Article *NIMHANS Journal*

Physiological Background of Speech Decoding Processes in the Brain

Guest lecture delivered at the First National Symposium-cum-Workshop on 'Central Auditory Disorders' held at NIMHANS from September 12-17, 1983

Volume: 02 Issue: 02 July 1984 Page: 93-112

Manfred Spreng, - *Institute of Physiology & Biocybernetic, University of Erlangen, Erlangen, West Germany*

The outstanding position of mankind to a large extent is based upon their given talent to form and comprehend conscious conceptions, to choose words as carriers of semantic meanings or thoughts and to produce and understand them as spoken language. It seems, if we believe in the Genesis, that the divine breath, making the only remarkable difference between human and animal energetic and information processing systems, partly reflects in the excellent ability of communication, using our vocal and auditory systems.

There is no such tremendous difference in the capability of the human visual system, compared with that of the animals. Neither the visual system is able to analyse such a distinct time-variable process as speech is: spoken words are quick singular events; they can never be articulated identically twice.

Therefore speech perception may be regarded as grouping of classes of signals to a meaning. Signals which carry at least two types of information; a semantic one and a phonetic one. The latter also helps to identify the meaning, really intended by the speaker. And it often lasts remarkably longer than a single word. The reception of the acoustic speech signals, their transportation to the sensory cells of the ear and the decoding processes within the neural part of the auditory system are basic requirements of speech communication.

In this presentation we will discuss examples of functions within the field of auditory information processing in relation to problems of linguistics. And in addition, we will emphasize the interactions between articulation and auditory perception. But, whatever we will find out of plausible partial explanations, we have to take the warning of Chomsky [1] and Teuber [2] very seriously: the tremendous achievements of the normal child in the acquisition of language must require an intricate, and essentially innate, apparatus, shared by all of human kind, for the extraction of linguistic feature, and for the production of grammatically ordered speech.

In other words, a motor theory of speech perception is a too weak and too simple model. But it should not be replaced by a hypotheses of comparable weakness, chaining simply complex or hypercomplex neural feature defectors within the auditory system.

In fact, it is still a completely open question in which way the connection mechanism between various decoding elements and abstract units acts, which are serving as linguistic detectors, if the latter ones do really exist. Almost the same question is discussed in the production theory: are there rules which mediate between abstract-discrete, static and context-free-linguistic units and concrete-dynamic, continuous and context-adjusted -production units or not? [3]. The obvious lack of success in answering those questions should compel all of us to concentrate modestly, basic experimental phonetics and neurophysiology, thus clearing up fine motor control behaviour and detailed processing of auditory excitations.

Time-Analysis of Speech Stream

If we regard the auditory system in detail it turns out to be a complex time-analyser, rather than a system to detect simply different frequencies. Besides, this time-analyser must not restricted to linguistic elements, but it should be able to detect even smaller segments. If the ear really acts as a system dividing the speech stream into such small segments, this is a very hard task. The time resolution for recognition of this kind of a grammatic speech is often assumed to be insufficient [4]. The efficiency of the auditory system as a processor of time-varying signals is pointed out by regarding the perception of the minimal pair (phoneme), necessary to distinguish two spoken words. For that the ear should recognize and roughly decode the information about the consonantal place of production in a patch of noise within 20 to 400 ms and a silent interval in the case of a stop consonant ranging from 30 to 80 ms. Immediately following the transitions, an analysis of the steady-state formant frequencies (differing 1 to 4 octaves within the frequency range 700 to 3000 Hz) should take place during a mean duration of about 300 ms. It is still an open question, whether the ear does some segmentation shorter than the syllable length. The essential meaning of the latter segment has been pointed out so clearly by experiments of speech synthesis [5], [6].

Segments of the size of phonemes can not be found "as segments" in the acoustic stream. Smaller ones, mentioned before, do not have clear boundaries to be regarded as "basic units" of perception, besides their obvious lack of invariance. Thus for instance, syllables [ba] and [ga]differ only in the direction of the second-formant transitions, rising for [b] and falling for [g]. But they are sounding like rising and falling glissandos or like chirps when presented along. Even more striking, two very different perceptions are produced simultaneously; the syllable [ba] (or [ga]) and a rising (or falling) chirp, when one ear receives the formant frequencies and the other just the transitions. Another example for the lack of invariance in the time-range of the transitions is the fact that in one case the rising transition at the beginning of the first formant causes the listener to hear a stop [k] in [ki] or ([p] in [pu]). In the other case a silent interval of 60 ms after a patch of noise [s] produces the perceptual equivalence [ski] (or [spu]) [7].

