained on regularities in syllable nucleus durations due to intra-syllable factors. The syllable types used can be derived from Fig. 1. The spoken monosyllables were tape-recorded in optimal recording conditions, and high quality normal and duplex oscillograms were made. The syllable nucleus durations were measured by hand, and the mean syllable nucleus durations for each monosyllable type are presented in Fig. 1.
Fig. 1. Syllable nucleus durations for monosyllables of the form /mVC/, for 51 different VC-combinations, each spoken 6 times by one native speaker of Dutch within one test series. Top: long vowels; bottom: short vowels.

Besides the great differences due to phonological vowel quantity we recognize the well-known effect of vowel height, the effect of the plosive, fricative or nasal character of the following consonant, and some other effects of specific vowel-consonant interactions. The differences due to intra-syllable factors in the case of monosyllables spoken in isolation may apparently be of the order of 40 ms for long vowels and 25 ms for short vowels, or 25% of the mean values for both long and short vowels. There is no doubt that such differences may play a role in the perception of speech. In fact, it is known that in Dutch differences in syllable nucleus durations due to the voiced or voiceless character of the postvocalic consonant, which are of the order of 25 ms, may contribute to the perception of the voiced character (Slis and Cohen, 1969).

some extra-syllable factors in syllable nucleus durations

Some data on extra-syllable factors acting on syllable nucleus durations were obtained in articulatory measurements with a lip contact, from the same speaker as in the above measurements. The test materials consisted of simple nonsense words of the form /mːaːm/, /mːaːmaːm/, etc. with different numbers of syllables and different stress placements. The durations of consonants and vowels were measured as the intervals during which the lips were closed and open respectively. The phonemic structure of the nonsense words was chosen as simple as possible in order to eliminate differences in segment durations due to intra-syllable factors. This has the disadvantage that these nonsense words may hardly be called acceptable phoneme sequences to native speakers of Dutch.
For example, four syllable words containing a long vowel in each syllable either do not occur or are very rare loan words (Fischer-Jørgensen). However, I feel that precisely this kind of simple nonsense words may reveal the effect of generalized underlying timing patterns more clearly than more realistically structured words would do. In this way reasonable hypotheses may be obtained which at a later stage may be tested as to their generality.

Some of the data on these nonsense words are presented in Fig. 2.

Fig. 2. Schematical durational build-up of simple nonsense words with word stress on the initial syllable and varying number of unstressed syllables. Closed and open periods of the mouth have been measured by lip contact.
In this case all words were spoken with the stress on the first syllable. The stressed syllable nucleus duration decreases as the number of syllables following in the word increases. It is as if the articulatory system is looking forward and increases its rate of action as a function of the amount of work to be done in the programme. As pointed out by Lindblom and Rapp (1972) this durational behaviour of the stressed syllable nucleus can be described as follows (at least for initial syllables):

\[ V = \frac{D}{m^a} \]  

(1)

in which \( V \) is the duration to be calculated, \( D \) is a standard duration, which has to be chosen differently for phonologically long and short vowels, \( m \) is the number of syllables in the word from the beginning of the syllable concerned, and \( a \) a factor smaller than 1, restricting the effect of \( m \). The data presented in Fig. 2 can be compared to values calculated from rule (1) in Fig. 3. Quantitative rules for unstressed vowels will not be discussed here (See Nooteboom 1972 b, 1973). The example given may suffice to show that there exist rather outspoken prosodic timing patterns in the temporal organization of words which govern the measurable durations of individual syllable nuclei. It may be asked whether such differences in articulatory timing are perceptually real.

**Fig. 3.** Syllable nucleus durations for stressed initial syllables in the words presented in Fig. 2, together with values calculated from rule (1) (see text), with \( D=200 \) ms for long vowels, 100 ms for short vowels, and \( a=0.5 \) and 0.4 for long and short vowels respectively.

**the perceptual reality of some prosodic durations**

In order to study the perceptual reality of some prosodic durations some perceptual experiments were carried out using a method of matching to internal criterion. The experiments and their results were discussed extensively elsewhere (Nooteboom 1972, 1973). Here I will confine myself to presenting some examples of the results. The examples given concern long and short stressed vowels in initial syllables of synthesized versions of common Dutch words, differing in the number of syllables coming later in the word. Two test series of words were used, viz.:

(a) māt, māte, māteloo, māteleoze

(b) pān, pānne, pānnekoek, pānnekoeken

The stressed initial syllables contain a long vowel /a:/ in series (a) and a short vowel /a/ in series (b).
Three naive subjects were asked to so adjust the durations of these syllable nuclei, that the words as a whole sounded as natural as possible. For each individual adjustment the word in question was made audible to the subject over headphones as many times as he wished. Each subject made 10 adjustments for each word. The duration could be altered by turning a knob. The relation between the knob position and the duration was non-linear and changed randomly after each trial. The same words were spoken two months later by the same subjects, each word twice by each subject. These spoken versions were recorded on tape, spectrograms were made and the durations of the stressed syllable nuclei measured. The results of these experiments are presented in Fig. 4. The adjusted durations averaged over all settings of all three subjects are indicated by crosses, the spoken durations by circles. The solid lines indicate calculated values as obtained from (1). The standard durations \( D \) were the same as in the description of the earlier mentioned articulatory data on nonsense words, namely 200 ms for long vowels, and 100 ms for short ones. The power exponent \( \alpha \) however, was made to fit the new data. Below I will return to this discrepancy between the two sets of data. In Fig. 4 the vertical axis gives the syllable nucleus durations in ms, the horizontal axis the number of syllables in the word counted from the beginning of the syllable concerned.

These results, and similar results for unstressed vowels show that the prosodic timing patterns, found in articulatory measurements of simple nonsense words, are of more general validity and perceptually real.

**interaction between some factors in syllable nucleus durations**

Above, a discrepancy was noticed between the quantitative effect of the number of syllables in the word on stressed syllable nucleus durations as measured articulatorily in simple nonsense words and as found in perceptual adjustments of syllable nucleus durations in synthesized common words. The cause of this discrepancy appears to be the different stress patterns of the words used in the two situations. The nonsense words had only one word stress, on the initial syllable, the three and four syllable words in the perceptual experiments had two stresses, a primary stress on the initial syllable, and a secondary word stress on the third.

One may notice that the unit within which the duration of the stressed syllable

---

*Fig. 4. Preferred adjusted, spoken and calculated durations of stressed /\( a: \)/ and /\( o/\) as a function of the number of syllables in the word counted from the beginning of the syllable concerned. Calculated values were obtained from rule (1) with \( D = 200 \) ms and 100 ms for long and short vowels respectively, and \( \alpha = 0.2 \) for both.*
nucleus durations is determined has until now been assumed to be the word. There is some controversy, however, in the phonetic literature whether the word or the foot (defined as a stressed syllable plus all the following unstressed ones) is the relevant unit for the description of the timing phenomena under discussion (See e.g. Abercrombie, 1964). It has been pointed out by Elisabeth Uldall that the data presented earlier in this paper could also be explained in terms of foot isochrony. If we now assume that a word like /mɑːtələbɔːz/ consists of two feet, each two syllables long, we would not expect to find the same syllable nucleus durations in the initial syllables. To study this question further, some simple measurements were performed on the following words:

<table>
<thead>
<tr>
<th>Word</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>/mɑːtə/</td>
<td></td>
</tr>
<tr>
<td>/mɑːtətə/</td>
<td></td>
</tr>
<tr>
<td>/mɑːtətətət/</td>
<td></td>
</tr>
</tbody>
</table>

These words are somewhat closer in their phonemic structure and stress pattern to normal Dutch words than the nonsense words used earlier. They were spoken by one speaker, the present author, each word six times, in a quasi random order, under optimal recording conditions. A tape recording was made and high quality normal and duplex oscillograms were made. The durations of all acoustic segment durations were measured by hand. The standard deviations were generally between 5 and 10 ms.

The results are presented schematically in Fig. 5. These results are relevant to the controversy between word and foot isochrony. If the word was the relevant unit within which the compensatory adjustment of syllable nucleus durations took place, one would expect the durations of the stressed initial syllable nuclei in the second and third words to be equal, just as the durations in the fifth and sixth words, because in those four words the number of syllables in the word is equal. If, on the other hand, the foot would be the relevant unit for compensatory adjustment, one would expect the durations of the stressed initial syllable nuclei in the first and second words to be equal, and those in the fourth and fifth words, because in those four cases we have feet of two syllables. The results rather indicate that both the word and the foot are appropriate units for the description of timing phenomena. We may express this formally by changing the formula (1) used earlier in the following way:

\[
V = D/(m^a \cdot p^b) \tag{2}
\]

in which \(V\) is again the duration to be calculated, \(D\) is a standard duration being different for long and short vowels, \(m\) is the number of syllables in the word counted from the beginning of the syllable concerned, and \(p\) is the number of syllables in the foot counted in the same way. \(a\) and \(b\) are positive and smaller than 1 and restrict the effects of \(m\) and \(p\) respectively. The dotted lines in Fig. 5 connect values calculated from formula (2) with appropriate values for the parameters (see caption to Fig. 5). The above example showed an interaction between two prosodic factors acting on stressed syllable nucleus durations. The following example concerns a case of interaction between intra- and extra-syllable factors.

Test materials were constructed from four syllable types, viz. /mɑːm/, /mɛːp/, /mʌm/ and /mɪp/, followed by a varying number of unstressed syllables, in the following way: /mɑːmətətətət/. These syllable types were found to show extensive differences in syllable nucleus durations from type to type when spoken as monosyllables in isolation (Cf. Fig. 1). The test words obtained were spoken six times each by the present author in optimal recording conditions, and high quality normal and duplex oscillograms were made. The stressed syllable nucleus durations were measured by hand. The standard deviations were generally between 5 and 10 ms.
Fig. 5. Schematic durational build-up of Dutch nonsense words with different stress patterns. Acoustic segment durations were measured in oscillograms, each presented duration being the mean over six individual measurements. The dotted lines connect values calculated for the stressed initial syllable nuclei. These values were obtained from rule (2), with $D=200$ and 100 ms for long and short vowels, $a$ and $\delta$ are 0.25 and 0.13 for long vowels, and 0.2 and 0.1 for short ones.

The results are presented in Fig. 6. It may be seen that the effect of growing number of syllables coming later in the word decreases as the duration of the syllable nucleus in the case of the monosyllable decreases. This is not only true for the absolute effect (in milliseconds) but also for the relative effect (in percent). In the case of the very short syllable /mɪp/ the shortening effect has nearly vanished.

We may interpret this as a greater resistance against prosodic shortening for inherently shorter syllable nuclei than for longer ones. To describe these data quantitatively an algorithm has been constructed that takes as input a sequence of phoneme symbols, word boundaries and stresses being indicated. The algorithm for each stressed vowel computes the number of syllables in the word end and to the next stress. It determines the character of the following consonant and assigns an inherent syllable nucleus duration to the vowel. Then the power exponents of rule (2) are appropriately determined as a function of the inherent syllable nucleus duration.
Finally, the actual syllable nucleus duration is calculated from rule (2). The inherent syllable nucleus duration is obtained from the data presented in Fig. 1. In Fig. 6 the solid lines connect values calculated from this algorithm, with a correction for a difference in tempo applied to the inherent syllable nucleus durations as obtained from Fig. 1.

To see whether the algorithm has any more general validity, a new series of measurements was carried out in the same way as the preceding ones, this time with the syllable types /mːːm/, /mːːt/, /mːːm/, /mːːt/ and /mːːm/, and a considerably faster tempo. Again an appropriate correction for tempo was applied to the inherent syllable nucleus durations. Nothing else was changed in the algorithm. The results of these measurements together with the calculated values are presented in Fig. 7. The fit seems to be quite satisfactorily.

---

Fig. 6. Syllable nucleus durations for stressed initial syllables of varying phonemic structure as a function of the number of unstressed syllables remaining to be produced in the word. Each mark gives the mean of six measurements. The curves connect values calculated from an algorithm described in the text.

Fig. 7. As Fig. 6 for new test materials. Calculated values were obtained from the same algorithm.
discussion

The main point of the present paper is to demonstrate that a possible way of achieving a rule system for the description of quantitative interactions between various factors acting on syllable nucleus durations in real speech, is to start from very simple nonsense words and then to proceed gradually to more realistically structured words. This naturally is a step-by-step procedure and only some minor steps have been taken. Even the results found in these first steps are not very well established because the amount of data on which they are based is rather limited. The results may suffice, however, to show that this kind of procedure works satisfactorily and may lead to useful quantitative rules, which can be subjected to further tests, and can be extended to cover new findings. A clear advantage of the present procedure is that it first leads to the description of factors in syllable nucleus durations in isolated and presumably optimal conditions, so that the underlying patterns are revealed most clearly, and only then shows how the quantitative effect of these factors interacts with that of others. Thus in the resulting rule system the various factors are presented in an abstract, non-polluted form. This may help in phrasing new and interesting questions concerning the origin, and the universality or language specificness of the various factors.

Many problems arise with respect to the perceptual functioning of the temporal factors in syllable nucleus durations. The strong interaction of many factors seems to obscure the perceptual identity of each of them. On the other hand, the interacting factors together put rather heavy constraints on the number of possible temporal patterns in the language. It may be hoped that the explicitation of these constraints in the form of a rule system for the temporal patterning in speech may help to further future investigations into the perceptual functioning of these temporal patterns.

summary

Research into the temporal organization of speech is hampered by the strong interaction between many factors acting on measurable durations. In this paper a possible way is proposed for approaching a rule system for the temporal organization of speech. This may be done by starting from the study of very simple nonsense words, in which single factors in nonsense words are studied in optimal conditions, and then, combining these factors in nonsense structures, gradually becoming more complex and closer to real linguistic structures in the language, see how these factors interact quantitatively. The approach is exemplified in a series of measurements on syllable nucleus durations.

The durational effect of the following consonant on a stressed syllable nucleus duration is studied, and it is shown how this effect interacts with that in position in word and foot. Also some data are presented on the perceptual reality of some prosodic syllable nucleus durations. The quantitative interaction between the various factors can be described quantitatively in the form of a simple algorithm.

references


Fischer-Jørgensen, E. Personal Communication.


Uldall, E. Personal Communication.
ON COARTICULATION OF ARTICULATORY EFFORT AND SYLLABLE BOUNDARIES

I.H. Slis

In previous reports we proposed a model on the relation between effort and the timing of articulation. In this model it was assumed that speech sounds, articulated with more effort compared to those articulated with less effort, are advanced in time with respect to the innervation of the muscles that perform the initial closure. (Slis, 1969, 1970, 1971, 1972).

In the literature, tendencies can be found that higher muscle activity is not restricted to the closing gestures of the vocal tract, but it can be shown also to be present in other muscles involved in the production of that particular speech sound. Literature on this point reviewed in Slis, 1969, 1971 and 1972 indicates a 'wave' of higher innervation affecting the complete set of muscles in operation. If the effort radiates spatially to other muscles, one might wonder whether effort radiates in time just as well. In other words, is it possible that indications of higher effort can be found in gestures belonging to other speech sounds in the temporal environment?

If the latter is the case, it may be asked whether this effect is restricted to the gestures of the same syllable or whether it is stopped by syllable boundaries. We specifically ask ourselves whether traces of effort are to be found in preceding sounds or syllables with regard to an advancement of the timing of articulatory gestures.

Before trying to find an answer to this question, we think it useful to define the syllable, since it is a point of discussion in the present literature. Three different categories of speech production models can be distinguished with respect to the role assigned to syllables:

1. Models which have no need for syllables. We think we can incorporate the model of Henke (as described by Daniloff and Moll, 1968) in this category. This model describes articulation as a process that is independent of syllable boundaries.

2. The second group of models can be represented by the view of Chistovich and Kozhevnikov (1966). They assume that speech is programmed in terms of consonant-vowel (CV) groups.

3. The last category is derived from a more traditional view of the syllable, which is based mainly on intuitive and distributional data. In it we find open syllables (CV or V) with long vowels on the one hand and closed syllables (CVC or VC) with short as well as long vowels on the other.

In line with our attitude held in our previous reports on effort oppositions, where we tried to account for the intuitive concept of effort, we assume that the third category, which is based, among other things, on intuitive grounds, is a good working hypothesis. In accordance with this model the word /pa:pa:pa:p/ can be divided into:

*This report is part of a paper read at the Second Annual Essex Phonetics Symposium (Slis, 1972).

---

Fig. 1. Schematic representation of the build-up of a word of open (upper part) and closed (lower part) syllables.
two open syllables /pa:/ and one closed syllable /pa:p/, the word /papapap/ into three closed syllables /pap/. This is visualised in Fig. 1. In the words with the open syllables the end of /a:/ is formed by the beginning of the next syllable, which leads to a concept of overlapping syllables during production. In words with the short vowel /a/ the syllable boundary lies somewhere within the lip closure of /p/.

We assume that speech production is organized on different successive levels. On the lowest level the syllables are the units for the organisation of speech sounds. In a model like this, it seems reasonable to expect that the temporal organisation of speech sounds within a syllable shows more cohesion - and thus less variability - than speech sounds belonging to different syllables, since the latter belong to different units of organisation. This leads us to expect that the articulatory segment durations which contain a syllable boundary show greater variability than those which do not.

This assumption on temporal cohesion may be applied to the hypothesised difference with respect to syllable boundaries between /pa:pa:pa:p/ and /papapap/. We may now expect on the one hand that in the first word the vowel durations (/a:/) show higher standard deviations than the /p/ closure durations, since the ending of the /a:/ is constituted by the closing movement of the following /p/; the opening and closing of the /a:/ belong to different syllables, and the closing and opening movements of /p/ belong to the same syllable. On the other hand we may expect in the second word (/papapap/) that the /p/ closure durations show higher standard deviations than the opening durations of the vowels (/a/), since the closing gesture of /p/ now belongs to another syllable than the opening gesture, and the opening and closing gestures of the vowel belong to one closed syllable.

This hypothesis was checked using experimental data obtained from previous measurements (Nooteboom and Slis, 1970) on three subjects. Each word was spoken 50 times. The measurements were done by means of lip contacts, which formed an electric circuit when the lips were closed (Willems, 1970). Indeed, in the first case we find that the standard deviation of /a:/ durations is higher than that of the embedded /p/ durations, and in the second case that the standard deviation of the embedded /p/ durations is higher than that of the preceding /a/ durations (Fig. 2). This trend was confirmed by measurements on words containing /o:/ and /ɔ/, /ɛ/ and /ɪ/, and /ʌ/ and /œ/. These data are not further reported on in this paper.

**Fig. 2.** Standard deviation in the measured duration of articulatory segments in words of the type /pVpVpVp/ in which the results obtained with the long vowel /a:/ and the short vowel /a/ are given separately.
If we compare words with a long vowel in a closed syllable, like /pa:tpa:tpa:t/ and /ta:pta:pta:p/, with similar words with /a/, we may expect that the standard deviation of both vowels will be lower than in the consonant clusters, since in both cases the syllable boundary lies within the consonant cluster. Each of the four words was spoken 40 times by two subjects. Moments of lip opening and closing and tongue opening and closing were measured by means of lip contacts and tongue contacts; the latter were mounted on an artificial palate. With this method of measuring the overlap of the successive syllables is measurable, being the period during which the lips in /p/ as well as the tongue in /t/ (belonging to different syllables) are close (Fig. 3).

Fig. 3. Schematic representation of the buildup of words of closed syllables with long and short vowels.

The results show that our expectation was justifiable (Fig. 4). Although these results can be used as additional evidence for our choice of syllables, we are still aware of the fact that it is a highly speculative choice.

Returning to our question whether effort is phoneme of syllable bound or not, we see in Fig. 5 that the opening gesture of the /d/ in /bapce/ is advanced as compared with /babce/. Effort of phonematic origin (voiceless consonant /p/) seems to be anticipated in the motor command of /a/, which belongs to another syllable than /p/. This strongly suggests that the higher effort of /p/ is not restricted to /p/ only, but is already anticipated in the preceding syllable.

The measurements show that the duration of the /a:/ in /pa:pa:pa:p/ of the first (unstressed) open syllable /pa:/ is shorter when the second syllable is stressed than when the third syllable is stressed (Fig. 6). This can be interpreted as an increasing overlap of the advanced stressed second syllable over the first (unstressed) open syllable. However, the same shortening of the vowel of the first (unstressed) syllable can be shown (to a lesser degree) in word pairs with the closed syllables /pap/.

In the words /pa:tpa:tpa:t/ and /ta:pta:pta:p/ we find that the /a:/ and the initial consonants (/p/ or /t/) of the first syllable are both slightly (about 6 ms) shortened if the second syllable is stressed (Fig. 7). This effect holds for all the eight comparisons made (2 subjects, 2 series of 20 measurements, 2 word types). In the latter cases the /a/ or /a:/ is not limited by the initial gesture of the second syllable, but by that of the final consonant of the same syllable. This shortening is therefore not due to increasing overlap caused by an advancement of the second syllable.

Fig. 5. Schematic representation of the buildup in time of the word pair /bapce/, /babce/ differing in voice character of the embedded consonant only.
These data show an influence of stress on the timing of the unstressed syllable preceding the stressed one. Therefore we think it justifiable to assume that the higher effort of stress is anticipated in the preceding syllables. On the basis of the data presented we would propose a model in which an increase in effort of phonematic origin as well as of prosodic origin is translated into an advancement of the articulatory command. The increase in effort is not restricted to the speech sound or syllable that causes the effort increase, but is anticipated in a coarticulation-like manner in the preceding commands.

**summary**

A syllable model with closed (VC or CVC) and open syllables (V or CV) is proposed. Closed syllables may contain long as well as short vowels, open ones may contain long ones only. It is assumed that two articulatory gestures belonging to one syllable are more closely related in the time programme of articulation than two gestures belonging to different syllables. Moments of lip- and tongue opening and closing were measured.
Intervals between two of these articulatory events show larger standard deviations when they belong to different syllables, than when both take place in one syllable, which is in accordance with the assumption. Previously we have shown that linguistic effort is correlated with an advancement in time of the initial gesture of a speech sound with effort. It is shown that this advancement is not restricted to the initial gesture of the speech sound at issue, but that it extends to preceding gestures as well. Syllable boundaries impose no constraints on this effect.

references

3 VISUAL PERCEPTION
INTRODUCTION I.P.O. RESEARCH ON VISION 1972

H. Bouma

This introduction intends to augment for the reader the structure of our 1972 research on vision by briefly sketching its perspective, also of research not explicitly dealt with in the present progress report.

In VISUAL DYNAMICS, the beautifully simple relationships between deLange flicker-fusion curves and flash thresholds have now been published in Vision Research in one paper on experiments and one on theory. Shortly, a third paper will appear in which thresholds for double flashes are shown to fit nicely the proposed theory. The low frequency portions of deLange curves, representing differentiating processes, still contain some unsolved problems and a rather fundamental difference between the dynamics of brightness vision (slow) and the dynamics of flicker vision (fast) will perhaps have to be made. If so, it should be concluded that the essentials of brightness dynamics are still rather obscure.

On SLANT PERCEPTION, a 19 minute movie has been finished, which intends to bring to a wider audience the reality of a piece of research, without becoming inaccurate by undue popularizing.

In eccentric vision, masking interactions occur between parallel lines, as reported in the 1971 issue of our Annual Progress Report, and these data have been corroborated this year. At the boundary of neuro-physiology and psychology of perception there are intriguing problems on how the multitude of excitations, produced by a single retinal stimulus, converges towards a single perceived object, such as a single line. A few years ago we built a hard-ware slant detector for this purpose, and for continuing the work in soft-ware for more complex configurations, we have completed an apparatus which links a 16 x 16 matrix of phototransistors to our P 9202 computer.

READING processes, as approached from the side of research on vision, can roughly be divided in eye saccade control, form recognition, and central integration. Saccade measurements carried out this year have failed to reveal substantial correlations between durations and sizes of successive reading saccades. Last year, we reported that when jumps of retinal images are under experimental control, reading is quite possible. It seems to follow that individual saccades are not controlled by proceeding recognition, but, less critically and more rapidly, at a lower level. Assuming word recognition to be a constituent of reading, we have an active interest in those perceived word properties which trigger word perception. Certain letters may perhaps function as such, particularly initial and final letters (paper forthcoming in Vision Research), but also global properties of words, whatever these may be. We are interested in what these may be and are studying the factor of word length. It has become clear by now that recognition of words slightly off fixation is not limited by visual acuity but by adverse interactions between line? stimuli. It is conceivable that certain reading difficulties find their basis in a high degree of such interference.
Visual CONSPICUITY, or prominence, of objects in their background seems to us an important factor in normal vision, since it can be assumed to be a determinant of the line of regard. The complex, or at least inhomogeneous, background is essential, and this is where research on conspicuity deviates from most of the visual experiments described in the literature. The influence of the background can be described as an adverse interference. By experimenting with a number of different subjective criteria of noticing a test object, it has now become clear that part of this interference is specific for the perceptual dimension involved: luminance-specific interference and size-specific interference have been demonstrated. Apart from this, there is also interference to which no clear-cut specificity can be attributed so far.

For coloured stimuli at rather low luminances, we have found that chromaticity contrast is secondary to luminance contrast. In that colour widens the so-called conspicuity area only if luminance differences are low.

The world of the VISUALLY IMPAIRED held our attention this year, since we had analyzed a close mutual contact as a prerequisite for the fruitful development of technological devices for improving adaptability and independence of those whose vision is poor or absent. We describe separately the factual progress made with two such devices, viz. a TV magnifier 5-25 X, now in production, and a drawing board for embossed relief figures. Also, we try to analyze the gap between research effort and devices actually available.

As in earlier years, many problems of directly APPLIED nature came to our attention in which visual perception of industrial products or tasks was critically involved (legibility, detail visibility, TV flicker, etc.). A few of these are reflected in separate contributions. It is fortunate that a number of these problems, often too complex for theoretical analysis, can be satisfactorily solved by direct experimentation, on the basis of existing knowledge.

TECHNOLOGICAL DEVICES FOR THE VISUALLY HANDICAPPED: GAP BETWEEN RESEARCH EFFORT AND AVAILABLE AIDS

H.Bouma, F.L.Engel and H.E.M.Mélotte

introduction

There is a general awareness that technology should come to the aid of people handicapped with perceptual or motor deficits. If such a handicapped individual happens to find his way to a technological research institute, he is typically met by inventive enthusiasm and after many hours or days of hard work, a useful apparatus may result which is put at his disposal free of cost. In addition to this, quite a few institutes manage to run a project aimed at constructing certain devices on their own initiative.

Despite of this, the majority of such handicapped people have only limited profit from advanced technology. Catalogues of available aids certainly do not reflect much research effort. As seen from the side of the research, reports and conference papers typically concern "prototypes under construction" or "prototypes to be tested shortly". This would suggest improvement of the situation in the near future, were it not for
the fact that this research situation seems rather stable over the years with a few notable exceptions. In the wake of this illusory progress some reflection on the gap between research effort and available aids may be useful.

In fact, we entered the problem area much in the way described. Over the years, a number of devices have been constructed in our Institute, most of them for a single handicapped person only. Examples are a guiding plate and an automatic sheet inserter both for a typewriter (Vredenbregt, Van den Ban, and Mélotte, 1968), and a Braille punch for labeling tape. A few projects of a somewhat wider scope, such as a punched-tape controlled Braille writer, did not reach a stage of sufficient reliability. One notable exception, for people with motor handicaps, was a nerve stimulator for drop-foot patients, which is now commercially available (Vredenbregt and Noordermeer, 1967).

When, in 1971, we improved a relief-drawing technique for the blind and put together a closed circuit television set for enlarging print ("TV-Magnifier"), we realized the regrettable inefficiency of testing the usefulness of the apparatus for a certain individual only rather than the usefulness for the general category of people concerned. In consequence, we discontinued a long term project on tactile image perception and spent the available time on testing two potential aids on a larger scale, on adapting them for a more general use, and on pulling them towards general availability. These projects concerned the above TV-Magnifier and a drawing system for embossed relief figures. Before describing our 1972 efforts in more detail, we shall first briefly survey the field for the visually handicapped, and outline some of the problems in a more general way.

visually handicapped: needs and aids

The term "visually handicapped" is understood to mean those persons whose vision is functioning below normal. Such deficiencies may range from total blindness to a level at which, with optical aids, still some reading remains possible. Broadly speaking, these visual impairments can be more or less characterized by a low visual acuity and/or a diminished visual field. The causes of impairment are manifold, congenital deficiencies, (traffic) accidents and (neural) diseases among which many concomitants of old age.

Two major aspects of present day life which cause difficulties for those with gross sight deficiencies are mobility and reading. Both involve an obvious loss of independence for the individual. The most relevant developments in visual aids are therefore related to these two abilities. In 1970, Nye and Bliss published a very informative survey on these topics. We shall indicate briefly recent developments in this field, with special emphasis on those that crossed the gap between prototype and generally available aid.

For reading and writing the braille code has been in use for about 150 years. A good braille reader, using (contracted) Braille II, can read at speeds of 200 words per minute. This illustrates that information can be transmitted through the tactile sense at rather high rates. More usual speeds are however 75 words/min. for Braille Grade I and 125 words/min. for Braille Grade II. For comparison: speeds of normal (silent) reading are 200-800 words/min., whereas e.g. talking books are recorded at an average speed of about 180 words/min. A speed below 20 words/min. or so is considered insufficient for interested reading and acceptable only for certain special tasks, such as reading bank cheques.
It should be noted that not all visually handicapped can read braille effectively. According to a survey of Grey and Todd (1968) covering the registered blind of England and Wales, only 40% of those aged between 16 and 64 could read braille effectively. Moreover, less formal materials like notes, letters and newspapers are generally not available in Braille. A typewriter-adaptor, supplying a simultaneous Braille copy has recently been announced (Phylab Brailler).

Ideas and efforts on direct conversion of inkprint text into audio or tactile signals date from the beginning of this century. Only recently has one of those devices, the "Optacon" (Linvill and Bliss, 1966), become commercially available. With this promising reading aid, producing an enlarged tactile facsimile display on a single finger, reading rates ranging from 20 to 70 words/min. have been achieved. By means of a small optical probe the Optacon translates an area of about the size of a letter of into a similar tactile image produced by a 24 x 6 array of small piezoelectrically driven rods.

Teaching programs, further applications, and possibilities for financial support, are currently under study (Optacon Newsletter no. 3, November 1972). Other reading aids for direct audible or tactile translation including reading machines based on more or less refined print recognition giving (spelled) speech output, are only at the laboratory stage. The possibilities of a relief drawing technique in which characters drawn, immediately appear embossed at the working side, will be described elsewhere in this contribution.

Situations for which mobility aids are required are considerably more variable than those which confront a reading aid. Major requirements for a mobility aid are adequate warning and reliable service. In this respect the cane can hardly be surpassed by an electronic aid as to simplicity, effectiveness and reliability. Therefore the devices currently under development aim at supplementing the cane. In general these devices are based on the emission of light or ultrasonic radiation and the detection of the reflected energy.

The ultrasonic torch developed by Kay (1963), which is in fact the only commercially available, transmits during short pulses an ultrasonic beam, with frequency sweeps from about 90 kHz to 40 kHz. The echo received is mixed with the transmitted signal and the resulting difference frequencies in the audible range are supplied to an earphone. Within a radius of a few meters, direction, distance and structure of surrounding objects can thus be estimated. In spite of the enthusiastic response of users of the torch, this aid does not yet seem to be in general use. Further developments, such as similar ranging devices in spectacles or in the cane, which leave the hands free, are "under development".

A general problem in designing sensory aids is to achieve an intimate interface, auditory or tactile, between the user and his aid. The work of Starkiewicz and Kuliszewski (1965) in improving these interface aspects should also be mentioned. They built a portable optical to tactile image conversion system with 144 tactile stimulators on the forehead, for scanning the environment, as did Bach y Rita et al. (1969) with 400 stimulators on the subjects back. However, these devices are not commercially available. Direct electrical stimulation of the visual cortex, causing electro-phosphenes, has recently been tried by Brindley and Lewin (1968), which research is in fact far from a practical realization.
So far, we described some of the most promising devices. However, one should be aware of the large number of developments that failed in spite of the technological ingenuity involved. In fact, problems of a different nature must be solved as well, such as evaluation, training, distribution, and financing. These will now be discussed.

**From laboratory model to industrial product**

The motive for pushing a certain apparatus towards general availability is its usefulness which should therefore first be assessed. In the stage of a laboratory model, a serious though subjective assessment by one or two people may suffice. If, however, production is considered, a more general and objective test is needed for assessing the usefulness of the apparatus in regard to the various categories of handicapped people who might profit from it. Also, information on the drawbacks of the laboratory model is needed for finding a good ergonomic design.

Therefore, a substantial number of potential users should try out the apparatus seriously, and sufficiently long to pass through a possible initial learning phase. It will be clear that contacts with a few handicapped persons cannot fill this need. Closer contacts are necessary with institutes and organizations of the handicapped and with the medical profession involved. In case of a loose organizational structure, this may prove to be a time-consuming activity. Let us assume that the prototype turns out useful in such a test, and that a number of shortcomings are revealed. Then, an improved prototype must be built. If, however, this prototype is actually intended to be produced, quite different requirements should be met as well. Materials to be used must be inexpensive, labour minimized, and tools for the production must be available in the actual production unit. At this stage, it should be known where the production will take place, especially since the rather small number of devices may lead to rather small production units with restricted possibilities. So contacts with a production unit come relatively early. Prior to production, the financing problem should be solved. This is in fact the task of a distributor, who should order the production, assess series-size and price, organize delivery and service, and distribute information. In case of a small potential market, commercial eagerness will perhaps have to be supplemented by social zeal. Needless to say that testing, development, production, and distribution need efficient coordination, which of course is rather time-consuming.

**Nature of the gap**

The observation that a number of technical instruments do not pass the stage of a laboratory prototype, can now be understood. The construction of a laboratory model constitutes only a fraction of the work necessary for making a device generally available. Moreover, the remaining work is of a much different nature, and falls outside the walls and usually beyond the scope of most research institutes. What is needed are strong contacts with the wider world of the handicapped involved, with producers and with distributors. Also needed is an appropriate industrial competence for the development of a suitable prototype. This comes as no surprise to those involved in industrial production. The problem is, in fact, that the drives which ordinarily lead to new products, are much weaker if the number of possible "consumers" is small, their purchasing power restricted, and most importantly, their "market" difficult to reach. In such a situation, there seems a pressing need for an initiative by the government, the industry or others, aimed at organizing professional activities of competent people to this end, in order not to let the disabled being unnecessarily charged with extra
loads in our society based as it is on making profit through mass production. A frame­work should be created, within which devices can be constructed, tested, developed, produced, and distributed through smooth channels. Developments in other countries can then be noticed early and followed up at short term, evaluated along existing pathways, adapted to local circumstances, and, where necessary, accompanied by suitable tests and training. Grateful as one should be for the present initiatives and financial contributions of individuals and organizations, a firmer base is necessary if handicapped people are to be served by technical aids to a degree acceptable for a human-directed, and wealthy society.

i.p.o. activities in 1972

TV-Magnifier 5-25 x

In the previous issue of the IPO Annual Progress Report, we described a laboratory model of a Closed-Circuit Television set for enlarging print (Van Dun, Bouma and Ale­wijne, 1971). Advantages of such a system over conventional optical magnifiers are higher magnification, a large field of view, sufficient brightness, acceptable image quality, and uncritical eye position. Early in 1972, we established contacts with institutes for low-vision people and during 1972 a prototype was tested in three of these institutes for periods of 2 - 4 weeks. Reactions were generally favourable. In April, we had the good fortune of coming into contact with a possible producer and a possible distributor. Starting in May, we had regular meetings with them. In the same month, an industrial prototype incorporating as many suggestions from users as possible, was constructed in our workshop.

In particular, controls were much simplified, dimensions optimized, sturdiness increased and less expensive materials used. In early summer, production of a first series of 10 TV Magnifiers started and in October the first production apparatus was finished. In December, it was demonstrated at a meeting of Dutchophthalmologists and a press release was issued. The apparatus is now commercially available in the Netherlands from Philips. (Fig. 1,2).