Both, the impossibility to find acceptable small segments and the lack of invariance, give reason to regard the acoustic inputs as continuous, and that the processing of it is not influenced by linguistic analysis into real segments [8]. recently Blumstein and Stevens [9] on the basis of short-time spectra (initial 10-20 ms of a stop consonant) argued for a segmental speech processing. The gross shape of the spectrum in this short period seems to provide the essential perceptual cues for place of articulation with remarkable invariance to the vowel contexts. The problems of short-time spectra, and the fact that auditory systems hardly act as short-time spectral analyses are reasons for further tests of that hypotheses, which on the other hand emphasises the importance of short-time analysis of the ear. However, it seems that decisions about any particular speech segment (at a meta-level)- from phonem to sentence-are made when enough information has been processed. The idea of parallel processing of a continuous acoustic input is also supported by the remarkable capabilities of the ear in temporal resolution, sequencing and temporal integration. Especially the latter is necessary to keep the continuously changing excitation, which is passed on to the upper parts of auditory system, for a period of time. Within this time period, which may be regarded as a gliding time window, various influences, forward and backward, within the parallel processing channels can be assumed.

Nooteboom [10] argued for a "primary auditory storage' or "gliding time window" of about 250 ms based upon silent gap experiments [11] forward masking in vowel-consonant series [12], speech pause perception [13], backward recognition masking [14] and measurements of rate of decay of auditory sensations. This decay-time of roughly 250 ms still seems too short to account for the temporal integration, needed for the recognition of words and morphemes. Perhaps the transformed auditory attributes of sound like timbre, pitch, loudness and duration may have longer survival times in a secondary auditory memory. However, the existence of temporal reversals in auditory information processing-recognition times are shorter for vowels than for the prevocalic consonant- and the fact that word recognition was not necessarily mediated by phonems [e.g. initial phonems in words (like bit) are reacted to faster than initial phonems in non-words (like bip)] point clearly to an overlapping, continuous, real-time selection process starting with fine auditory patterns of the beginning of a word. (Active Direct Access Model according to Marslen-Wilson, 1979).

Indeed, words are often only recognized on the basis of the acoustic beginnings. And the essential acoustical qualities of phonetic identity usually are preserved despite variation in the rate of articulation [15].

Special Functions of the System

Speical functions of the auditory information processing system with regard to speech perception can only be summarized briefly [16], [17], [18].

If we have a look at the total auditory information processing system (Fig. 1) between the sound source and the conscious perception we will notice some neural layers or neural computing systems, 2000 to 40000 element each, with different tasks. Thus for instance the quadrigeminal body calculating bilateral time differences for the directional hearing can be mentioned. The different layers are multi-connected with a complex neural network Special functions as those of so called tuned elements and coincidence systems as well as lateral inhibitions and interconnections between space and time patterns are basic requirements for speech recognition.

Simplified block scheme of the auditory information processing system neglecting descending control loops. [Spreng [18] ***1978]***

After passing through the mechanics of the middle ear and through the hydrodynamic system with the cochlea, the acoustic signal influences the transducing hair cells with a typical distribution of displacements along the basilar membrane (depending upon the frequencies) and in addition stimulates most of the pair cells with the corresponding time-pattern. As shown in Fig. 2 the broad displacement of the basilar membrane caused by a single frequency is enlarged, if the intensity increases. As a consequence the peaks of the displacements can be shifted away from a given hair cell to produce an excitation just reaching its threshold. This is the way, so called tuning curves are measured (drawn thick) giving the threshold intensity for excitation if the stimulating frequencies move up and down the critical frequency of the unit under consideration. Tuning curves exist for all elements, not only for those of the acoustic nerve, and they show quite different shapes, steepness of their slopes and behaviour. Besides this processing of spatial distribution of excitation an analysis of periodicity takes place within the auditory elements. This is explained by Fig. 3. During the acoustic signal of a constant frequency of 1600 Hz (625μ s interval), either switched on for several times or continuously running (small bars in Fib. 3). Evaluation of the intervals between the spikes (in a larger period) results in histograms with peak distances equivalent to the periodicity of the acoustic signal (lower part of Fig.

3).

Sketch of the measure of the tuning curves (threshold curve) of neural elements [Spreng [18] ***1978]***

Schematic explanation of the principle of periodicity analysis [Spreng [18] ***1978]***

Consequently, the auditory system has four basic possibilities of parallel processing the time and intensity parameters of the acoustic signal. Increasing intensity (at a constant frequency) produces
1) increasing rate of action potentials in a nerve and
2) increased number of excited elements according to the enlarged area of displacement.