On the basis of a questionnaire, we have now reports of over a hundred low-vision people, most of them with visual acuity below 0.2. From this, we derive that the TV Magnifier is a useful aid in reading for the majority (80%) of people with acuity between 0.015 - 0.15 or so, unless there are large deficits in the retinal center. For about 10 people it was a unique aid in a sense that they could read with the TV Magnifier only. These few may perhaps now shift from Braille to normal text. In case of very low acuity, a large 24" screen is to be preferred. Contrast reversal was preferred by some 10%, relatively more often for ocular opacities than for retinal diseases. A demand which we could not fulfil satisfactorily so far is a window through which only one line of print would be visible, so as to eliminate confusion caused by adjacent lines. So far we have not tested the usefulness for elderly people, but recent literature suggests it to be a useful apparatus for them as well (Genensky et al, 1972). As we learned in the course of the year an approach similar to ours had been taken earlier by others, in particular by Genensky in USA and by the Heidelberg Clinique in Germany (Blankenagel and Jaeger, 1972).
Fig. 1. (left) Final prototype of TV Magnifier.

Fig. 2 (below) Scheme of TV Magnifier with controls. Focusing occurs through shifting the camera, magnification can be adjusted via a motor-driven zoom objective.

1) contrast reversal
2) power switch
3) mirror
4) magnification
5) Focus
6) x-y platform
7) height platform.
As a test for knowing if a certain individual is helped by the apparatus, we follow Sloan's advice (1972) to let her or him read highly enlarged print, letter height between 10 and 50 mm. Next, a potential user should try the TV Magnifier at ease, using his own type of text. For this purpose, we were visited by some 15 low vision people during 1972. Governmental social agencies have already provided financial provisions in a number of instances.

relief drawing system

Basically this aid consists of a rubber-coated drawing board with special plastic writing-sheets. The drawings immediately are embossed on the working side of the sheet when writing with an ordinary ball-point with some pressure, so that no reversing is required. This technique is not new, the American Foundation for the Blind mentioned a "Raised line drawing Kit" in their catalogue (Clark, 1963). We discovered however, a better writing material with an imprinted texture, which did not stick to the fingers sensing the drawings (Engel and Van den Ban, 1970). This material was received rather enthusiastically. Thereafter, it was decided to improve the accompanying drawing board as well. The major applications of this drawing board were thought to be:

- as an educational aid emphasizing abstract functioning e.g. to be used for drawing, mathematics and geography;
- as a means for exchanging written messages between those with normal vision and people with visual impairments;
- as a computer programming aid, where flow charts can be illustrated with the help of a regular Braille writer;
- as a scratch pad for the elderly blind who find it difficult to learn Braille.

Our prototype was outfitted with an easy to handle magnetic clamping frame as the sheets should remain fixed during drawing. Tactile cm-divisions were made on the edges of the board. Together with additional sheets, a T-square, a triangle and a ball-point, the drawing board could be accommodated in a regular A4-writing case.

During 1972 we contacted a number of institutes for visually impaired in the Netherlands and tested the prototype at two primary schools for the blind, a rehabilitation center for visually handicapped adults and at a library for the blind. Since their responses were generally positive, a second prototype was constructed at the end of 1972 (see Fig. 3,4).

Improvements suggested during the testing have now been realized in this second prototype of a more simplified construction. Negotiations on production and distribution are presently in progress.

conclusion

In conclusion we propose that more emphasis be put on the pursuit of promising research projects with suitable testing, development, production, and distribution. In 1972 we have utilized our limited capacity for this type of work completely to these ends. It is our opinion that modern technology should be deliberately organized for bridging the gap between research efforts and available aids.
Fig. 3. Relief-drawing System, final prototype. A regular A4-writing case accommodates the drawing board together with additional writing-sheets, a T-square, a triangle and a ball-point.

Fig. 4. The easy to handle magnetic clamping frame. Owing to the tactile cm - divisions and a complementary provision in the T-square, this T-square can be shifted in discrete steps along the edges of the board.
references


CONSPICUITY AND COLOUR
F.L. Engel, H.G. Ligtenbarg and T.M. Bos

Introduction
In general only a small fraction of what can be seen during an eye fixation is actually perceived. Some objects attract attention and are consequently perceived, most others are overlooked unless attention is directed to them. So attention can be taken to perform a certain selective task.

The factors influencing selective attention can be divided into object and subject factors. We consider visual conspicuity to be an object factor, while, e.g. expectation can be conceived as a subject factor. In a previous paper (Engel, 1969) we proposed as a measure of conspicuity the size of the so called "Conspicuity Area", the retinal field in which the object can be discovered from its background during a single brief presentation. With structured backgrounds, advance information on object location generally increases the retinal field in which it can be perceived. Therefore, peripheral visibility and conspicuity appeared to be linked by directed attention (Engel, 1970, 1971).

Exploring object factors influencing the size of the conspicuity area, we reported previously on the combined influence of size and luminance contrast of a test disk among a number of mutually identical background disks (Engel and Bos, 1971). In these experiments no additivity was observed between the two factors. Depending on their relative magnitude, one of the two factors appeared dominant. This contribution reports on the effect of colour on the size of the conspicuity area.

Experimental
Experimentally it was convenient to use surface colour standards for the test disk, actually the chips of the Munsell Color System. In this system the samples are ordered subjectively into uniformly graded series of hue, saturation (Munsell Chroma) and brightness (Munsell Value). Fig. 1 shows the arrangement of chips and the Munsell notation.

Fig. 1. The Munsell Color System, specifying object colors on scale of Hue, Value (brightness) and Chroma (saturation). The Hue scale contains 5 principal and 5 intermediate hues, the Value scale contains 10 steps from black to white and the Chroma scale shows up to 16 steps from the equivalent grey.
The background of the test disk was a black sheet measuring 92 x 62 cm (brightness: Munsell Value 2.5) with randomly located identical white disks of 5 mm in dia (brightness: Munsell Value 9.5). The minimum center to center distance between the latter was 7.5 mm. One of the disks represented the test disk by inserting the relevant Munsell chip. Fig. 2 gives an impression of this test-disk/background combination.

![Fig. 2. The test-disk/background combination used in the experiments. The actual size of the black field was 92 x 62 cm, containing about 4000 white 5-mm background disks and one test disk (grey in this figure).](image)

The test field containing the test disk and part (40 x 50 cm) of the background was presented tachistoscopically to the observer by means of a semi-reflecting mirror for 100 ms, see Fig. 3. The test disk could be located at any point in the test field by moving the test-disk/background combination.

The luminance of the background disks as measured through the semi-reflecting mirror corresponded with 2.9 cd/m², while the black background part equalled 0.3 cd/m².

The colour temperature of the tachistoscopic illumination, again measured through the semi-reflecting mirror, was 3100°K, which is somewhat higher than that of the C.I.E. standard source A (2854°K).

The luminance of the rest field (0.65 cd/m²), presented for the purpose of adaptation when no test field was shown, corresponded with the average luminance of this field.

A viewing distance of 47 cm was maintained by means of a head rest, which simultaneously occluded the left eye of the observer. Consequently, the diameter of the 5-mm disks corresponded with 0.6° visual angle, while the size of the test field equalled...
The procedure corresponded to previously described conspicuity area experiments (e.g. Engel and Bos, 1971). The observer was informed in advance about the test-disk properties and had to discover it in a single trial. In the centre of the tachistoscope field, in front of the observer's right eye, a fixation spot was provided. When ready while fixating this spot, the observer controlled the moment of exposure. When he discovered the test disk, he indicated its location as a check. A training session preceded each area determination. In some instances the observer was also asked to mention the perceived chromaticity of the test disk. The size of the conspicuity area was expressed by $R$, the mean eccentricity of the border in 8 directions (horizontal, vertical and diagonal).

The results to be presented were obtained from the observers T.B. and H.L. Both had adequate visual acuity and colour vision. No spectacles were worn and no artificial pupil was used.

**influence of brightness**

First of all we determined the conspicuity areas as a function of the brightness (Munsell Value) of the test disk with the achromatic chips of the Munsell Color System. These chips divide the brightness scale into 10 subjectively equal intervals from black (Munsell Value : 0) to white (Munsell Value : 10), see Fig. 1.

The results obtained are plotted in Fig. 4. The results of the two observers agreed well. Self-evidently, the size of the conspicuity areas became zero when the brightness of the test disk equalled either that of the background disks or that of the black area. The size of the conspicuous area appeared to be maximum at Munsell Value 5.

![Fig. 4. Size $R$ of the conspicuous area as a function of brightness of the test disk, using the achromatic chips of the Munsell Color System.](image)

--- observer T.B.
--- observer H.L.
I standard deviation

**influence of chromaticity**

Fig. 5 represents, for several hues and 3 brightnesses of the test disk, the size of the conspicuous area as a function of saturation (Munsell Chroma). As the results of the two observers agreed reasonably well, only the averages are presented here.
Fig. 5. Influence of saturation of the chromatic test disk for several hues and 3 brightnesses (Munsell Value 5 = ▲, 7 = ■, 8 = ◆) on the size \( \bar{R} \) of the conspicuity area. The plotted points represent averaged results of the observers T.B. and H.L.

At Munsell Value 3 no reliable conspicuity areas could be obtained because the chromaticity of the test disk was hardly perceivable. At Munsell Value 5, where \( \bar{R} (V) \) is maximum (see Fig. 4), the influence of colour appeared to be small. Only with the Munsell Hues 5B, 5R and 5G, did \( \bar{R} \) increase somewhat with saturation. At Munsell Values 7 and 8, where brightness contrast relative to the background disks is less, the effect of saturation on \( \bar{R} \) is more pronounced. At Munsell Value 9 the available number of chromatic chips was too small to produce significant results.

The saturation steps (Munsell Chroma), which in foveal presentation are subjectively equal for all hues, were not equally effective in eccentric vision. The saturation steps in Munsell Hues 5B and 5R yielded larger increases in conspicuity area than, e.g. 5P and 5GY. This corresponded with the less saturated colour perception experienced with 5P and 5GY with increasing eccentricity. The hues 5B and 5R remained chromatic practically up to the borders of the conspicuity area, while 5P and 5GY did not. Slight changes in hue were also perceived with increasing eccentricity; red, for instance, became yellowish.

To explore the desaturation effect more closely, retinal fields were also determined in which the observer perceived the test disk as coloured. Fig. 6 gives the size of these fields for the principal Munsell Hues 5R, 5Y, 5G, and 5B, again averaged over both observers.
Fig. 6. Size of the retinal field at Value 5, in which the test disk was perceived as coloured (○) and the corresponding conspicuity area (□). The plotted points represent averaged results of the observers T.B. and H.L.

The borders of the retinal fields in Fig. 6, where colour could be perceived, were difficult to establish and, accordingly, relatively inaccurate. The results give nevertheless an indication of the influence of perceived chromaticity on $\bar{R}$.

**interpretation**

For achromatic test disks (Fig. 4) the conspicuity areas are mainly based on two different luminance contrasts, viz. relative to the white background disks and relative to the black area of the background. With luminance on a logarithmic scale, a symmetric function can be obtained (Fig. 7), in line with previous results of Engel and Bos (1971).

Fig. 7. Size $\bar{R}$ of the conspicuity area as a function of the test-disk luminance on a logarithmic scale, using the achromatic chips.
The colorimetric specifications of the Munsell chips are known under several illuminants including tungsten light (Nickerson and Wilson, 1950). The chips provide nearly uniform interval scales under day light illumination. Applying tungsten light, as we did, the resultant colour shifts are relatively small (Judd, 1956), because the colorimetric shifts are largely counteracted by chromatic adaptation. No simple relation exists, however, between the colorimetric specification of the chips and the subjective colour attributes hue and saturation. We therefore preferred to express our results in the independent variables of the Munsell Color System.

The interaction between contrast in brightness and in saturation of the test disk relative to the background disks is in principle additive since \( R = 0 \) only when both contrasts are zero. As a first approximation we assume here that a logical summation applies: \( R(V, C) = R(V) \odot R(C) \).

This means that the variable leading to the largest conspicuity area determines \( R(V, C) \), i.e. the size \( R \) of the conspicuity area as a function of Munsell Value (V) and Chroma (C). This suggests the possibility of selection between the two factors in the visual system.

Fig. 8. The experimental data of Fig. 5 approximated by \( R(V, C) = R(V) \odot R(C) \) being a hue dependent logarithmic function.

In Fig. 8 the logical summation has been applied, using the data of Fig. 4 for \( R(V) \).

In Fig. 8 we assumed \( R(C) \) to be a function of \( \log C \), with a hue-dependent weighting factor, see also Table I. The weighting factors were determined by straight-line fitting of the \( R(V, C) \) data as a function of \( \log C \).
No precise weighting factors could be determined for 5 GY and 5P, because for these colours no influence of saturation was found. As a cross-validation, the functions for $\bar{R}(C)$ in Table I appeared also to apply to the data in Fig. 6, which shows the retinal fields in which the test disk appeared coloured (see Fig. 9). The fit of these data is reasonably accurate.

Table I. Approximations of $\bar{R}(C)$ applied in Fig. 8.

<table>
<thead>
<tr>
<th>Hue</th>
<th>$\bar{R}(C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5B</td>
<td>13 log C</td>
</tr>
<tr>
<td>5G</td>
<td>11 log C</td>
</tr>
<tr>
<td>5YR</td>
<td>11 log C</td>
</tr>
<tr>
<td>5R</td>
<td>11 log C</td>
</tr>
<tr>
<td>5Y</td>
<td>8 log C</td>
</tr>
<tr>
<td>(5BG)</td>
<td>(13 log C)</td>
</tr>
</tbody>
</table>

The weighting factors reflect, for example, that blue (5B) was a more conspicuous colour in our experiments than e.g. yellow (5Y). Owing to the small number of 5BG samples, the corresponding weighting factor is inaccurate.

The luminances applied were rather low (between 3 and 0.3 cd/m²), nevertheless sufficient for photopic vision. For instance, Smith - Kinney (1958) established at as low as 0.03 cd/m² a regular foveal photopic luminosity curve, while at an eccentricity of 10° visual angle the shift from scotopic vision (with a peak-sensitivity near 510 n.m.) towards photopic vision (with a peak sensitivity near 540 n.m.) had been completed at 0.3 cd/m². In spite of this shift the periphery remains still more sensitive to the short and less sensitive to the long wave lengths than the fovea.

According to Weale (1953) this difference still exists at 300 cd/m². Since his spectral sensitivity curves, determined at several eccentricities, remained practically unchanged between about 5 and 300 cd/m², it is expected that our results also apply at higher luminance levels.

With respect to the conspicuity areas, blue, green, and red (orange) were the most effective hues.

Weale (1953) also demonstrated relatively higher sensitivity to these colours in eccentric vision. In line with our experiences with pigment colours, he reported,
even for monochromatic lights, hue changes and desaturation effects with increasing eccentricity of presentation.

summary

Explorative results have been obtained concerning the conspicuity of coloured test disks among white background disks. The size of the conspicuity area was taken as a measure for this property. The results indicate that:

When brightness contrast of the test disk relative to the background of white disks on a black surface is chosen optimally (viz. mid-grey, Munsell Value 5), the influence of test disk colour is slight. The colour influence is more pronounced when brightness contrast is low.

Although the Munsell Chroma steps are subjectively equal for all hues in direct vision, their effectivity is different with respect to the size of the conspicuity area. Relatively most effective are blue, red (orange) and green, while no effect at all has been established for purple and yellow-green.

The experimental data are interpreted in terms of a feature selection model.

references


REFLECTION PROBLEMS OF A PASSIVE LX DISPLAY PROVIDED WITH A FLAT MIRROR

H. Bouma and A. L. M. van Rens

In a passive liquid crystal (LX) display, application of a certain voltage causes a local separation in a thin LX layer. At these positions, light from outside sources is now scattered, mainly in forward directions. For sufficient contrast, it is necessary that the eyes receive these forwardly scattered rays. This can be realized either by illumination of the transparent display from the back (transillumination), or by illumination from the front if the light is reflected by a mirror immediately behind the LX layer. The LX layer will now be passed twice. We restrict this contribution to the reflecting LX display provided with a flat mirror.

A passive display offers the advantage of being visible under almost any prevailing illumination without special lamps, just as, for example, this text. However, the mirror behind the LX layer superimposes part of the outside world on the symbols displayed, and makes the value of the contrast and even its sign, dependent on circumstances as accidental as room illumination, wall reflection, window position and sunshine. We describe here a shield which prevents the reflection from degrading the contrast.

For ensuring legibility, all light reflected by the mirror towards the eye should originate either from the symbols or from a black surface, and under no circumstances from the outside world. Of course the black surface shields the display from being visible from certain directions. Design should be such that:

a) the angle under which the display is visible (display angle) is large and the angle shielded from view small.

b) the most backward part of the screen should send its rays in horizontal direction after reflection.

c) by relocating of the front edge to point C, perpendicular viewing directions are shielded from view.

d) combination of a and c in one figure. For $\alpha = 70^\circ$, $\beta = 50^\circ$.

Fig. 1. Relative positions of reflecting display and screen.

IPO annual progress report 7 1972
b) within the display angle, no part of the outside world is reflected towards the eye.

A solution is depicted in Fig. 1a - d. LX is the reflecting display with height $a$, at an angle $a$ with the vertical. If the lowest viewing direction is the horizontal, the most backward part of the screen should produce the rays in horizontal direction (Fig. 1a). At the highest viewing direction, at an angle $\beta$ with the horizontal, the front part of the screen should produce the reflected rays (Fig. 1b). However, the arrow in Fig. 1b indicates that the observer can look at the display from an angle larger than $\beta$ and then he will see part of the outside world superimposed on the symbols. This is not conform the second requirement, and, consequently, it should be prevented that the display can be viewed from angles larger than $\beta$, i.e. from directions perpendicular to the display. Therefore, we shift the front part of the screen to point C (Fig. 1c). The front part of the screen should therefore be located at the intersection of the direct ray from the upper edge of the display and the reflected ray from the lower edge, both at viewing directions closest to perpendicular. In fact, C is located right above the display center at a height $h = \frac{1}{2}a \cotg (a - \beta)$.

Fig. 1d combines Figs. 1a and 1c for $a = 70^0$, $\beta = 55^0$. If $a = 20$ mm. C is located at a distance of $h = 38$ mm from the screen centre.