Variation of the frequency causes
1) moving range of displacement or excitation (place theory) and
2) changed periodicity in the action potential series within the nerves (periodicity analysis).

It must be mentioned, that adaptation takes place already during the transducer process in the sensory cells as well as in higher centres. This means an overshooting excitation for quick changes of the acoustic stimulus accompanied with a threshold shift due to a momentary increased steepness of the dynamic intensity function. The static intensity function is recovered during readaptation, a process consuming as more time as more frequent the adapting dynamic stimulus has occurred. Another interesting function within the peripheral part of the system is the two-tone inhibition. Fig. 4b present sa tuning curve of an element whose firing rate is reduced if an additional sound is present with frequencies lying in the left/right shadowed ranges. As shown in part a) of this figure a tone with the critical frequency (360 Hz, 40 dB) of that element produces a remarkable firing rate (uppermost rows). Also louder tone bursts with another frequency (720 Hz, 60 dB) produce excitation (middle rows). If both stimuli are presented simultaneously a clear reduction or inhibition takes place as to be seen in the lower most rows of this figure.

Example of two-tone supression (a) and the inhibition ranges (shadowed) of an auditory element. [Keidel and Kallert, [17] ***1979]***

Regarding the function of higher centers of the auditory information processing system we will find increased steepness of the slopes of tuning curves by lateral inhibition (increased selectivity), the existence of on-and off-elements, longer lasting (30-70 ms) feed-forward inhibition, clocks and coincidence systems. Besides them, especially those elements are of interest, which show sensitivity to frequency modulation alone [19], [20], [21].

As a typical example the dynamic response curve (approximately comparable with the inverse tuning curve) of an element in the geniculate body of the cat is given with Fig. 5. The element do not show any excitation following a continuous tone stimulation. If the frequency is modulated from 100 Hz up to 10 kHz (scheme in the lower part), however, the element produces strong excitations at different frequencies (upper part). The same effort occurs if a downward frequency modulation stimulates that unit.

Dynamic response curve of a single unit of the geniculate body (cat) to frequency modulated stimuli. The symbol of the stimulus indicates the direction of frequency modulation, namely first from 100 Hz to 10,000 Hz and then back from 10,000 Hz to 100 Hz [Kallert, cit. in Keidel [20] ***1974]***

In a very first approximation we may assume such elements to act as parts of a decoding circuit for on-and offsets of consonants. Beyond that more specialized elements could be found in the cats geniculate body. Fig. 6 presents a unit with an asymmetric behaviour during frequency modulation in the range 2 to 5 kHz. The maximum of sensitivity during upward modulation is in the range of 2, 6 kHz and for downward modulation at 3,3 kHz. That unit does not respond to steady state stimulation at all.

Asymmetric responses pattern to an upward and downward frequency modulated stimulus [Keidel [16] ***1975]***

Even more complex units are presented with the next figure (Fig. 7). The right part gives additional examples of elements a symmetrically sensitive to frequency modulation in the range 100 Hz to 10 kHz and 2 to 5 kHz, and thus being able to decode the direction of the modulation. Surprisingly, this unit is more sensitive for upward modulation if the frequencies sweep from 100 Hz to 10 kHz (right side, upper part). It turns to be most sensitive during downward modulation if the frequencies change from 5 to 2 kHz, respectively.

Dependence of the response pattern upon sweep rate (left) and direction of modulation (right). The sweep rate can be taken from the time scale of the abscissa, and direction of frequency modulation can be seen from the stimulus symbol below the abscissa [Kallert, cit. in Keidel [20] ***1974]***

In addition, as proved with experimental results shown in the left part of Fig. 7, elements exist with an excitatory pattern clearly dependant on the speed of the frequency modulation. Decreasing the sweep time for a 100 Hz to 10 kHz change from 500 ms to 10000 ms (from top to bottom in Fig. 7) produces remarkable changes in the behaviour of that unit.

Based upon this specialized behaviour it is a reasonable assumption to regard those elements as a main part of a decoding system for transitions [21].

The complexity of the elements of that kind being responsible for this important decoding process needs some more explanations. Before a detailed discussion of Fig. 8, it should be mentioned that facilitation and inhibition normally act with considerable delays. And such activities often last for a longer time. In addition there exists feed-forward inhibition (a primary sound suppresses the expected activity of a succeeding one) as well as feed-backward inhibition (the activity of a primary stimulus is reduced or extinguished by a following one). The facilitating effects behave as well. The scheme in this figure (Fig. 8) outlines the realisation of three different types of elements by the facilitating and inhibiting influence of nerves coming from a lower neural layer and converging in the upper layer, under consideration. The elements of the lower network may possess the tuning curves drawn above, with critical frequencies in near neighborhood. It can be seen how narrowing, widening and multi-peaks results from this in the upper neural layer (drawn below). Furthermore those elements with inhibition zones at both sides (shadowed) do not respond to frequency modulations. The inhibition induced during the passage through these zones lasts for a longer time, acting as feed-forward inhibition. Therefore inhibition is active even if the frequency range is passed which should evoke excitations. Similarly an element with only one zone of inhibition will be sensitive to that frequency modulation which will not pass primarily its inhibition zone. This, as shown in the middle part of Fig. 8, the element is only sensitive to increasing (or upward) frequency modulation. This explains the microelectrode recordings presented with the preceding figure (Fig. 7).

Scheme to explain the different behaviour of the tuning curves of neural elements and their sensitivity to the direction of frequency modulation [Spreng [18]***]***

In addition those neurons are less excitable by broadband noise with considerable energy at frequencies in the inhibition ranges. Their excitation is also reduced in the case harmonics are emerging in different zones of the element.

Naturally the rate of modulation plays an important role, depending on the way of neural connections. We must assume the inhibition to be only effective for quick passages of the inhibition zones. During very slow frequency modulations the inhibitory effects (inhibitory postsynaptic potentials) may die away before excitatory influences will start. On the other side a very rapid frequency modulation will produce neither large inhibitions nor considerable excitations. This explains the result presented in the left column of the preceding figure, where a medium speed of the frequency modulation (sweep time 2000 ms to 4000 ms) reveals the largest excitations. If there exist more complex connections, including delaying inter-neurons, paradoxical effects may arise. For instance elements, which do not respond to continuous tones but show activity for quick frequency modulations ranging only outside or only within the inhibition zones.

Experiments carried out in bats [19] result a small number of elements responding to broadband noise signals only. Neither continuous tones nor frequency modulations produce any excitations. Elements of that type may be regarded as pure consonant detectors. Elements with multi-peaked frequency response curves and the results of microelectrode studies in the cat as presented with the next figure (Fig. 9) make us almost suppose the existence of pure vowel decoding elements. The element presented with Fig. 9, however, acts as a multi-peaked one (lower part) only in the presence of an additional stimulus (6 kHz tone). It is easy to speculate upon a special task of this decoding element namely to identify immediate groupings of vowels as for instance an [u] after an [o]. Elements of this kind are switched on only if an additional formant is present.

.Change of selectivity (single-peaked to multi-peaked) of a single element in the presence of an additional stimulus (lower part) [Kallert, cit. in Keidel, [20] ***1975]***

Finally neuronal elements exist in the geniculate body which are very sensitive to time differences between the stimuli. As presented in the right part of Fig. 10, element B responds if the distance of two stimuli exceeds 50 ms. The element A is excited only if the two stimuli are more than 120 ms apart [22]. Indicator elements for the duration of the silent intervals in the speech flow can be based upon those neurons. It is of interest to notice the 250 to 300 ms delay of the excitation produced by a noise burst (left part of Fig. 10). This behaviour indicates the long survival time of consonantal information and throws some light upon temporal reversals within the processing of auditory information, which we have mentioned before.

.Examples of neural elements which are sensitive to time differences (right part) [Atkin and Dunlop [22] ***1968]***

Before detailed comparisons between results of linguistic experiments and physiological findings will be made, the behaviour of decoding elements following more complex stimulation and in the awake animal should be regarded. Thus for instance Fig. 11 presents an element for a combined transition-formant configuration. In the upper left part is shown that no excitation (even suppression) takes place if only a single sinusoidal tone (1 kHz) is present. Likewise a single transition from 10 kHz to 2 kHz does not excite the element, as to be seen in the lower left part. Presenting both stimuli at the same time (right part), however, produces a large activity of the element, which perhaps may be a part of a transition-formant decoding circuit. Another example of excitations following complex stimulations is given with the next figure (Fig. 12). Part (a) presents a clear activity of that neuron following a tone stimulation. Using a stimulating noise burst does not produce any excitation, as shown in part (b). If the noise burst immediately follows the tone stimulation part (c) the excitation is considerably reduced (feed-backward inhibition). And, shown in part (d), the preceding noise burst completely abolishes the excitation which might be evoked by the sinusoidal tone [23].