Since the display cannot be viewed from perpendicular directions, the symbols appear lower, by a factor $\cos (a - \gamma)$ if $\gamma$ is the viewing direction. This can be compensated by a corresponding increase of actual symbol height at constant symbol width. For $a = 70^0$ and $\gamma = 40^0$, height should increase by 15%. The front edge of the screen should have the indicated position C. The back edge should intercept the indicated rays (Fig. 1a). Otherwise, there are no reflection requirements and sufficient freedom is left for a handsome design. The space between LX display and screen should be used for a suitable background.

If a large display angle is not a requirement, light may well enter from above (Fig. 2 : $a = 30^0$, $\beta = 20^0$).

In case the display can be given a slightly concave form, a smaller screen will do.

We have tried out the solutions of Figs. 1 and 2, in a laboratory set-up and found satisfactory legibility under a wide variety of outside illuminations. We could always maintain a luminance ratio between displayed symbols and background of a factor of 10.

Fig. 2. If only a small viewing angle is needed, illumination may come from behind the screen. $a = 30^0$, $\beta = 20^0$.

*The LX display was provided by Messrs Engelman and van Houten (Philips Research Laboratories).*
PREFERRED VIEWING DISTANCE AND SIZE OF THE TV PICTURE

J.J. Andriessen

introduction

Viewing distance is an important parameter in watching television or judging television picture quality. Literature often relates viewing distance to the visual angle of a separate line of the TV frame. With a sufficiently large viewing distance, the line structure of the picture is no longer visible on account of the limited resolution of the eye (Bodeker, 1962). At large viewing distances, however, the presented pictures are small, which is usually considered to be a disadvantage.

A question hardly discussed in literature is whether the common viewer perhaps has a subjective preference for a definite viewing distance. The viewer is perhaps not so critical as to the line structure being visible, in favour of a somewhat larger picture. The question of the preferred viewing distance in watching television has been investigated in the following for a number of picture sizes.

method

Experiments were carried out with 7 (monochrome) TV monitors of different size (625-lines). The height of the pictures varied between about 8 cm and 41 cm. Sometimes the diagonal of the picture is used. With the usual height-width ratio of 3a : 4a, the diagonal is 5a.

The TV monitors were lined up before a homogeneously illuminated background. The centres of the picture tubes were at eye-height of the seated subject. During each session only one TV monitor was visible at a time. The 7 TV monitors were presented in random order. They were so adjusted that their pictures were of corresponding luminances and contrasts. The luminances were between about 1 cd/m² ("black") and 50 cd/m² ("white"). The luminance of the background was 5 cd/m². These luminances agree well both with the usual values in domestic television viewing and those as applied in experiments on picture quality (Balder, 1957; Grosskopf, 1966; Prosser et al., 1964). The TV pictures were of five different types, including moving scenes ("Film"), stationary scenes ("Static") and lettertexts.

The lettertexts were of two types, viz. periodical text printed in Mercator 10 and TV display letters according to a design of Bouma and Van Rens (1971). All letters were in lower case, except for a few initial letters of sentences in the periodical text. The short letters (for instance a, o, etc.) had a height of 5 lines of the TV frame. The pictures were recorded on a video tape recorder under identical conditions. Hence it was possible to present the pictures reproducibly to the subjects. The order of presentation was always the same. The subject was seated on a wheel-chair for an easy adjustment of the preferred viewing distance.

The distance between the picture tube and the subject's eye, the viewing distance, was measured with an accuracy of 2 cm.

The experiments were carried out by 30 subjects, aged between 20 and 55 years. No television experts were among them. The subjects wore their spectacles, if necessary.

results

The results have been averaged over the 30 subjects. For each picture size, it appeared that no significant difference in viewing distance existed for the situation "Film" and "Static".

IPO annual progress report 7 1972
Neither was there a substantial difference between the results of the two experiments on lettertexts. As a result of this the experimental results of picture type "Film" and "Static" were pooled, just like those for the lettertexts. Abbreviated as results on "Picture" and on "Text".

Fig. 1. Preferred viewing distance as a function of picture height of 7 TV monitors, for two main types of pictures denoted as "Picture" and "Text". Vertical bars indicate 95% confidence levels.

In Fig. 1 the average of the preferred viewing distance V is plotted as a function of the picture height H of the applied TV monitor, for both "Picture" and "Text". With vertical bars, 95% confidence levels have been indicated. It appears from Fig. 1 that a linear relationship exists between viewing distance V and picture height H, both for "Picture" and for "Text". The standard deviation of the preferred viewing distances turns out to be relatively low. Mathematically, the relationship for "Picture" can be described with (V and H in cm):

\[ V = 6.75H + 40 \]

For "Text" holds: \( V = 6.75H \)

discussion

From a quantitative point of view the relationships found agree fairly well with literature. In C.C.I.R. report 405 - 1 (1970) a summary is given of viewing conditions for the subjective judgment of TV picture quality. The ratio viewing distance over picture height is recommended here as between 6 and 8, for so-called non-experts. Experts choose somewhat smaller distances. Jesty (1958) mentions 6 times the picture height as the preferred viewing distance, for TV pictures of 3 - 7 MHz bandwidth.
From the present experimental results the "viewing angle" can be determined, i.e. the visual angle of the TV picture, based on its height. The viewing angles appear to be between 5 and 7.5 degrees for the situation "Picture" and between 7 and 10 degrees for the situation "Text". These viewing angles are somewhat smaller than those found in the experiments on the preferred viewing conditions of projected Röntgenograms, which came out at about 12 degrees (Andriessen and Bouma, 1971). The main reason for the larger viewing distances with TV pictures might be their disturbing line structure. As in the experiments of Jesty (1958) with pictures of low bandwidth (< 3 MHz), the subjects choose larger viewing distances for acceptable pictures.

Yet, on being asked, our subjects generally choose viewing distances at which they probably still could perceive something of the line structure of the TV pictures. From the experiments of Bödeker (1962) it appears that the ratio $\frac{V}{H}$ has to be about 9.5 at a minimum for the line structure to be below the visual acuity threshold. The viewing conditions during his experiments are well comparable with ours. The line for $\frac{V}{H} = 9.5$ is shown in Fig. 1.

So in choosing the preferred viewing distance with TV pictures a compromise was made between "a scanty visibility of the line structure" on the one hand and, for instance, "a picture as large as appears optimal", or "legible text" on the other. In the experiments on "Text" viewing distances were chosen which were somewhat smaller than those found in the experiments on "Picture", probably because of the small height of the letters in the used text. At the preferred viewing distances a short letter subtends about 5 minutes of arc, roughly the threshold at which text can be read.

**summary**

This article deals with the preferred viewing distance of monochrome TV pictures. Experiments were carried out with 7 different monitor types, the height of the TV pictures ranging between 8 cm and 41 cm. Different types of pictures were used, including moving scenes, stationary scenes and lettertexts. Thirty subjects took part in the experiments.

The preferred viewing distance turns out to be roughly 7 times the picture height. Generally, it appears that the preferred viewing distance is a compromise between a distance at which some of the line structure of the TV picture still can be perceived and a distance at which a picture as large as appears optimal can be watched. In the case of text the preferred viewing distance appears to be somewhat smaller, probably as a result of the relatively small letter height.

**references**


ON THE OPTIMAL LUMINANCE TRANSFER FUNCTION IN TELEVISION

J.J. Andriessen

Introduction

In television the luminance transfer function is defined as the relationship between the luminances in the scene to be transmitted and the luminances of corresponding points in the TV picture. In designing the camera and monitor (receiver) system a linear relationship for a wide range of luminances is aimed at. Hence the luminance transfer function is a straight line with a given slope (Fig. 1).

The coefficient of slope is defined as $\gamma = \tan \alpha$. One would assume the most logical value to be $\gamma = 1$, i.e. the luminance relations in the scene are reproduced identically on the TV picture tube. There are a few aspects of watching television necessitating a somewhat larger value of $\gamma$. One important factor is the relatively small luminance range attainable on the picture tube. Another important factor is the fact that watching television often occurs at night with a relatively low surround luminance.

In the following these aspects will be discussed briefly on the basis of literature. The discussions are only related to stationary pictures.

Fig. 1. The relationship between the luminance in the TV picture and that in the scene.

Monochrome television

In a number of publications the value of $\gamma$ is discussed. Grosskopf (1963) states that a 1:1 relative luminance transfer of the "middle greys" is essential in order to get optimal TV pictures. So for this range of luminances $\gamma = 1$. His TV pictures were watched within a light surround of, on the average, equal luminances as that in the picture. Bartleson and Breneman (1967a, b) concluded from their experiments that $\gamma = 1.2$ for TV pictures watched in a relatively dark surround. Indeed, a $\gamma > 1$ seems to be plausible because of the fact that the luminances on the picture tube generally will be lower than that in the original scene. As a result of this, some loss of brightness contrast can result. This loss of contrast can be compensated by a higher luminance contrast on the picture tube, achieved with a $\gamma > 1$.

Luminance level of the surround

Grosskopf (1963, 1966) and Bartleson and Breneman (1967a, b) discuss the influence of the luminance level of the surround to the picture tube. The luminance of the surround is an important factor, because the state of adaptation of the retina is largely influenced by it. The state of adaptation of the retina has significant consequences for the brightness perception of details on the picture tube. From the experiments of Bartleson and Breneman (1967a) it appears that the brightness estimates of image points in photographic reproductions with a light surround follow a steeper function than those with photographic reproductions in a dark surround. The experiments of Jameson and Hurvich (1961) show corresponding results.
Here, the subject had to adjust the luminance of a matching field in order to match the brightness of a testfield the surround of which could have a number of luminance levels. At a high level of surround luminance the adjusted matching field luminances followed a steeper function than at a low level of surround luminance. Consequently, watching television with a dark surround requires a larger value of \( \gamma \) than with a light surround to compensate for the then resulting limited contrast range. A value of \( \gamma = 1.2 \) as found by Bartleson and Breneman (1967 a, b) is certainly acceptable. For TV pictures with a light surround \( \gamma = 1 \) may be sufficient. An important condition is here that the optical density of the monitor screen is such that incident light scarcely affects the luminance contrast in the picture (Grosskopf, 1963; 1966).

**colour television**

In colour television perception problems are still slightly more complex. As a result of the generally somewhat lesser sharpness of colour pictures compared with that of monochrome pictures, the brightness contrasts appeared to be lower (Bartleson, 1968). Secondly, the maximum luminance which can be realised on the picture tube is smaller than that of the monochrome picture tube, which may involve some loss of brightness contrast. In order to counterbalance these aspects the value of \( \gamma \) for colour television should be somewhat larger than that for monochrome television. A value of \( \gamma \approx 1.3 \) appears to satisfy (Wood and Sproson, 1971).

An essential point here is the similarity of the three colour channels with respect to luminance output. Any dissimilarity will be particularly of influence in the lower range of luminances with the effect of a more or less colouring of the dark greys (Grosskopf, 1967).

**summary**

The luminance transfer function of a television system is generally represented by a straight line in a log-log plot of luminance in the original scene and luminance on the picture tube. The coefficient of slope is \( \gamma \). The value of \( \gamma \) is discussed on the basis of literature. An important parameter appeared to be the luminance of the surround to the picture tube. For monochrome television and a dim surround \( \gamma = 1.2 \). For colour television \( \gamma \) has to be somewhat larger, \( \gamma = 1.3 \) seems satisfactory.

**references**


4 PERCEPTUAL AND MOTOR SKILLS
The studies on *stereotyped responses* as described earlier (Van Nes, 1970) have been continued, and several aspects have been investigated, viz. (a) Learning, by measuring reaction times and error score, (b) retention after two years, (c) influence of emphasizing response speed or error score, and (d) effect of order of learning one code and its reverse with various degrees of stimulus–response compatibilities (Van Nes and Van Schuur, this issue).

In a previous issue the effect of *advance information* in a *visual synchronisation task* (Koster and Boonstra, 1971) was studied. These studies have been extended to auditive stimuli. Two ways of presenting advance information have been tested, viz., a continuously increasing signal intensity that suddenly returns to its original value and then increases again, and the same procedure carried out with the pitch of the signal. These continuous ways of presenting advance information failed because they were very awkward for the subjects. Only a discrete way of presenting advance information appeared to be successful. Before the presentation of the stimulus to synchronize with, three weak and short signals were given with equidistant time intervals. The intervals I between the stimuli were varied (240, 480, 720, 960, 1440 and 1920 ms) as well as the intervals between the advance information signals (60, 120, 240, and 480 ms). As a control no advance information was provided. As with the visual signals, the help of advance information was greatest with maximum I. For I greater than 480 ms the best results were obtained with I/i = 4, i.e. the total interval is subdivided into four equal parts. As the best performance is obtained at I = 480 ms, a value of i = 480 ms will probably be the optimum value for interstimulus intervals greater than 2 seconds.

The studies on *handwriting* have been hampered by lack of an automatic device for analyzing displacement-time curves. A computer programme on the P 9202 computer which makes such an analysis possible, is nearly ready. In the mean time the study of writing-errors, as reported in the previous issue (Van Nes, 1971) has been continued by determining error frequencies for various letters of a written word and by making a comparative analysis of typing errors.

**Medical physics**

In a previous volume (Rau and Vredenbregt, 1970) studies on the M. biceps have been reported relating the *EMG* of the muscle to the *force* and the *endurance time*. As such measurements might have interesting applications, the studies have been extended to the back muscles (Vredenbregt, this issue).

The development of a *tremometer* as described two years ago, has been the impetus for some basic studies on *tremor* (Rau and Vredenbregt, 1971).

These have been continued, and extended to pathological tremor (Koster and Van Schuur, this issue).

In 1966 Schouten and Koster reported about high-frequency analysis (100–1000 Hz) of *electrocardiographic* (ECG) signals. In a more extended study, some years later, the ECG signals of a group of healthy subjects were compared with those of a group of patients with a myocardial infarction. No clear difference could be established. The ECG signals were picked up at the extremities. As the choice of the leads may
influence the high frequency components in the ECG, a new experiment was carried out. Three groups of eight subjects each participated.

(1) Patients with a myocardial infarction. The mean age of the group was about 60 years.

(2) Subjects of the same age without any diagnostic indication of an infarction.

(3) A control group of subjects of about 25 years without any diagnostic indication of an infarction.

Again no clear differences in high-frequency activity could be established. At the request of the Medical Systems Division of Philips an appliance has been developed and tested. On the basis of the actual heart rate of the subject, it enables information on the task performance to be fed back to the subject. This can be achieved either through an indication of the required speed of task performance, or, more directly, by altering the load of the task as a function of the heart rate. Such a device can be used for training cardiac patients (Van Schuur and Burema, this issue).

ergonomics

A highly important fact in the ergonomics practice is the moment within the development of a product or system at which the ergonomist is consulted. We have experienced that in an increasing number of cases the ergonomists of the IPO have been consulted at a very early stage of development. This is especially true for professional equipment and systems. Examples are the videophone project, telephone keying, and medical systems.

In the videophone project important contributions have been made to the trial network and its evaluation.

In behalf of the International Consulting Committee on Telegraphy and Telephony experiments have been carried out in various countries to study the effect of an additional column of four keys (for data transmission or facilities) next to the existing three columns of four keys each. In the Netherlands these experiments have been performed at the IPO and at the Netherlands PTT administration. The performance in keying conventional telephone numbers appears to be not adversely influenced by the four additional keys.

In 1972 a cooperation was established between the Medical Systems Division of Philips and the ergonomics team of the IPO. Studies have been made in cooperation with designers and clients. A simple technique that proved very useful is video-recording. In a hospital a recording is made of a system in operation. Comparison of the recordings of several existing systems often reveals various undesirable situations. This knowledge is used in the development of new systems.

references


Schuur, R. van and Burema, Th. (1972) On Controlling the Heart Rate, this issue.
SPEED AND ERRORS OF WELL LEARNED VERBAL RESPONSE SETS
F.L. van Nes and R. van Schuur

introduction

The method of paired associates has been used to study verbal learning since the end of the nineteenth century (Calkins, 1894). In these experiments a number of pairs, e.g. consisting of two words, is presented to the subject. Upon later presentation of the pair members which serve as stimuli, he has to respond with the associate member. His learning score is essentially determined by the percentage of associates recalled correctly, e.g. as a function of trials.

When attention is given to speed of responding, the paired-associate paradigm can be used to investigate a type of motor learning: the development of stereotyped responding to often repeated stimulus presentation. Such stereotyped responses may well occur in a variety of routine activities. The processes involved, from discrimination of the relevant stimulus to selection of the ultimate response, can, with repeated stimulus presentation, gradually become integrated and carried out under the control of "neural subroutines". Learning thus will be reflected in more efficient neural programming, which we will suppose to result in (i) faster responding, (ii) fewer errors.

An example of a stimulus-response association (S-R) of stereotyped character is to be found in reading, i.e. producing the letter-name responses R₁ to visually presented letter shapes S. Speed and correctness of this stereotyped association S - R₁ provide means to judge the development of a stereotype S - R₂ to be learned, although with the possibility of negative transfer from the existing S - R₁ (Gibson, 1941). Negative transfer would, in this case, cause slower responding and/or more errors. The amount of negative transfer from S - R₂ to learning another "new" stereotyped response, S - R₃, may be a criterion for the degree of development of a neural subroutine controlling the association S - R₂. Another criterion may be the amount of retention, since motor skills are known to be largely retained for a long time (Hovland, 1962).

In a previous communication (Van Nes, 1970) the time needed for reading lists of letters vs. coding them with number names was reported. The present paper deals with paced presentation of letters from the set S = {A, B, C, D} and verbal reaction time (RT) as well as error rate as dependent variables.

method

Stimuli consisted of capital letters, projected as luminous characters in the same place by a rear-projection read-out. The letters subtended visual angles of about 1° horizontally and 11° vertically. The stimuli were presented for about 100 ms through a mechanical shutter which opened fully in about 10 ms.

The subject was seated in a sound-proof box, viewing the stimuli through a window. He could start or stop stimulus presentations by means of two buttons. His verbal responses were recorded on tape. This facilitated their classification in (1) correct responses, (2) "full errors", i.e. fully pronounced wrong responses, (3) "half-errors", such as the response "thr .. two" when it should be "two". Actually, all responses were given in Dutch. Verbal RT's were measured with a voice key operated from a throat microphone.

IPO annual progress report 7 1972
An experimental session consisted of 320 trials, in which each stimulus was presented 80 times; all possible consecutive stimulus pairs occurred about equally often. With some exceptions, stimulus presentation was machine-paced, with an interstimulus interval of one second. Three response sets were used: (a) the letter names (A, B, C, D), (b) the number names \{1, 2, 3, 4\}, and (c) the number names \{4, 3, 2, 1\}. The first response set employed, \(R_1\), was that of the letter names, (a). The second response set \(R_2\), which subjects learned first and the third set \(R_3\), which they learned after \(R_2\), were either (b) or (c).

results

Experiment I: Development of stereotyped responses.

Three subjects participated. The only instruction to them was which response set they had to use. Figs. 1 and 2 show the data of the most extensively trained subject, VS. In Fig. 1 RT's, averaged over the four responses, are plotted as a function of the number of sessions for the three response sets, used in the order A-A, A-4, A-1. For different stimuli, no large differences in RT's were found. Fig. 2 shows the corresponding error rates. Although the data in Figs. 1 and 2 were obtained under self-paced conditions, further sessions under machine-paced conditions showed no significant differences. All other experiments were carried out under machine-paced conditions.

In 10 to 15 sessions RT's for A-4 had reached the RT-level for A-A. However, even after the response set A-4 had been learned during 69 sessions, i.e. some 5500 trials for each of the four stimuli, the error rate was still slightly above that for A-A.

After these 69 A-4 sessions, the subject responded according to A-1. In Fig. 1 RT's can be seen to remain virtually constant during 20 sessions, so, as regards response speed, no learning seems to occur. The error rate for A-1, however, in Fig. 2 is shown to decrease to the level for A-A in about 10 sessions.

Subject VS received no feedback on his RT's, neither during nor after a session. As to errors, no feedback was given during a session, but the subject was informed of his error percentage after each session. For that matter, he usually had a rather accurate impression of number and kinds of error he had made. The exclusive feedback on errors as well as a dislike to making them probably induced subject VS to try to minimize his error rate.

The two other subjects, DJ and KE, who used the response sets \(R_1\), \(R_2\) and \(R_3\), showed the same trend in their RT's: (i) after some learning, the average RT's for \(R_2\) and \(R_1\) were equal, (ii) the average RT's for \(R_3\) were higher than for \(R_2\), although this difference was smaller than in the case of subject VS. In contrast to subject VS, the other subjects made consistently more errors at \(R_3\) than at \(R_2\). Subject DJ, with the same learning sequence as VS, made 65 full errors in the first 21 sessions with \(R_2\), the set \{4, 3, 2, 1\}, compared with 264 full errors, i.e. 4 times as many, in the corresponding sessions with \(R_3\), the set \{1, 2, 3, 4\}.

Subject KE made 8 full errors in the first 21 sessions with his \(R_2\), being the set \{1, 2, 3, 4\}, and 130 full errors, i.e. 16 times as many, in the corresponding sessions with \(R_3\), the set \{4, 3, 2, 1\}. Subjects DJ and KE also tried to minimize their error rates for the same reasons as subject VS.

After the second learning process, A-1, subject VS once more used the response set \(R_2\): A-4, during four sessions. The increase in RT's and errors of sessions 72 and 73 shown in Figs. 1 and 2, can be interpreted as retroactive inhibition from the second learning process, A-1.
Fig. 1. Average reaction times against number of sessions, Experiment I. Three response sets were used, viz. A-A, A-4 and A-1 successively. The data for A-4 from sessions 72-75 were obtained after the 20 sessions with A-1.

Fig. 2. Error percentages against number of sessions, Experiment I. The error scores are made up of "full errors" and "half-errors", two half-errors contributing as much to the score as one full error.
Fig. 3. Average reaction times against number of sessions, Experiment II. Full symbols and solid lines pertain to the first response set A-4; empty symbols and dotted lines to the second response set A-1. Differences in shape of the symbols denote differences in the subject's instruction for the session concerned: circles = "minimize errors"; squares = "give equal attention to speed and correctness"; triangles = "minimize RT".

Fig. 4. Error percentages against number of sessions, Experiment II. The error scores are made up of "full errors" and "half-errors" as in Fig. 2. The various symbols and lines have the same meanings as in Fig. 3.
Experiment II: Retention and influence of instruction.

Retention. Almost two years after subject VS participated in Experiment I, his retention of the stereotyped response A-4 was tested. The data from the instruction-less sessions 1 to 4 in Figs. 3 and 4, as compared with sessions 1 to 4 in Figs. 1 and 2, show some retention, not in terms of RT, but in error rates.

Instruction. For the other sessions of Experiment II, subject VS was either instructed:
(i) to try to minimize his error score, or
(ii) to give equal attention to speed and correctness of his responses, or
(iii) to try to minimize RT, regardless of making errors.

First, the influence of these different instructions on responding according to A-4 was investigated. In Figs. 3 and 4 the results regarding the instructions "minimize errors" and "minimize RT" can be clearly differentiated. Equal attention to both aspects of responding yielded results that were hardly different from those under the instruction "minimize errors".

After 24 sessions of A-4, the subject had to respond according to A-1 again. Apart from an initial five sessions with high error rates, it seemed that the relations from Figs. 1 and 2 still applied: equal error rates, but considerably higher RT's for A-1 than for A-4. However, from the last group of data for A-1 under the "minimize errors" instruction (session 35 to 40), it is apparent that, after a learning process in which progressively more attention was given to speed of responding (session 9 to 34), the subject was able to respond equally fast, with about the same error rate, to both the first-learned set A-4 and the secondly-learned set A-1.

When, however, the triangular symbols in Figs. 3 and 4 are compared, it appears that under the "minimize RT" instruction the subject was also able to respond equally fast according to A-1 as according to A-4, but this at a somewhat higher error rate.

Analysis of errors. Almost all "full errors" that were made by subject VS in Experiment II are classified in the confusion matrices of Table I and Table II. "Half-errors" (see Method) followed similar confusion patterns. Correct responses belong in the main diagonal, but have been omitted for the sake of clarity.

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table Ia. Errors from A-4 sessions under "minimize errors" instruction. The 16 full errors represented out of a total of 213 relate to 640 trials per stimulus.

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>3</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>-</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>9</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>22</td>
<td>4</td>
<td>-</td>
</tr>
</tbody>
</table>

Table Ib. Errors from A-4 sessions under "minimize RT" instruction. The 82 full errors represented out of a total of 110 relate to 480 trials per stimulus.

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>5</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>-</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>14</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>43</td>
<td>6</td>
<td>-</td>
</tr>
</tbody>
</table>

Table Ic. Errors from all A-4 sessions. The 161 full errors represented out of a total of 205 relate to 1920 trials per stimulus.

Tables Ia, Ib and Ic contain the errors made while responding to the first-learned association A-4. The full errors made in sessions under the "minimize errors" instruction are summarised in Table Ia, those made under the "minimize RT" instruction in Table Ib, and the errors made in all the A-4 sessions in Table Ic.
The negative diagonals contain the errors "A-1", "B-2", "C-3" and "D-4". This category represents 50% of the errors in Table Ia and 44% of those in Table Ib. The errors of this category may be brought about by the wrong set of associations A-1. The errors "B-1" and "D-3" together cover 38% of the errors in Table Ia and 35% of those in Table Ib. They may result from confusion of the responses to B and D, due to the visual, acoustical and/or phonetical similarity of these stimuli. Further, the frequently occurring error "D-3", especially under the "minimize RT" instruction, may be induced by the fact that of the Dutch names for the numbers 1 to 4, only "3" begins with /d/.

The full errors made in the A-1 sessions are tabulated in the confusion matrices of Table II. Table IIa contains the errors from the sessions under "minimize errors" instruction, Table IIb those from the "minimize RT" sessions and Table IIc those made in all A-1 sessions together. The negative diagonals now comprise errors that belong to the first-learned set of associations, A-4. They represent 91% of all errors in Table IIa, and 72% of those in Table IIb. The errors "B-4" and "D-2", of equal frequency within each of the tables IIa, IIb and IIc, now cover 6% of all errors in Table IIa, and 19% of those in Table IIb. Again, these errors may result from confusion of the responses to the stimuli B and D.

The confusion matrices for subject VS and subject DJ in Experiment I show the same trend as in tables I and II. Subject KE of Experiment I first learned the set A-1. The majority of errors in his first learning process was made up of the types "B-4" and "D-2". During his second learning process, of the set A-4, the errors belonging to the first set, i.e. A-1, also became predominant.

Conclusions

1) The two response variables that were chosen to monitor the learning processes, RT's and error rate, are interdependent and must be considered together. This is demonstrated best by the effect of giving different instructions to a subject, as shown in Figs. 3 and 4. The two variables together can, indeed, be used to study the development of a stereotyped response along the lines described in the introduction. It is possible that other variables, not studied in the present experiment - like smoothness of a response or mental load caused by responding - would provide a better means of studying such stereotypes. However, such variables appear considerably more difficult to quantify.

2) Errors are of different kinds and therefore bear more information than RT's.
The type of errors made can tell something about the original (in)compatibility of an association which is learned, or about associations which subjects have learned earlier, as shown in Table I and Table II.

3) Negative transfer from a recently learned association S-R₂ to learning another association S-R₃, resulting in longer RT's and/or higher error rates, may be one of the best criteria to judge the development of a stereotype for S-R₂. Especially Figs. 1 and 2 show to what extent learning the A-1 associations is hampered by the A-4 stereotyped responses already developed, even though the A-1 associations can be considered as more compatible than the A-4 associations.

summary

The paired-associate paradigm is used to study the development of stereotyped responses by measuring reaction times and error rates of verbal responses to visual stimuli. Letters are used as stimuli, S, letter and number names as responses, R. A well-learned or stereotyped set of associations S-R₂ inhibits learning of another set S-R₃.

references

Gibson, E.J. (1941) Retroactive Inhibition as a Function of Degree of Generalization between Tasks, J.Exp.Psychol. 28, 93-115.
ON THE ORIGIN OF PHYSIOLOGICAL TREMOR

W.G. Koster and R. van Schuur

The various parts of the human or animal body exhibit small movements even when the individual is completely relaxed. In case the hand is resting on a table the most distal part of the extended forefinger vibrates with an amplitude of 0.1 to 1.0 mm and with a frequency of 10 to 25 Hz depending on the individual and on the experimental situation. This phenomenon is referred to as tremor. A distinction is made between physiological tremor which is observed in normal individuals and pathological tremor (cerebellar or intention tremor and parkinsonian tremors) as observed in patients.

In the literature a great many studies on tremor have been published since the work of Schäfer et al. (1886). The mechanical properties of a limb-muscle system can be described with an underdamped second-order system. This system is activated through signals, the origin of which is the source of much discussion in the literature.

Several hypotheses have been put forward:

1. The alpha motoneuron hypothesis assumes some pacemaker in the brain to make the motoneurones discharge synchronically (Jasper and Andrews, 1938, Marshall and Calne, 1956, Gatev and Ivanov, 1971).

2. The ballistocardiographic hypothesis states that the impact of the blood ejected from the heart shakes the body, resulting in a trembling of all limbs (Brumlik, 1962).

3. The servomechanism hypothesis holds that a movement of a limb is activating the stretch reflex, resulting in activity of the antagonist muscles and thus reversing the movement, and vice versa. (Halliday and Redfearn, 1956, Lippold, Redfearn and Vučo, 1957, Lippold, 1970, 1971).

4. The mechanical filter hypothesis emanates from the idea that the limb-muscle system is activated through a broad band noise signal (Stiles and Randall, 1967, Fox and Randall, 1970, Rau and Vredenbregt, 1971).

The resulting tremor is only dependent on the mechanical properties of the filter. The contribution of each of these causes may depend upon the experimental conditions. Lindqvist (1941) provides evidence against the first hypothesis as far as the cortex is concerned.

The first as well as the second hypothesis fails to explain the results of Marsden et al. (1969) indicating that the tremor of the two hands of a subject are not related.

In recent literature emphasis is laid most on the last two hypotheses. The available data, however, are not consistent enough to choose either of these hypotheses or to quantify the contribution of each source in a particular experimental set-up.

A possible way to discriminate between the servomechanism explanation and the mechanical filter hypothesis is by loading the limb with various weights. The mechanical filter hypothesis predicts the tremor frequency to decrease with added weights without a shift in the lower frequencies of the EMG spectrum.

Such results have been in fact obtained by Fox and Randall (1970) and Rau and Vredenbregt (1971). The servomechanism hypothesis predicts the peak in the tremor spectrum and the low frequency peak in the EMG spectrum to be the same, both shifting to lower frequencies with added weights.

The experiments to be described here are an extension of the study of Rau and Vredenbregt (1971).
The pronated forearm and hand rested in a relaxed position on the table; only the index was extended. The tremor oscillations were picked up via a specially developed accelerometer with a weight of 4 g, which was attached to the phalanx distalis. In the range of 5 - 40 Hz the instrument operates as a real accelerometer (Pacela, 1968). Only the vertical component of the oscillations was recorded. As additional weights, pieces of lead were attached to the phalanx distalis. Six additional weights were used (30, 60, 90, 120, 150 and 180 g). The EMG signals were picked up with surface electrodes (8 mm diameter) from the extensor muscles (M.extensor digitorum and M. extensor indicis). Each measurement lasted about one minute. The EMG amplifiers have a band-pass characteristic which is flat between 15 and 1000 Hz (Vredenbregt, 1971). The tremor oscillations and the EMG signals were simultaneously recorded on an instrumentation tape recorder (Philips Analog 7). The EMG signals were fed into a high-pass filter (24dB/oct, 50 Hz) to eliminate zero-line shifts, a full-wave rectifier and a low-pass filter (24 dB/oct, 50 Hz), the result being the envelope of the EMG signals (Fig. 1).

The envelope of the EMG signal is used to reflect the low-frequency activity of the EMG signal. The tremor signal, the EMG signal and the EMG envelope were fed into a hardware spectrum analyser (General Radio FFT Analyzer Time/Data). Two subjects participated in the experiment.

The results obtained with the two subjects appeared to be very similar. In Fig. 2 the power spectra of the tremor oscillations, the EMG signal from the M. extensor indicis and the EMG envelope are shown for one subject with four different loads. The spectrum of the tremor oscillations (Fig. 2a) clearly changed with the additional loads. With the minimum load (4 g) the spectrum shows a slight peak at about 20 Hz. Increasing the total load up to 184 g makes the peak to shift to a lower frequency, about 5 Hz. The total energy is about ten times the energy with minimum load. The bandwidth decreases linearly with decreasing frequency. The spectra of the EMG

*We would like to thank Mr. E.van Walwijk and Mr. E.F. van Looy of the Eindhoven University of Technology for their permission to use the apparatus and for their help during the analyses.
Fig. 2. Power spectra of the tremor signals (a), the EMG envelope (b) and the total EMG-signal (c). The energy is represented in arbitrary units. The gain of each of the signals (a), (b) and (c) was kept constant.
envelope are shown in Fig. 2b. With small loads, a slight maximum is noticed at about 15 Hz. With increasing load more activity at lower frequencies appears. The total energy increases about five times. The EMG spectrum is given in Fig. 2c. No shift in peak frequency with increasing load is noticed. The bandwidth exhibits a slight decrease, whereas the energy increases in a very pronounced way; it becomes about five times as large.

The tremor signals were sampled with a frequency of 100 Hz; 8 replications of 5 seconds each were made. The EMG envelope was sampled with a frequency of 100 Hz and 16 replications of 2.5 seconds each were analysed. The EMG signals were sampled with a frequency of 500 Hz and 128 replications of 0.5 second each were made.

**discussion**

The most striking effect in the data is the decrease in tremor frequency and the relative stability of the frequency of the EMG envelope as a function of the additional weights. This effect points in the direction of the mechanical filter hypothesis. The EMG signal may be considered as white noise filtered in a frequency band of 10 - 200 Hz (Fig. 2c). In case of a white noise signal the amplitude variations will be completely random. This will result in a flat envelope spectrum. In the EMG signals measured here the envelope spectrum is slightly peaked (Fig. 2b), which is an indication that there is some periodicity in the amplitude variation of the original EMG signal.

In static condition there exists a very stable relation between the size of the EMG signals and the force exerted by the muscle (Vredenbregt and Koster, 1966b, Rau and Vredenbregt, 1970). The very tiny movements during tremor makes the condition pseudo-static.

![EMG signals](image)

**Fig. 3.** The size of the EMG signals (rms in arbitrary units) as a function of the total load in g, i.e. equivalent weight of the finger plus the weight of the accelerometer plus additional load.
In Fig. 3 the square root of the energy of the EMG signals is plotted against the load. The shape of the curve is remarkably similar to that presented in the articles mentioned. Therefore, in this tremor situation, too, the condition can mainly be considered as a static one. The curve is slightly bent towards the EMG axis. The curvature is slight as the maximum load is only 10% of the maximum load the finger is capable of supporting. The curve does not pass through zero.

A relation, identical to that in Fig. 3, is found between the square root of the energy of the EMG envelope and the load. Thus the energy of the EMG envelope is a constant percentage, about 50%, of the energy of the total EMG signal.

As a first approximation, we assume the shape of the energy spectrum of the EMG envelope to be constant. Its shape is broad in relation to that of the tremor signal. The mean EMG signal reflects the constant force with which the finger is extended. The EMG envelope reflects the variations in that force.

We may now try to make a more quantitative analysis of the mechanical filter hypothesis. The input signal may be written as:

\[
F_i = c(m)F_c + c(m)F_v(\omega)\sin(\omega t + \phi)
\]

\[
F_i = c(m)F(a + \frac{1}{2}b(\omega)\sin(\omega t + \phi))
\]

where \(F_i\) is the input force that is linearly related to the total load \(m\), \(F_c\) is the constant force and \(F_v(\omega)\) are the amplitudes of the variable forces.

The hand-finger system will be assumed to be a mass-spring system with damping, for which the following differential equation holds:

\[
m\ddot{x} + d\dot{x} + s\dot{x} = c(m)F(a + \frac{1}{2}b(\omega)\sin(\omega t + \phi))
\]

where \(m\) is the equivalent mass, \(d\) the damping factor, \(s\) the stiffness, \(\omega\) the angular frequency and \(x\) represents the displacement.