.Influence of a static (formant) and a dynamic (transition) combination of stimuli. - Left: Pure tone and frequency modulated sound presented separately - Right: Pure tone and frequency modualted sound

presented simultaneously (or over lapping) [Kallert cit. in Keidle [20] ***1975]***

.Examples of feedforward and feedbackward inhibition. - a) Excitation caused by a tone-burst - b) No excitation caused by a noise-burst - c) Reduction of the excitation by the noise burst presented after the tone-burst - d) Abolition Complete abolition caused by the noise burst presented before the tone-burst [Watanabe and Simada [23] ***1971]***

In our opinion elements of that kind are able to differentiate between a pure vocalic section and a closer consonant-vowel or vowel-consonant configuration as for instance the syllable [ad] or [da].

At least highly specialized measurements with implanted microelectrodes in the awake cat, also done by my colleague [20], are shown in the last figure of this series (Fig. 13). The four different rows always present in their upper recording trace the excitation of the elements (action potentials) and in their lower recording trace the stimulus. Stimuli are the German words [FEIN], [DEIN], [MEIN] sounding similar with exception of the beginning consonant. The three elements of the geniculate body of the awake cat react quite differently. The element P2S1 is stimulated by the word [FEIN] only. Element P2S2 responds to all of other stimuli. And element P2S3 shows excitations only in case of the word [DEIN]. The latter element is also excited if the consonant [d] or [t] emerges in the middle or at the end of a word. Reading a text this element clearly answers if that specific consonant is spoken, even with a raised and changed voice. Scarcely other plosies like [k] and [p] lead to excitations. As far as vowels are concerned, the element P2S4 presented in the lowest row shows a clear selectivity to the vowel [a] and nearly no reaction if the other vowels are presented.

.Telemetric microelectrode records from single units of the geniculate body in the awake cat and corresponding stimulus traces (below) [Kallert cit. in Keidle [20] ***1974]***

Besides the description of the function of the auditory information processing system and continuing the examples mentioned above, some special comparisons should be added.

Special Comparison

There are existing some interesting results of linguistic experiments which can be closely related to the physiological decoding processes of the auditory system, mentioned before.

Selective Adaptation

Selective adaptation experiments show that prolonged repetition of a particular phonetic feature under investigation changes the appropriate detector and lessens its sensitivity [24]. As a consequence, the assignment of a series of stimuli to phonetic categories would be altered; more stimuli would be assigned to the unadapted category. Thus regarding the continuum [ba]-[da]-[ga] being adapted with [ga] results in a marked shift in the category boundary toward [ba]: After adaptation mainly [da] is identified. Moreover, the effect occurred when the adapting stimulus came from a different series from the identification series. This latter result argues against the hypotheses that the adapting unit was the entire phone or syllable [25] and support a sensory explanation of the selective adaptation effect [26]. This latter effect is demonstrated with the scheme of in Fig. 14. Normally the dotted curves of equal loudness (upper part, left) and the dotted intensity functions of the two frequencies A and B (lower part, left) are valid. After adaptation nearby the frequency A, however, the threshold is increased, the equal loudness contours are compressed and the intensity function for that frequency A has become steeper (full drawn). In the normal case a frequency transition form B to A causes a normal change of excitation, as shown by the dotted arrow. After adaptation this change of excitation is reduced as indicated by the broken arrow in the lower part of Fig. 14. With respect to the final point A of the transition, a reduced excitation during a frequency modulation is related to a smaller change of frequency, e.g. from B' to A (broken arrow in the upper part of Fig. 14). Assuming a correct decoding of the second formant (F2) by aid of periodicity analysis the transition B-A is identified as transition B'-A after adaptation (Fig. 14 right). The configuration, normally sounding like [ga], sounds [da] after adaptation. The latter may be caused by different sounds, not only by repeated presentation of [ga], because the different intensity functions are responsible rather than a change in phonetic decision factors.

Physiological explanation of the change of perception in case of adaptation (selective adaptation)

A strong remaining shift in the equal loudness contours and therefore a resting steep intensity function, starting from an increased threshold and reaching the normal ones at higher levels, is well known as recruitment in clinical audiology. In that extreme case, as indicated with Fig. 15, the transition B-A even may cause an inverse change of excitation as normally expected. The excitation behave as if the transition has started from a place of articulation corresponding to the secondary formant frequency F2. And it is hard to say, what is going on in the decoding system processing quasi-static frequencies (vowels) and abnormal changes of excitations caused by transient frequency modulations of that kind. Although the auditory system slowly gets accustomed to these disfunctions those patients have severe difficulties to understand speech correctly, especially when hearing aids are used, which only compensate the threshold shift.