From (1) we may calculate at what frequency the maximum output will be found, viz.:

\[
\omega_{x_{\text{max}}}^2 = \frac{s}{m} - \frac{d^2}{2m^2}
\]

As a first approximation, equation (2) may be written as:

\[
\omega_{x_{\text{max}}}^2 = \frac{s}{m}
\]

As \(\omega = 2\pi f\), equation (3) may be written as:

\[
f_{x_{\text{max}}}^2 = \frac{1}{(2\pi)^2} \frac{s}{m}
\]

In case \(s\) remains constant, this means that

\[
mf^2 = \text{constant}
\]

Or, with additional masses attached to the finger, the frequency will decrease with the square root of the mass. With equation (5) the equivalent mass \(m_o\) of the finger may be estimated when

\[
m = m_o + m_a + m_w
\]

where \(m_a\) is the mass of the accelerometer and \(m_w\) the mass of the added weight. In the same way the natural frequency \(f_o\) can be estimated. In Table I the experimental data are compared with the calculated values which are based on the estimated values of \(m_o\) and \(f_o\) for each of the two subjects.
Table I. Experimental values represent the frequency of the maximum peak in the power spectra of the tremor signals. The calculated values are based on equation (5) with $m_o$ and $f_o$ estimated from the data. It appears from the good agreement between the experimental and the calculated results that $mf^2 = \text{constant}$, see eq. (5), gives a rather good description of the experimental data. The frequencies found in this experiment are of the same value as "the die-away oscillations following a sharp tap to the finger" obtained by Halliday and Redfearn (1956). These results together with the fact that the peak frequency of the EMG spectrum and that of the EMG envelope spectrum are independent of the load, suggest that tremor is nothing but the mechanical response to a random forcing signal. In cases in which the muscle is not completely relaxed, the forcing will result from asynchronous firing of the individual muscle fibres or from synchronic discharges of alpha motoneurons (Hamoen, 1962). The rather stable peak in the EMG envelope spectrum (Fig. 2b) supports the latter process. In the completely relaxed muscle the ballistocardiogram could produce the forcing (Stiles and Randall, 1967).

There is, however, some evidence against the mechanical filter explanation:
- The stiffness. In completely static condition (x is constant) the only way a muscle can increase its force is by increasing its stiffness. In the experiments described above the EMG signal increases by a factor three when the load is increased. Thus we must assume that the driving force has increased. From other studies (Wilkie, 1956; Granit, 1958; Vredenbregt and Koster, 1966a) it is known that there exists a positive correlation between force and stiffness. Therefore we must assume that stiffness increases with load. We observe in these experiments that the tremor frequency decreases inversely with the square root of the load. In terms of equation (5) this means that the stiffness remains approximately constant. These two results are in contrast with each other.
To avoid that problem Stiles and Randall (1967) instructed their subjects to keep the EMG signal constant (i.e. on the level of the maximum load). The rationale was to keep the stiffness independent of the load. In fact, the same quadratic relation between frequency and load was found.
- The bandwidth. Another aspect in which the model deviates from the data is the bandwidth. The model predicts the bandwidth to decrease linearly with m and thus to increase linearly with $f^2$. In fact the bandwidth increases linearly with f.
The envelope spectrum. A more detailed analysis of the EMG envelope spectrum shows that the shape is not independent of the load. The results for both subjects show that with small loads (4 and 34 g) a peak in the envelope spectrum is noticed which is different from that in the tremor spectrum. With increasing load a low frequency peak in the envelope spectrum becomes visible; the peak appears to coincide with the main tremor frequency. This might be an indication that the stretch reflex becomes active, the more so as the low frequency peak decreases in frequency with increasing load, but increases in amplitude. Apart from this "tremor" peak, others appear in the envelope spectrum. The data are not accurate enough to analyse these peaks in more detail. Although the frequency response renders a mechanical filter explanation plausible, some other data made us reject the simple filter hypothesis. With increasing load the stretch reflex, too, probably plays an important part.

Summary

In two subjects the acceleration caused by tremor in the vertical plane of the indexfinger have been measured as well as the electromyographic (EMG) signals of the extensor muscles under various loading conditions (up to 180 g).

The tremor frequency decreases reciprocally with the square root of the weight. The amplitude increases with load.

With small loads the spectrum of the EMG envelope is broad without clear peaks. Above 90 g its spectrum exhibits a peak at the tremor frequency. It is argued that the tremor normally is caused by random activation of the limb-muscle system, whereas with increased load the stretch reflex becomes active.

References


A NOTE ON PARKINSONIAN TREMOR

W.G.Koster and R. van Schuur

introduction

In the preceding article a method has been described of discriminating between possible causes of physiological tremor. It consists in loading the subject's finger with various weights and measuring the tremor oscillations as well as the electromyographic (EMG) signals of the relevant muscles. There is strong evidence that physiological tremor is caused by local, i.e. peripheral, sources. In the present article the same method is used in a pilot study of patients with Parkinsonism. It is generally accepted that the so-called Parkinsonian tremor is of central origin (Wachs and Boshes, 1961; Stuart, Eldred and Wild, 1966). From this and other sources it can be concluded that a tremor registration of patients seriously affected by Parkinsonism exhibits a pattern completely different from normal physiological tremor. Its amplitude is high and a clear periodicity of low frequency (4 - 8 Hz) is observed. The EMG signals also differ from the normal ones, high amplitudes being observed together with clear periodic bursts of activity.

methods

The methods and apparatus were the same as those described in the previous article (Koster and Van Schuur, this issue). The tremor is measured with an accelerometer attached to the back of the hand near the Articulatio carpometacarpea. Two patients participated in the experiment, which comprised three tests. The tests were carried out in three loading conditions, one unloaded and two with different weights on the hand. For one patient weights of 35 and 60 g were used. The other patient exhibited such a strong tremor that proper fixation of the lower arm appeared to be impossible. Thus the mass of the lower arm also participated in the movement. It was decided to use heavier loads, viz. 65 and 330 g. The electromyographic (EMG) signals of the hand flexors were picked up with surface electrodes. The EMG signals were fed through a high-pass filter to eliminate zero-line shifts, then full-wave rectified, and finally fed into a low-pass filter to eliminate high-frequency activity. The result was the EMG envelope. These signals, tremor, EMG, and EMG envelope were fed into an FFT analyser to obtain power spectra. The duration of each recording was about 60 seconds. The tremor signals were analyzed with a resolution of 0.1 Hz, unless otherwise indicated, the EMG envelope with 0.1 Hz and the EMG signals with a resolution of 1.6 Hz.

results

In Fig. 1a a hand-copied recording of patient Sch. is shown. A clear periodicity in the EMG signal is obvious. The tremor exhibits an almost sinusoidal pattern. The suppression of the zero-line shift is clearly noticed. In Fig. 1b the results of patient Van W. are represented. The duration of the bursts is somewhat longer and the tremor shows a periodic, but less sinusoidal, character. Figs. 2a, b and c show the power spectra of the tremor signals, the EMG envelopes, and the EMG signals of patient Sch. under three loading conditions. For each of

We would like to express our gratitude to Professor Dr. J. Droogleever Fortuyn and Dr. J. van Manen (Department of Neurology, University of Groningen) for the opportunity to do this study and for their help in the performance of the experiments.
Fig. 1a and 1b. A hand-copied recording of the results from two patients. The vertical scale is arbitrary. The hand was not loaded.

Fig. 3. Tremor data of patient Sch. with a load of 55 g. The analysis has been made in much finer detail than in Fig. 2a. The resolution is 0.02 Hz. The energy is in arbitrary units.
Fig. 2. Power spectra of the tremor signals (a), the EMG envelopes (b), and the total EMG signals (c) of patient Sch. under various loading conditions. The energy is represented in arbitrary units. The gain of each of the signals a, b and c was kept constant.
the signals the gain was kept constant. The most conspicuous phenomenon in the curves of Fig. 2 is the nearly discrete character of the spectra. The power spectra of the tremor signals and of the EMG envelopes show clear "spikes" with no activity in between. The frequencies of tremor and EMG envelope coincide. The sinusoidal character of the tremor of patient Sch. is shown even better in Fig. 3. A recording of about 60 seconds was analyzed with a resolution of 0.02 Hz. The resolution appeared to be not high enough for this kind of tremor. The restricted length of the recording made a higher resolution impossible. From spectra like that in Fig. 3 the peak frequency of the tremor can be measured with high accuracy. Such measurements show that the frequency decreases somewhat with increasing load (4.85 Hz unloaded; 4.65 Hz with 65 g and 4.50 Hz with 330 g). The EMG envelope closely satisfies the same description. The "high"-frequency activity of the EMG signal (20-150 Hz) decreases somewhat with loading.

The data of patient Van W. are virtually the same. The tremor frequency was 6.5 Hz and it did not change when loaded. A strong decrease in high-frequency activity in the EMG signal was remarkable.

discussion

The data presented here must be considered as indications rather than as final results. As the tremor movements were very strong, proper fixation could not be achieved. Thus, the recorded signals cannot be ascribed solely to hand tremor. This is in particular true for patient Sch. The low-frequency cut-off of the EMG amplifier is 6 dB/oct from 5-Hz downwards, causing the 5Hz bursts to be reduced in size. The strong periodicity in the EMG signals and, hence, in the EMG envelope points to an origin of Parkinsonian tremor that is different from that of physiological tremor. From the four possible sources mentioned by Koster and Van Schuur (this issue) only the alphamotoneuron hypothesis and the stretch reflex hypothesis remain plausible. The slight shift in tremor frequency when loaded makes a pure alphamotoneuron hypothesis unlikely, as the "tremorogenic" centres will not be influenced by changes in conditions at the periphery. More emphasis can be laid on the stretch reflex that might be activated through some "central nervous system pacemaker" in the motor tract. In that hypothesis the movement is seen as the cause and the muscle activity as the resulting action. More systematic studies must be made to test the stretch reflex hypothesis.

summary

From the right hand of two patients with Parkinsonism the acceleration caused by tremor in the vertical plane and the electromyographic (EMG) signals of the relevant muscles have been measured under various loading conditions. The power spectra show the tremor signals to exhibit an almost sinusoidal pattern. The frequencies of the tremor signal and of the EMG envelope coincide. For one subject the tremor frequency as well as that of the EMG envelope decrease inversely with loading. The data provide evidence that the tremor originates from a stretch reflex activated through some central nervous pacemaker in the motor tract.

references


ON CONTROLLING THE HEART RATE*
R.van Schuur and Th.Burema**

**introduction**

Physical stress of the human body causes a load of the heart. The energy needed for this stress is supplied by the blood. More muscle activity is attended with more need of blood per time unit and the heart realizes this by adjusting the heart rate. In training for a better physical condition the limiting of the heart rate may be of great importance. Too strong a physical stress may lead to fatal increase in the heart rate in people who have had some heart disease.

Another reason for controlling the heart rate may lie in the training schedules which are often made up so as to load the heart increasingly. Therefore it is desirable to control the stress. With this end in view we have investigated on theoretical grounds three fundamental control systems. The system which turned out to be the most promising was applied in an experiment.

For this purpose we consider the human organism as a controlled system, with as the input the amount of power the subject has to supply and as the output the heart rate. For this kind of system Wigertz (1970) found the following transfer function:

\[
H(j\omega) = \frac{F(\text{output})}{F(\text{input})} = \frac{H_1}{1+j\omega\tau_1} + \frac{H_2}{1+j\omega\tau_2}
\]

According to this function it appears that the system consists of two parallel first-order systems. For this system holds \(\tau_2 - \tau_1\). The first-order system with \(\tau_2\) will play its part only for signals which, in comparison with the other first-order system, vary slowly. As it is essential that during training a specified value of the heart rate should not be exceeded rapidly, the rapid part of the system plays the major role. Therefore the transfer function is simplified to:

\[
H(j\omega) = H_0 \frac{1}{1+j\omega\tau_H}
\]

with \(\tau_H = \tau_1\).

The limiting process is such that the acting power is controlled as a function of the heart rate. An ergometer-bicycle is used to supply the power. The subject has to cycle in the rhythm set by the controller. A higher tempo corresponds with a heavier load.

The task may also be controlled in an electric way, viz. by varying the resistance the subject has to overcome during cycling at the same tempo.

The control circuit is represented diagrammatically in Fig. 1.

*The research was performed on request of the Department of Medical Systems and Devices of Philips.

**Student of the Eindhoven University of Technology.
The heartrate (u) of the subject is compared with the ultimate heartrate (f). The difference specifies the heaviness of the load (i) imposed by the controller R. By changing the task the heartrate (u) of the system H will also change, until a steady state is reached. The choice of controller is the subject of this article.

**choice of controller**

The influence of each of the elementary controllers on the system will be checked successively. The relation between heartrate (u) and the established heartrate (f) will be considered on the basis of Fig. 1. The notation used here will be applied during the rest of this article.

In general

\[ u = i H \] \hspace{1cm} (2)

\[ i = (f-u) R \] \hspace{1cm} (3)

Eliminating the heaviness of the task (i) leads to

\[ u = \frac{HR}{1+HR} f \] \hspace{1cm} (4)

If both H and R are assumed to be positive, the established heartrate (f) will be approached better according as HR is greater than 1.

**proportional controller**

By proportional controlling is understood that controller in which the heaviness of the task performed is linearly proportional to the difference between output value and established value.

For this control applies:

\[ R = R_p \] \hspace{1cm} (5)

This gives the relation between u and f as

\[ u = \frac{HR_p}{1+HR_p} f \] \hspace{1cm} (6)

The product \( HR_p \) defines how near the established level is approached. From a pilot study we found \( H \) to be about 0.7 (beats/min)/W. The value \( R_p \) is also limited because one is not able to perform an infinitely heavy task. A non-trained subject can hardly yield a load of 500 W for some time. Starting the controlling, the difference between the heartrate and the established value will be about 50 beats/min. With the maximum load of 500 W the value of \( R_p \) will be: \( R_p = \frac{500}{50} = 10 \) W/(beats/min). The product \( HR_p \) is about 0.7x10 = 7, which value does not suit the condition \( HR>1 \), needed to make \( u = f \). For this reason the established value will not be reached when using the proportional controller. Besides the level reached depends on \( H_o \) with a specified \( R_p \) value, and since the \( H_o \) value depends on the subject this type of control is not a practical one for the purpose in view.

**differential controller**

With the differential controller the heaviness of the task depends on the changes in the difference between u and f. In case of a big change the task will be heavier than with a small one. When there are no changes no task should be performed. As the established value of f does not change during the session, the control only reacts on changes in the value of the heartrate u. Whenever a steady state is reached this equilibrium will be maintained in spite of the value of u. This value then has no relation to the established value of f. Therefore this controller cannot be used either for our purposes.
integrating control

In the integrating controlling process the heaviness of the task depends on the value reached before and on the difference between $u$ and $f$. If the difference between the momentary value $(u)$ and the established heartrate value $(f)$ is positive, then the heaviness of the task will decrease owing to the controller, and increase if the difference is negative. If there is no difference the task will be maintained on the level reached at the preceding moment.

During controlling, the heaviness of the task will increase until the heartrate reaches the established value, after which the task will maintain that level.

For this type of control one has:

$$R = R_i / j\omega \tau_i$$  

With (4) this relation gives

$$u = \frac{H R_i / j\omega \tau_i}{1 + H R_i / j\omega \tau_i}$$  

Reaching the required equilibrium means that changes round this situation take place so slowly that they are not recognized as disturbances. In the ideal case the disturbance has an infinitely long period ($T$). This corresponds with $\omega = 2\pi / T = 0$ rad/s and makes $R$ of (7) infinite. In the practical situation $R >> 1$. Since the value of $H$ is about 1, the product $HR$ will be much greater than 1, so that with (4)

$$u = \frac{H R_i / j\omega \tau_i}{1 + H R_i / j\omega \tau_i}$$  

The established value will be approached very closely, so that this type of control suits the demands.

The parameters $\tau_i$ and $R_i$ will be so chosen that the established value of $f$ is not yet exceeded.

optimal control

Optimal control is obtained when:

(a) the established value is not exceeded;
(b) the established value is reached as rapidly as possible.

When both requirements are met the system is on the point of the oscillation. Oscillation may appear when the system is damped too little; then the established value is exceeded. On the other hand, if the system is damped too strongly, the established value is reached after a longer period than in the system with optimal damping. In the integrating system damping depends on the time constant $\tau_i$.

From (1) and (8) follows:

$$u = \frac{H_0 R_i}{-\omega^2 \tau_i^2 + j\omega \tau_i + H_0 R_i}$$

and

$$((j\omega)^2 \tau_i H_i - j\omega \tau_i + H_0 R_i)u = H_0 R_i f.$$  

This is a special kind of Laplace transformation, for the harmonic functions, of a second-order differential equation. The general differential equation reads:

$$\tau_i H_i u + \tau_i H_0 R_i u = H_0 R_i f$$  

The solution of (10) that satisfies the condition mentioned above is:

$$\tau_i^2 - 4\tau_i H_i H_0 R_i = 0$$

hence:

$$\tau_{icrit} = 4H_0 R_i / H$$  

This is called the critical adjustment.
When $\tau_i < \tau_{i, \text{crit}}$ oscillation will occur and when $\tau_i > \tau_{i, \text{crit}}$ the system acts more slowly than in optimal condition. To test whether this functional description holds in the human organism, an experiment was set up.

**EXPERIMENT.**

The purpose of the experiment was to consider whether the established heartrate is reached or not, and whether the integrating system controls the adjustment of the task in an optimal way. With differently chosen integrating constants it was checked whether and how the desired value is obtained. The subject used an ergometercycle to perform the task. By means of a pair of headphones clicks were presented to him. He had to adjust his rhythm of cycling to the clicks. Because the heartrate was varied within relatively narrow limits it is supposed that the power supplied is proportional to the cycling rhythm. During the experiment the heartrate was converted into a proportional voltage, which was subtracted from a reference voltage in a differential amplifier. The reference voltage corresponded with the established heartrate. After integration the difference voltage controlled the frequency generator, which gave the impulses at a repetition frequency proportional to the controlling voltage. To choose the value of $\tau_i$, the results of Wigertz (1970) were used so that in the critically damped situation $\tau_i = \tau_{i, \text{crit}} = 22$ s with $H_2 R_1$ about .25. For $H_0$ we took the value of .7 (beats/min)/W according to the pilot study. The heartrate ultimately to be obtained was established on 120 beats/min. Each of the sessions lasted 5 minutes. Fig. 2 shows the curve of the heartrate during these 5 minutes.