Physiological explanation of the change of perception in case of recruitment

It must be added that the small acoustic differences in some voiceless consonants are better perceived due to the bending of the equal loudness contours. This ability may also be reduced in the case of adaptation and recruitment.

Backward Recognition Masking

Backward recognition masking experiments show that a masking sound following a test stimulus influences that identification of the latter in a complex manner, mainly depending upon the interstimulus interval and duration. There exists a much cited difference between consonants and vowels, in that the masking effect depends upon the relative acoustic confusability, and is not tied to a fundamental distinction between consonants and vowels [27]. In those experiments, the quality of auditory information necessary to distinguish prevocalic stop consonants may have been degraded too much after the processing of the second masking sound to make correct identification possible, whereas the quality of auditory information necessary to identify vowels may generally have a longer survival time [10].

This fact as well as the temporal reversals mentioned previously give rise to the assumption that consonants are processed by analysis of place and vowels by analysis of periodicity, both done in different channels with different decay times. A similar conclusion was made by Doman et al [28] judging their data from a linguistic point of view.

Lack of Invariance and Equivalence of Acoustic Cues

There is the variant behaviour of transitions (if we compare the syllables [du] and [di] the [d] in one case is represented by a falling [u] and in the other case by a rising [i] transition) as a function of vowel context and position in the syllable (comparing the syllables [dud] and [did] the [d] in the final position shows opposite transition cues: rising in one case [u] and falling in the other [I]). In addition exists the equivalence of silence on the one hand, and on the other, such spectral cues as the frequency at which the first formant starts and the extent of the first-formant transition [7] (a rising transition at the beginning of the first formant of [I] causes the listener to hear the stoop [k] in [ki], an appropriate interval of silence between a patch of s-noise and the vowel [I] will also cause the listener to hear [k] in [ski]). Both effects demand highly specialized decoding systems. In fact, they do exist in grouping neural elements with different tuning curves in an inhibitory/excitatory manner as described above.

Let us regard the [ba]-[ga] example mentioned before, where only the rising and falling transitions (upward and downward frequency modulation: FM) make the difference (Fig. 16). If there are higher elements with a tuning curve suitable to be involved in the direction of the second formant of the vowel [a] (part of a F2-a-detector), drawn thick in Fig. 16, they will be both activated by the vowel [a] alone. Let us further assume that the grouping of neural elements is such as two different inhibition zones exist (shadowed in this Fig. 16). Then, during a preceding upward FM-transition caused by the syllable [ba] one element will be inhibited for a longer time period (right part), the other not (left part). This will be reversed during the syllable [ga], thus giving rise for the buildup of a part of a [ga]-detector function rather than a part of a simple [a]-detector in that case.

Behaviour of hypothetical speech decoding elements during variant transitions

In addition it might be possible that corresponding to temporal reversals in the auditory processing the inhibitions show some delay. That means both these types of elements may even act as vowel-detectors for a short time, before one group of them is inhibited by the earlier patch of noise and the specific transition. In other words, some f the vowel detectors may be switched on or off after a while by the accompanying transition.

Another example is given with Fig. 17. As mentioned before, a silent gap between a noise patch, associated with fricative [s] and a vocalic section (vowel [a]) must be small enough to produce the fricative-vowel syllable [sa], shown in the middle part. It should be noted about the format transitions at the beginning of the vocalic section that they are also appropriate, at least approximately, for the stop consonant [t] (left-hand part).

Behaviour of hypothetical speech decoding elements during different noise-patches and silent intervals preceding the same transition

If the silent interval between the patch of noise and the vocalic section is increased (50 ms) the listener will hear this stop consonant [t]where none was before, this time in [sta)].

To try an explanation we may look at neural elements with characteristic frequencies around the formants vowel [a]. In the lower part of Fig. 17 (left side) is shown that such elements with a two-side inhibition may act as a part of a [ta]-detecting circuit, if the transitions are present without a noise. The middle part of this figure indicates what may happen if a patch of s-noise immediately precedes the vocalic section. A feed-forward disinhibition makes the element under consideration more sensitive to the transitions, thus switching it to a part of a [sa]-detector circuit. That kind of feed-forward disinhibition is more reduced as large the gap becomes between the noise and the vocalic section. Therefore the right-hand part of Fig. 17 presents the renewed increase of inhibition and the elements joins again the [ta]-detector systems. As a result the syllable [sta] is heard by the listener, although the place of production are rather the same for [s] and [t]. Comparable to these two examples there are a lot of interesting effects which can easily be explained by simple two-tone suppression or complex interactions resulting in changed sensitivity and tuning curves of neural elements.