*Fig. 2* Heartrate as a function of time during controlled task. The time constant of the integrating control is 22s and the closed-loop gain .25.
Fig. 3 Heart rate as a function of time during controlled task. The time constant of the integrating control is 4.7 s and the closed-loop gain .25.

Fig. 4 Heart rate as a function of time during controlled task. The time constant of the integrating control is 47 s and the closed-loop gain .25.
The established value was reached within 40 seconds without overshoot. The fluctuations round this value were less than 10%. They were probably due to lack of accuracy on the part of the subject.

In a following condition, with a shorter integrating time constant, the curve of the heartrate as a function of time was explored. The value of $\tau_i$ was now 4.7 s. The result is presented in Fig. 3. The established value of 120 beats/min was reached very rapidly viz. within 20 s. This value, however, was exceeded by an overshoot of 10%. Besides periodical fluctuations arose round the established value. The fluctuations had an oscillatory character, with a significant difference of those found with $\tau_i = 22$ s. As the established value was exceeded regulary, the integrating time constant of 4.7 s does not suit the demands.

In a third condition the integrating constant was 47 s. Fig. 4 shows that the established value was reached later then in the first condition (≈200 s), while during the whole session there was no overshoot.

**conclusion**

On the basis of the preceding theoretical considerations and the results of the experiment the integrating controller seems to give good results when the heartrate is controlled by means of a task. If the closed-loop gain is .25 and the heartrate is to reach the established value critically the time constant of the controller should be equal to $T_H$, i.e. the time constant of the human system. The average value of $T_H$ is 22 s according to the experiment of Wigertz. In our experiment we found a good agreement with this value.

**summary**

In schedules for revalidation or condition training for sportsmen limiting the heartrate plays an important role. It is desirable that the various established values of the heartrate are reached rapidly without overshoot. On the basis of considerations of control it apppears that the integrating controller gives the best results. This statement was confirmed in an experiment. The time constant of the controller was found to be 22 s, with a closed-loop gain 0.25. This is in good agreement with the value expected by the results of Wigertz.

**references**


EMG APPLICATIONS IN LOW BACKPAIN

J.Vredenbregt and C.M.J. Hulst

Introduction

Among factory people working at a conventional lathe, the number of them suffering from low back pain is significantly greater than the number out of any other group of workers of various kinds of duties. The main cause seems to be the long periods of fixed bent posture of the body, when at work, the muscles of the back, especially of the low back, are constantly contracted isometrically.

In a case like this surface electromyography (EMG) of the back muscles might give information on the load of the muscles involved for various positions of the back during work, especially with regards to the posture when working at the lathe.

On this idea an EMG study was set up to find out whether the EMG can indicate differences satisfactorily when loading the back muscles during standing in a bent position. Two criteria will be treated, namely the EMG-load relation and the EMG-endurance relation, the former of which is less complicated to establish than the latter. However, the latter has the advantage of studying the muscle behaviour during loading over a long period of time.

Method

Before starting the measurements, it was necessary to check whether the proposed method offered reproducible data. This resulted in the following arrangements.

A rod with adjustable length was coupled to a bar in the form of a T. The bar was to be held horizontally by the subject standing on a platform, his arms stretched downwards. The end of the rod was fastened to the platform between the subject's feet. The rod was adjusted so as to force the subject to bend slightly forward. The subject was instructed to exert an upward force on the bar by stretching his body. This action causes the back muscles to be loaded. To enable this load to be measured, the bar was provided with strain gauges.

The EMG signals were picked up from the M. trapezius left and right, M. latissimus dorsi left and right, and M. erector spinae left and right (Fig. 1) with the aid of miniaturized EMG amplifiers provided with suction cup skin electrodes (Vredenbregt, 1971). The signals were recorded on an instrumentation recorder (Philips Analog 7). After full-wave rectifying, the mean value of the rectified signal was taken, representing a measure of muscle activity. Simultaneously with the EMG...
Fig. 2a and b. EMG-force relation of the muscles of one of the healthy subjects. The curves are single trials. Fig. 2a refers to the first trial and Fig. 2b to the second one.

Fig. 2c and d. EMG-force relation of the muscles of one of the subjects suffering from low back pain. The curves are single trials. Fig. 2c refers to the first trial and Fig. 2d to the second one.
nals the force exerted was measured and recorded. The latter was fed back visually to the subject by an indicator in order to obtain a constant force (load). Using this method, the contractions were isometric and the posture well defined. The following conditions were applied:
- the body stretched, unloaded
- the body bent forward, unloaded
- the body bent and variously loaded.

Six subjects participated in the experiment, three of whom were healthy and three were suffering from low back pain. Under these conditions two trials were performed. Between the trials a time interval of at least a few hours was inserted during which the electrodes were removed.

**emg-load relation**

Fig. 2 presents the EMG-load relation of the muscles mentioned, from zero to maximum effort. Fig. 2a and b show the result of the first and second trial of a healthy subject, while Fig. 2c and d represent the results of a subject with low back pain. Fig. 2c refers to the first trial and 2d to the second. From the results of all subjects, of which Fig. 2 is an example, the following can be concluded:
- The EMG activity of each subject obtained in the first and second trial, is largely the same.
- The EMG-load relation of the muscles involved differs from one subject to another.
- For all subjects the EMG-load relation of the M. erector spinae can, within the reliability of the data, be represented by a straight line. This is not the case for the M. latissimus dorsi and the M. trapezius. It is our experience that for relatively long muscles with long parallel fibres the EMG-load relation seems to be more linear than for relatively short ones.
- For some subjects the shape of the EMG-load relation of the same muscle left and right differs reproducibly in both trials. This indicates that identical muscles, located symmetrically with respect to the spinal column, should not always be active in the same measure. In this respect the healthy and the "low back pain" subjects did not differ systematically.

The linear EMG-load relation of the M. erector spinae, obtained under the condition of "body stretched" enabled a fictitious load at the bar to be extrapolated which would cause the same muscle activity as the weight of the trunk in the unloaded bent position. The results of the measurements of the individual EMG-load relations are such that no systematic difference between healthy and "low back pain" subjects could be established.

**emg-endurance relation**

In Fig. 3a, b and c the EMG-endurance relation of the back muscles of one of the healthy subjects is shown. Fig. 3a represents the relation of the M. trapezius, Fig. 3b that of the M. latissimus dorsi and Fig. 3c the relation of the M. erector spinae. The data have been obtained by instructing the subject to exert an upwards force at the bar equal to a load of about 75, 55 and 30% of the maximum force that he was able to exert and to keep the force constant as long as possible. To this end the value of the output was visualized by an indicator.
Fig. 3. EMG-endurance relation obtained from the back muscles during constant loading. The curves are single trials and refer to the M. trapezius (Fig. 3a and d), the M. latissimus dorsi (Fig. 3b and e) and the M. erector spinae (Fig. 3c and f). The Figs. 2a, b and c (left side) represent the results of a healthy subject while Figs. 2d, e and f represent the results of a subject suffering from low back pain.
During each trial the EMG activity of the M. trapezius, M. latissimus dorsi and M. erector spinae, as well as the actual force were recorded simultaneously. The processing of the signals was the same as mentioned under "EMG-load relation". In Fig. 3d, e and f the EMG-endurance relation of the back muscles of one of the subjects with low back pain is presented for the same muscles and in the same sequence as mentioned above.

Comparing the results, it is obvious that the curves in Fig. 3a and d are in good agreement, apart from the amplitude of the EMG. They are identical to those found by various authors for a number of other muscle groups (Lippold, 1952; Morrioka, 1964; de Vries, 1968; Vredenbregt et al., 1973).

The same agreement can be established between Fig. 3b and 3e. However, between Fig. 3c and 3f there is a small discrepancy. While the healthy subject shows an EMG-relation for the erector spinae muscle, which seems to be less dependent on a loss of contractibility during the time period of observation, the subject with low back pain shows a slight increase in the EMG activity, very well noticeable for low values of the load. In this respect the two groups differs significantly. It ought to be mentioned that in the "low back pain" group only two subjects participated in the EMG-endurance experiments.

discussion

As can be seen in Fig. 1, the force exerted on the bar is transmitted to the spinal column via the arms and shoulder girdle. The part of the M. trapezius that can transmit effectively the force from the shoulder downwards to the spinal column is relatively small, through which this part might show some indication of fatigue very soon. It is this muscle group which shows for all subjects the greatest effect of increasing EMG activity. Hence, it might be that for all subjects the total endurance time is determined by the effective part of the trapezius muscle.

In the body posture applied, the erector spinae muscles must transmit nearly the total force from the trunk to the pelvis and further to the legs. However, the EMG activity of these muscles showed hardly any increasing activity for the group of healthy subjects and only a small increase for the group of subjects with low back pain. One of the functions of the erector spinae muscle is to keep the back in the proper position rather than move the trunk with respect to the pelvis. So the M. erector spinae has for an important part the functions of a postural muscle. It is known that in postural muscles the number of slow fibres (C.fibres) is relatively high compared with fast operating muscles (e.g. M. biceps) (Smalbruch, 1970). These slow fibres are less sensitive to loss of contractibility. This makes it possible for the muscle to exert a constant force over a longer period than a comparable fast muscle. Whether the relative number of slow fibres in the M. trapezius will be the same as that in the M. erector spinae or not, is difficult to say.

Another possibility might be that, because in the M. trapezius the available fibers in the direction of the force are much less numerous in comparison with those of the M. erector spinae in which all fibers can contribute, the erector spinae muscles are supposed to be underloaded. Then it is to be expected that, in the posture assumed, the loss of contractibility during the available observation time (endurance-time) will be very small.

In the case of the subjects with low back pain, it is to be expected that a number of muscle fibres of the M. erector spinae are overstretched, owing to many years of isometric contracting for keeping the trunk bent. This would cause a greater part of the muscles to be active for compensating the load of the trunk and hence a faster loss of contractibility.
The actual force contribution of the M. latissimus dorsi is to be expected to be small because the direction of the muscle fibres is nearly perpendicular to the spinal column.

It would be desirable to verify the results with those to be obtained with a greater number of subjects.

At the same time experiments could then be carried out regarding a position of the body in which only the M. erector spinae would be loaded.

summary

Electromyographic techniques were used to observe the activity of the back muscles M. trapezius, M. latissimus dorsi and M. erector spinae, in subjects when standing with bodies stretched, bent forward, and under various loads.

Two small groups of subjects participated in the experiment. One group consisted of healthy people, the other of people suffering from low back pain.

Two criteria were studied, viz. the EMG-load relation and the EMG-endurance relation. From the data a linear relation between EMG activity and load was found for the M. erector spinae and a non-linear relation for the M. latissimus dorsi and M. trapezius. The results of the measurements of the individual EMG-load relations are such that no systematic differences between healthy and "low back pain" subjects could be established. In the case of the EMG-endurance relation a slight difference in EMG activity of the M. erector spinae could be observed.

references


ELECTROMYOPHONE

In the Annual Progress Report nr. 6, 1971 a miniaturized electromyographic (EMG) amplifier was reported which has been very valuable in our research work. Combined with a normal portable radio/cassette recorder (Philips Radiorecorder RR512), the amplifier proves to be a handy and simple electromyophone (Fig. 1).

A small modification in the recorder allows the amplifier to be connected to it, the batteries of which are also used to feed the amplifier. The combination enables the electromyographic activity of the muscles to be made audible and a subject can observe his own muscle activity. This offers the following possibilities:

- Fully relaxing of the muscle by minimizing the EMG noise, which, to our experience, is very difficult without any feedback. This is for example helpful in diminishing headaches.

- Training of the muscle function when for some reason a muscle has been more or less out of use. This will be achieved by repeatedly increasing the EMG noise of the muscle.

J. Vredenbregt.
5 INSTRUMENTATION
INTRODUCTION I.P.O. INSTRUMENTATION 1972

D.J.H. Admiraal

The arrival in 1970 of the computer type P 9202 has shown clearly that such a device does not remove work; on the contrary it attracts work. The motivation to buy a computer was to be able to control experiments, its use for simulation purposes and finally to carry out computations.

For the first application a number of interfaces is necessary as owing to the various experiments which mostly make very specific demands of electronic devices. The specialization often goes so far that commercially available instruments do not meet the requirements or they are prohibitively expensive. Besides a computer can be so arranged that all the various experiments requiring on-line use of a computer can be carried out in their respective rooms instead of in the computer room. For this purpose digital data transport from the computer to every laboratory room in the Institute and vice versa is necessary. At the computer end the interfaces for data transmitter and receiver were realized already in 1971, as mentioned by Muller in the IPO Annual Progress Report No. 6 - 1971.

The transmitter and the receiver in the experimenter's room is called MARIE, a Dutch abbreviation of "Modulaire Aanpassing tussen Rekenmachine Interfaces en een Experiment", in English, Modular Matching Unit between Computer Interfaces and an Experiment.

The design of MARIE is correlated to the requirements of the experiment itself, e.g. in an auditive experiment where tonebursts with a certain carrier frequency are presented, their number, their frequency and intensity can be controlled automatically and responses be recorded. The construction of MARIE is in its final phase and the device will be described in the next Progress Report.

The well-known principle of Pulse Width Modulation has been used in our laboratory for linear modulators in the form of signal gates for auditive, phonetic, and visual purposes. A survey of various gate systems and circuits has been described by Admiraal in the 1971 issue of the Progress Report. Now the same principle has been applied in driver devices for what are known as Sylvania Glow Modulator Tubes. Analogous circuits have been used for the purpose as described by Alewijnse in the Progress Report No. 4 - 1969. Because of the non-linear relation between light output and tube DC current, a diode correction network in the feedback path was introduced for compensating purposes. The current of the tube is switched on and off by the PWM method according to a fixed frequency of, e.g., 1000 Hz or more. The non-linear characteristic of the tube as mentioned above is of no account with this 0-1 modulation method. This results in a simple circuit without critical adjustments. The properties of the device are adequate to most of the visual experiments done in the Institute.

If necessary a light feedback can be introduced, suppressing almost completely the influence of the temperature of the tube on the light output. A description of this apparatus is given elsewhere in this issue.

The employment of the electromyograph, described by Vredenbregt and Fubini in the Progress Report No. 4 - 1969 and in a more evaluated version by Vredenbregt and Basten in Report No. 6 - 1971, has assumed such large proportions in the study of complex motoric activities that a processing device for 6 channels has been developed. All provisions required for proper operation have been made such as a vacuum pump, vacuum tubes for the suction of the electrodes, power supply etc.
There are two measuring methods:

a) lowpass filtering of the double-phase rectified signal,

b) integration of the same double-phase rectified signal.

The device is discussed in this issue.

An original design for specific requirements in pattern recognition studies, developed in conjunction with members of the IPO visual group, is the Photomatrix Computer Interface. It is discussed in this issue and it aims at on-line registration of optical configurations with the help of a 16 x 16 photomatrix.
Introduction.

In visual experiments where a modulatable light source is needed, often the Sylvania Glow Modulator tube R1131C is employed. Its use, however, poses some technical problems in designing driving circuits because of the non-linearity of the light output versus the beam current of the tube. This non-linearity results from the following effects:

- Non-linearity of luminance versus current.
- Shift of spectral distribution causing additional and varying non-linearity of luminance versus current, depending on the bandpass wavelength of monochromatic filters, used to prevent colour shift of the stimuli presented to the eye. (Fig. 1a).
- Temperature drift of luminance, which is especially disturbing when switching from low currents to high currents, or the reverse (Fig. 1b).

Fig. 1a Relative luminance versus current measured via different monochromatic filters.

b Temperature drift of luminance. Luminance may vary by about a factor 1.5 during the first minute after switching to another steady-state condition.
Experiments and accurate description of the stimuli are disturbed by these effects; therefore driving circuits have to be so designed that these effects are minimized. A few years ago a purely analogue voltage-to-current convertor to drive Glow Modulator tubes was described (Alewijnse, 1969). Non-linearities were compensated either by means of non-linear feedback, or at choice, by means of light-feedback. Last year some experiments were carried out applying pulse-width modulation (PWM) of the tube current, initiated by some remarks in the literature (Troelstra, 1971). The principal advantage of PWM is that it diminishes greatly the difficulties associated with the non-linear characteristics and also the colour shift. The remaining temperature effects can be fully compensated by means of a simple photodiode light-feedback circuit.

![Fig. 2 The principle of Pulse-width Modulation, applied to the Glow Modulator tube.](image)

With PWM only two points of the characteristic of the tube are used (Fig. 2) : the zero-current point and a point in the high current region (30-50mA). The tube is switched on and off between these points at a frequency of about 1kHz and the duty cycle of the switching process determines the mean light level seen by the eye. So low-frequency modulation of the pulse-width is interpreted subjectively as a modulation of the mean light level, while at analogue modulation a purely objective modulation of the luminance is presented to the eye. Nevertheless, it may be expected that the perceptual effects of PWM or analogues modulation are equivalent. Because of the restricted bandwidth of the Glow Modulator tube, the pulse-carrier frequency can not be chosen much higher than 1kHz. Therefore, the modulation bandwidth is restricted to about one-tenth of 1kHz. But in visual research this bandwidth is sufficiently large.

**operating principles**

Fig. 3 shows the basic set-up of the PWM circuit. It consists of a free-running sawtooth generator determining the PWM carrier frequency, an input amplifier, a comparator, a tube-driver transistor acting as an on-off switch, and a low-pass filter for inspection of the modulation process. A feedback circuitry is incorporated to provide a stable and linear PWM converting process.
The sawtooth generator is a simple operational amplifier integrator circuit, provided with automatic reset. The sawtooth frequency has been chosen about 1kHz. The maximum current through the output transistor is limited to about 35mA. A pulse-width range from 0.008 to 0.8ms corresponds to a mean current range from about 0.3 to 30mA. At pulse-widths shorter than about 0.008ms the amplitude of the light pulses decreases because of the restricted bandwidth of the Glow Modulator tube. At pulse-widths shorter than about 0.1ms the colour of the light tends to red.

The comparator converts the sawtooth voltage and the output voltage swing of the input amplifier into a PWM signal, driving the output transistor on and off. The PWM current through the output transistor produces a PWM voltage across the emitter resistor. The dc-component of this voltage is roughly filtered out by means of network RC and fed back to the input amplifier via resistor R. Together with the stronger integratory action of capacitor C across the input amplifier an overall low-pass characteristic is obtained, identical to that of the active low-pass filter shown in Fig. 4.

The frequency response of the PWM system, however, not only depends on the RC times chosen but also on the sensitivity of the PWM convertor. This sensitivity is inversely proportional to the amplitude of the sawtooth voltage. In the circuit of Fig. 4 this sensitivity amounts to a certain amplification or attenuation, indicated by a factor 'S'. It is possible to make S=1 at a certain sawtooth amplitude. In this
case there is a one to one relation between the swing of the dc-component of the PWM signal at the feedback point of the output stage and the input swing at the comparator. But at other sawtooth amplitudes $S$ will be different from 1, thus influencing the frequency response curves in the same way as a change of $n=C_1/C_2$ would do.

Two useful low-pass characteristics were realized, shown in Fig. 4, one causing no overshoot in case of pulse modulation (80 Hz, -3 dB) and one maximum flat response curve (160 Hz, -3 dB), causing about 5% overshoot in case of pulse modulation. This flat characteristic is primarily intended for use with sine-wave modulation. The two characteristics are obtained only by switching another value of $C_1$ into the circuit.

The whole circuit, together with a simple unregulated 200 volt supply is built into a small plug-in unit, intended for use in our modular frame systems. The 12 volt regulated supply is already available in this frame.

**light feedback**

The light-feedback has to compensate the temperature effects on the light output only. Fig. 5 shows the set-up of the light-feedback system, placed in a small housing together with the Glow Modulator tube and some optical arrangements. A small part of light is picked up by a silicon photo-diode provided with a blue filter to prevent an overestimation of the red part of the light spectrum.

![Light-feedback system](image)

The PWM photocurrent generated by the photodiode is amplified and integrated to a rough value of the mean light intensity by filter $RC_3$. This signal is fed back into the input circuitry of the PWM circuit. The internal current feedback factor of the PWM circuit is lowered by about a factor 7 to provide the same input sensitivity as without light-feedback. The RC time of the filter $RC_3$ is also adapted to such a value as to obtain the same overall low-pass characteristics. The light-feedback fully compensate the temperature effects, shown in Fig. 1b. The harmonic modulation distortion, with or without light-feedback is below 1%.

**conclusion**

The PWM system can indeed compete with the analogous set-up as regards simplicity and technical features. Visual experiments will be carried out to find possible perceptual differences between the two systems. Especially the aspect of the colour shift within the entire modulation range will be of interest in this experiments. If two or more PWM systems have to be used together in one experimental set-up, it is advisable to couple the carrier frequencies in the well-known master and slave operation to prevent visible low-frequency beats. The application of PWM will be also useful for driving Xenon compact sources.
### technical features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWM carrier frequency</td>
<td>ca. 1 kHz</td>
</tr>
<tr>
<td>Modulation range</td>
<td>0.3 - 30mA (0 - 35 mA max.)</td>
</tr>
<tr>
<td>Modulation bandwidth</td>
<td>0 - 80 Hz, -3 dB 0% overshoot</td>
</tr>
<tr>
<td></td>
<td>0 - 160 Hz, -3 dB 5% overshoot</td>
</tr>
<tr>
<td>Modulation input</td>
<td>1 V peak to peak for 100% mod. depth at 15 mA,</td>
</tr>
<tr>
<td></td>
<td>across 600 Ohm or 50 kOhm. Phase or anti-phase.</td>
</tr>
<tr>
<td>Harmonic distortion</td>
<td>&lt; 0.4% at 95% mod. depth</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.1% at 50% mod. depth</td>
</tr>
<tr>
<td>Control filter</td>
<td>200 Hz, -3dB, 18 dB/oct. Bessel characteristic.</td>
</tr>
<tr>
<td>Light-feedback</td>
<td>via silicon photodiode with blue filter.</td>
</tr>
</tbody>
</table>

### summary

The glow modulator light source used as the stimulus in visual experiments has a non-linear light output characteristic. To reduce this nonlinearity the glow modulator tube may be driven in a way known as Pulse Width Modulation. This principle has been realized in the described instrument. It is supplemented by a light feedback system reducing temperature effects.

### references


A 6-CHANNEL ELECTRO-MYOGRAPHIC MEASURING SYSTEM

M.A. Alewijnse and J. Vredenbregt

The electro-myographic signals picked up at the skin from an active muscle are very much like low-frequency noise with a spectral maximum at about 80 Hz. To evaluate the relative amplitude of such an EMG signal a well-known representation is given by the integration of a rectified EMG signal. Another often used technique is low-pass filtering of a rectified signal. Fig. 1 shows the two methods of measuring technique and the resulting output signals.

Fig. 1. Myographic measuring system. The output signals shown correspond to step-like EMG activity bursts (qualitative).

Both measuring techniques only differ in the way of representation of the EMG activity. In case of integration the slope of the output signal represents the degree of muscle activity, while in case of low-pass filtering the representation is given by the amplitude of the output signal.
RC times, both of the integrator and the low-pass filter have to be chosen as small as possible to prevent too much loss in the representation of small variations in the EMG activity. On the other hand, too small RC times cause too much ripple on the output signals.

Especially with short step-like EMG activity bursts, integration may be more favourable than filtering, but in general both measuring techniques provide the same information. A disadvantage of the integration is that it needs more processing after registration.

At the IPO, small myographic units provided with vacuum suction cup electrodes and differential amplifiers are used for picking up the EMG activity from the skin (Vredenburg, 1971). Especially for these 'myographs' a six-channel measuring system has been constructed, in which use is made of the Philips modular frame system (see photograph). The measuring system makes possible simultaneous detection and processing of the EMG activities from different muscles in studying complex movements, for instance handwriting. The system contains the measuring circuit shown in Fig. 1 and also provides the supply voltages and the vacuum for the myographs.

The main plug-in unit incorporates all general functions:
- a membrane vacuum pump, regulated by a small variac,
- a battery pack of rechargeable NiCd batteries, plus and minus 6 volts,
- a charging unit, and
- a measuring circuit and front panel meter enabling individual measurement of the rms values of the EMG signals and battery check to be made.

Only the vacuum pump is connected to the mains supply. All other circuits including the myographs connected to the system are supplied by the battery pack to prevent danger to subjects.

The charging unit is connected only if charging has to take place, during which the system is not available for measurements.

The system also contains three small plug-in units, each provided with two identical measuring circuits. Each channel has its own electrical and vacuum connection to a myograph. Battery and vacuum supply can be switched on and off per channel to prevent vacuum leakage and loss of electric power. Fig. 2 shows a more detailed circuit diagram of a measuring channel.

![Circuit diagram of an EMG measuring channel.](image-url)
Each channel consist of:
- a variable amplifier, amplification 1-2-5-10-20-50-100 x at choice,
- a linear double-phase rectifier circuit,
- an active low-pass filter, cut-off frequency 50 Hz, 18 dB/oct. (Butterworth characteristic), and
- an 'ideal' integrator circuit, RC time 0.05 s, provided with automatic reset at an output level of 4 volts.

All output signals as well as the reset pulses of the integrators are available on the front panel connectors.

The rear panel of the frame contains the plug-in connectors for electric power and vacuum supply. A vacuum buffer tank is also mounted on the rear panel.

In the set-up described all integrator circuits reset at random times, depending on the individual differences in EMG activities. In some cases it might be better to apply the principle of the 'first arrival' in such a way that the integrator which reaches its reset level first, resets all other integrator circuits, thus providing a clearer picture on the recordings.

It is also possible to use the system on-line with the computer for direct analysis and comparison of the output signals.

The front panel meter on the main unit is primarily used for calibration and alignment of the overall sensitivity of the measuring channels.

**summary**

An apparatus is described for the derivation of electromyograms of several muscle groups simultaneously. Up to six electromyographs may be connected to the unit, which provides either integration or low-pass filtering of the signals.

In addition it supplies the vacuum needed for the suction cup electrodes of the myographs. The unit may also be used in field work by means of a rechargeable battery pack.

**references**

PHOTOMATRIX COMPUTER INTERFACE
on line registration of optical configurations for visual pattern recognition

P.W. Verhagen* and M.A. Alewijnse

introduction

Research on human pattern recognition has led to several hypotheses on the analyzing properties of the visual system. An important theory has been put forward by Hubel and Wiesel (e.g. 1965) based on physiological experiments. They found cells in the animal visual cortex that were sensitive to the orientation of a line or edge projected on a part of the retina connected to such a cell. Starting from further research, they supposed these direction-detecting cells to be a part of a hierarchically analyzing mechanism. Bouma and Andriessen (1968, 1970) of our Institute supposed from results gained with psycho-physical experiments on human slant perception that, in case of simultaneous excitation of many such cells, the cell with maximum excitation decides the perceived orientation of the stimulus. Research on human pattern recognition is, however, still in its initial phase. In this situation the need was felt for a model making possible simulation of such hypothetical analyzing functions of the visual system. A few years ago Cosijn and Hoeks (1970) built a model in hardware, consisting of 16 light sensitive cells (LDR’s) together with some electronics circuits.

*Student at the Eindhoven University of Technology.
Their apparatus essentially 'recognized' one out of 4 orientations of the stimuli projected in the receptive field (horizontal, vertical, and + or -45 degrees), independent of stimulus position in this field.

In carrying on this kind of work, a choice had to be made between hardware and software models. Pure hardware models necessitate fairly complicated electronic circuits, the sheer number of which leads to voluminous constructions. These are not flexible in the sense that modifications are difficult to realize.

On the other hand a purely software model is not attractive because of the problems encountered if complex stimulus patterns have to be programmed. So it was decided to build a hardware receptive field consisting of many light sensitive cells, connected on-line to a computer through a special interface. The hardware part of the system is purely passive, it only converts light into voltages compatible with the computer system.

Analysis of the voltage patterns resulting from a certain stimulus configuration projected on the receptive field will be performed by appropriate software programs. In this way it is possible to experiment in an easy and tangible way with many different, simple or complex, stimulus configurations; in addition, the model itself, now being a software computer program, is easy to modify.

The matrix input to a computer has, of course, a much wider application than just for the detection of a slant.

**practical design**

The P9202 computer in use at our laboratory is provided with a modular Input-Output System (MIOS), described in the previous issue (Muller, 1971).

It contains 16 analogue and digital input and output modules.

For our purpose only the 16 analogue inputs (AISM) and the 16 digital outputs (DOSM) are of interest.

Because of the number of inputs available, the receptive field is built as a 16x16 matrix, consisting of 256 silicon photo-transistors (BPX29).

Each column of the matrix is connected - via the interface amplifiers - to one of the AISM inputs. By switching on row after row of the matrix by means of the digital output information programmed, all matrix elements are successively connected to the AISM for reading in the corresponding voltages of the column amplifiers.

So the photomatrix and the interface have two general functions:

- Translation of the luminances on each receptor cell into appropriate voltage levels for the 16 analogue input channels of the MIOS.
- Translation of the 16-bit digital output information of the MIOS into operational functions, e.g. switching the rows of the matrix, setting the amplification of the column amplifiers, displaying the detected presence or absence of certain configuration properties, etc.

Fig. 1 shows the set-up of the system in general.
For practical reasons a part of the interface is incorporated close to the photomatrix. The photomatrix unit actually consists of 3 printed circuit boards, one containing the matrix on the front of the unit. The second contains 16 electronic switches (FET's) together with a decoder circuit, and the third 16 column preamplifiers. Fig. 2 shows the lay-out.

During the investigations we received information on commercially available integrated photomatrices from Reticon Inc. For our purpose however, these circuits are unnecessarily small, and, moreover rather expensive.

For the conversion of luminance into voltage we used the well-known linear photometer circuit shown in Fig. 3a. Here, a silicon photo-diode is connected to the feedback input of a simple operational amplifier. The diode current produces an output voltage across the feedback resistor, proportional to the light intensity on the diode. Because the dc-voltage across the diode is approximately zero, influence of temperature and leakage is negligible. Such a simple circuit is linear over about 3 decades of luminous intensity.

In one column of the 16x16 matrix, 16 photo-diodes can be connected to the same preamplifier, but each diode has to be provided with its own on-off switch, making for an unacceptable number of 256 switches. This can be avoided when a normal diode is placed in series with each photo-diode (Fig. 3b). The diode prevents leakage currents to and from other photo-diodes in the same row, resulting in only one switch per row.

Such a series diode is also formed by the base-emitter junction of a photo-transistor, which is in this case a normal npn-transistor mounted together with a photodiode, connected between collector and base (Fig. 3b).
The collector of the photo-transistor needs a few volts for suitable operation. The photometer circuit has, therefore, to be changed in the way shown in Fig. 3c. The additional resistors in the circuit make the collector supply possible without causing a dc-shift of the output voltage. Owing to the current amplification of the transistor the circuit is much more sensitive with respect to the photo-diode circuit, but also more sensitive to changes of temperature. This, however, causes no trouble in practice. Fig. 3d shows the set-up of a simple 2x2 matrix as an example. The actual 16x16 matrix is built up according to this principle, providing 16 preamplifiers and 16 switches. All photo-transistors have the same collector voltage of about 3 volts. Spread of transistor data (although selected) causes different sensitivity and gradient between the receptors for which a correction is made in the computer program. The FET switches are operated by appropriate gate voltages to provide a pure on-off function. These voltages are delivered by a decoder receiving 4 bits from the digital output of the MIAS. One extra bit is used to switch off all the rows as well as the collector voltage of the photo-transistors, thus enabling the offset voltages at the amplifier outputs to be measured for correction purposes.

![Diagram](attachment:image.png)

Fig. 3a Basic photometer circuit with photo-diode.
3b Photo-diode with series diode, photo-transistor.
3c Photometer circuit adapted for application of a photo transistor.
3d Simple 2x2 matrix

All signals to and from the computer are handled by the interface. This unit contains:
- A regulated power supply for all electronic circuits, including the photomatrix.
16 line amplifiers, the amplification of which can be varied in six steps, identical for all amplifiers, by manual control or automatically by means of 3 bits from the digital information from MIOS. This makes possible the use of the photomatrix at very different light intensities. Automatic adjustment of the amplification is realized by means of a subroutine in the program. The process starts from the highest amplification. If, during reading the matrix, overloading is detected on one or more outputs, the amplification is automatically adjusted to a lower value, and this continues till the overloading has disappeared.

A bank of preset thumbwheel switches providing a range of constant voltages via the amplifier outputs. These voltages are used in the program as operational constants or for interruptions or alterations in a running program. Reading of the constants is performed during the time the photomatrix is switched off, and, during reading the matrix these voltages are automatically set to zero, thus not affecting the matrix output levels.

Digital circuits receiving and translating the digital information programmed. The 16 bits available are used in the following way:

- Bit Nos. 1-4: switching on and of the rows of the matrix successively.
- Bit No. 5: switching off the matrix for reading offset voltages.
- Bit No. 6: switching off the matrix for reading of preset constants.
- Bit Nos. 7-9: setting the amplification level.
- Bit Nos. 10-16: not in use yet, reserved for driving a hardware display.

The operational bits 1-9 can be inspected by an octal LED-display, mounted on the front panel of the interface (See photograph at top).

To correct for the individual differences of the matrix cells a test program is made providing a correction matrix which is always available by means of the computer disc memory. Such a matrix is derived by computing the least-squares line through a number of measured points on the characteristic of each matrix cell, corresponding to different levels of illuminance, e.g. 10-100-1000 lux. The individual line of each cell is then translated into a standardized line by computing the correction factors for slope and zero-level. Thus, a uniform illumination of the photo-matrix results in 256 computed numbers related to the individual output voltages of the matrix cells spreading only a few per cent.

Testing and trial experiments can also be done off-line, to reduce computer time.

Acknowledgement

We thank Mr. H. Hendriksen, student at the Technical Highschool of Zwolle, who joined us for 3 months to make the Photo-matrix Interface and to test the electronic circuits.

Summary

For the simulation of recognition processes of the human visual system a receptor field has been developed, the output of which can be analyzed on line by a computer. The receptor field consists of a 16x16 matrix of phototransistors with switching and amplifying electronics incorporated. It is connected to the analogue and digital I/O system of a computer, by means of an interface.

The interface decodes the instructions of the computer program and controls the operations of the receptor field. The analysis of optical configurations presented to the receptor field may be performed by appropriate programming techniques.
references


6 IPO PUBLICATIONS 1972
6 IPO PUBLICATIONS 1972

P 217 J.A.J. Roufs
Vision Research 12, 1972, 261 - 278.

P 218 J.A.J. Roufs
Vision Research 12, 1972, 279 - 292.

P 219 J.P.M. Eggermont
Systematiek in de verwarring bij het spraakafzien.
Universiteit van Amsterdam 1972.

P 220 J.F. Schouten
Wat, Waarneming en Werkelijkheid.
Organorama 8, 1971, 13 - 17.
Also appeared in English, German, French and Spanish.

P 221 M.J. de Vriendt- De Man
S. de Vriendt
J.P.M. Eggermont
J. van Nes
M. Wambach
Méthode audio-visuelle de Néerlandais I:
Excésises de Réemploi.
Marcel Didier, Paris 1972.

P 222 I.H. Slis
Enkele aspecten van de produktie en waarneming van spraakklanken.

P 223 A. van Katwijk
Een goed verstaander heeft maar een half woord nodig: de vraag is welke helft?

P 224 D.J.H. Admiraal
Een tienkanaals Fourier-synthesizer: De pansfluit.

P 225 F.L. van Nes
Determining Temporal Differences with Analogue and Digital Time Displays.
Ergonomics 15, 1972, 73 - 79.

P 226 J. Eggermont
Some aspects of the Speech Reading Process

P 227 J. Vredenbregt
Ein Muskelreizgerät für ein- und doppelseitig Gehbehinderte.
5. Internationaler Kurs der Schweizerischen Arbeitsgemeinschaft für Prothesen und Orthesen (AFO) Zürich.

P 228 H. Duifhuis

P 229 A. van Katwijk
Faktoren in de waarneming van klemtoon.
Verslagen van de voorjaarsvergadering van de Ned. Ver. v. Fonetische Wetenschappen. p. 29 - 34.

P 230 F.L. van Nes

P 238 J. 't Hart

DOCTORAL THESES

H. Bouma
Receptive Systems Mediating Certain Light Reactions of the Pupil of the Human Eye.

H. Duifhuis
Perceptual Analysis of Sound.

R. Collier
From Pitch to Intonation.
Catholic University of Louvain, June 1972.

S.G. Nooteboom
University of Utrecht, October 1972.