Integration Time and Temporal Resolution

Time consuming processes are involved in the build-up of facilitation, inhibition and disinhibition effects, which also show longer delay times. Hence, it is not by chance that we found integration times in the range 200 to 300 ms, based upon primary inhibited elements, when regarding auditory evoked cortical responses in human-subjects [29]. This is just the time of the "running window" assumed by Nooteboom [10] comparing some linguistic and psychoacoustic experimental results. Additionally, dealing with the order threshold for identifying the correct sequence of events, in a recent paper Poppel [30] demands for a cerebral clock generator producing a period of approximately 20-40 ms. Such mechanisms do really exist as shown with Fig. 18, which presents a neuron in the cat's thalamic region. The joint interval histogram of its spontaneous activity has the main peaks for intervals around 20 ms or 40 ms.

A periodicity of 20 ms is present in the activity of neural elements (upper part left) as to be seen in the joint interval histogram (lower part left). Possible coincidence system are shown in the right part

Although the auditory system is able to distinguish binaural time differences in the order of 20µ s a period of 20 ms may play an important role in speech decoding. Not only as a basic clock for coincidence circuits detecting frequencies out of the incoming periodicities (right part of Fig. 18), but also as basic time unit for feed-forward and feed-backward actions in the decoding systems.

Thus, for instance Albert and Bear [31] showed in a study on a patient with word deafness that understanding of speech was made possible for this patient whenever the examiner spoke very slowly.

Minor Importance of Interaction between Articulation and Acoustic Perception

The production of sounds and speech needs a complex cooperation between respiratory, laryngeal and articulatory muscles and control centres (Fig. 19). the post-natal development of control and regulation programs requires the intact function of the auditory system. Early deafness excludes completely learning of speech without extraneous help. Once developed, the neuro-muscular control system works quite autonomously, besides the audiophonate feedback. The kinesthetic reflexional control of phonation is able to realise a phonetical goal without the auditory feedback as well, showing a sufficient accuracy for communication. This has been confirmed with recently published results of binaural masking experiments with vocalists, normal subjects and dysphonic patients [32]. Although in the longer run there exist both, a remarkable difference between the frequency reproduction of the vocalists and the other groups, and an influence of the masking white noise (Fig. 20), the precision of the tone-onset does not differ so much (Fig. 21) and shows no variance caused by the intensity of the masker.

.***Block-scheme of the speech perception and production [Keidle,*** [17] ***1980]***

.***Precision of voice production during binaural masking (HT: error in semi tones, Sanger = vocalist) [Schultz-Coulon*** [32] ***1980]***

.***Precision of voice onset during singing of pure tones in the presence of a masker (HT: semitone, Sanger = vocalist, Standardabweichung = standard deviation) [Schultz-Coulon*** [32] ***1980]***

Pursuit tracking experiments using the electromyogram of the lip muscles [32], studies in dysarthric patients [33], and stutterers [34] reveal the minor importance of auditory feed back as against the surface sensibility of the lingual system respectively. These studies support the assumption of speech as a totally pre-programmed process, if established with the aid of a well operating auditory system.

Concluding Remarks

The fact that a well operating auditory system is the basis for the development of speech, and that late on audio-phonate control seems to be of minor importance for speech production supports the idea that there could be a genetically anchord preadaption of what we can learn to hear in a rapid and complex flow of speech. The various possibilities to switch on and to connect neural decoding systems enable us to hear "speech as speech" because there exist internal templates for the distinctive features of a continuous flow of spoken language. This may explain the findings of categorical perceptions of phonems by infants, even before their babbling stage [35]. The preverbal child with his auditory decoding system is able to classify speech sounds, that he cannot yet articulate.

However, we should not believe in a fixed set of complex "feature detectors" or decoding systems. Things are mostly variable in the brain, especially in the awake organism. Thus we observe changing dominant processing of a continuous stimulation. Highly variable decoding systems are at least necessary to avoid or suppress the influence of other parameters as for instance speaker's absolute pitch, masking noise etc.

Summarizing, we have a rapid, highly flexible, adaptive system based upon decoding mechanisms mentioned earlier. A system of auditory word which are activated or deactivated if the acoustic input matches or dismatches their internal specifications. There is not central decision mechanism selecting the best candidate after some time of processing out of a given lexicon. On the contrary the beginning of a word triggers in a parallel a list of possible candidates, which begin with the particular acoustic sequence. Each of these activated word-decoding units will then continue to monitor the subsequent acoustic input signals. As the acoustic signal proceeds, more and more word-candidates will remove themselves from the pool, until only one candidate remains. This is a real-time process until the best fit survives and thus word recognition often is completed before the acoustic ends of words. The obvious perfect working of our word recognition system must not nearly be so good, because always the possibility exists to inquire again. And in most cases we can hope that our partner will be so kind and will repeat, thus establishing a real communication not with the aim to understand speech but to understand the fellow-being.

1.Chomsky N, *Language and Mind. Harcourt, Brace World, New York*1968

2.Teuber H L, The brain and human behaviour
*In: R Held, H W Leioo witz and H L Teuber (Ed) Perception: Handbook of Sensory Physiology. Vol 8, Springer, Berlin-Heidelberg-New York*1978

3.Fowler C, Rubin P, Remez R, Turvey M, Implications for speech production of a general theory of action
*In: B ButterWorth (Ed). Language Production. Academic Press, New York*1978

4.Liberman A M & Studdert-Kennedy M, "Phonetic perception"
*In: R Held, W Leibowitz & H L Teuber (Ed) Perception: Hand-book of Sensory Physiology Vol. 8 Springer, Berlin-Heidelberg-New York*1978

5.Fujimura O, Syllable as a unit of speech recognition
IEEE Trans. Acoust. Sp. Sig. Proc Page: 23: 82-87, 1975

6.Mattingly L G, Syllable synthesis
J. Acoust. Soc. Am Page: 60: 75A, 1976

7.Summerfield A Q & Bailey P J, On the dissociation of spectral and temporal cues for stop consonant manner
J. Acoust. Soc. Am Page: 61, A, 1977

8.Barry W J, Temporal aspects of speech production: Primary vs. secondary processing
*In: W J Barry & K J Kohler (Ed). "Time" in the Production and the Perception of Speech. Report. Phonetics Dept., University of Kiel*1979

9.Blumstein S E & Stevens K N, Perceptual and onset spectra for stop consonants in different vowel environment
J. Acoust. Soc. Am Page: 67: 648-662, 1980

10.Nooteboom S G, The time course of speech perception
*In: W J Barry & K J Kohler (Ed) "Time" in the Production and the Perception of Speech. Report, Phonetics Dept. University of Kiel*1979

11.Huggins A W F, Temporally segmented speech and "echoic" storage
*In: A Cohen & S G Nooteboom (Ed) Structure and Process in Speech Perception. Springer, Berlin-Heidelberg-New York*1975

12.Slis I H & D J P J van Nicrop, On the forward masking thresholds of vowels in VC-combinations. Institute of Perception Research, Eindhoven
Ann. Prog. Report Page: 5: 68-72, 1975

13.Butcher A, *Experiment zur Pausenperception, Arbeitsberichte. (Universitat Kiel (AIPUK)*1973

14.Massaro D W, Preperceptual images, processing time, and perceptual unit in auditory perception
Psychological Review Page: 79: 124-145, 1972

15.Summerfield M R C, Timing in phonetic perception: Extrinsic or intrinsic
*In: W J Barry and K J Kohler (Ed) "Time" in the Production and the Perception of Speech. Report. Phonetics Dept., University of Kiel*1979

16.Keidel W D, *Psyiologie des Gehors. Georg Thieme, Stuttgart*1975

17.Keidel W D & Kallert S, Physiologis des afferenten akustischem Systems
*In: J. Berendes: R Link & F Zollner (Eds.) Hals-Nasen-Ohren-Heilkunde in Praxis und Klinik Ohr I: Band 5, G. Thieme, Stuttgart*1979

18.Spreng M, Horen (Aufbau und Funktion des Gehors)
*In: P Groner & H Schmidtke (Ed.) Handbuch der Ergonomie. Luftfahrtverlag W Zurel, Steinebach/Worthsee*1978

19.Suga N, Feature extraction in the auditory systems of bats

*In: A R Moller (Ed.) Basic Mechanisms in Hearing. Academic Press, New York*1973
20.Kallert S, *Telemetrische mikroelektrodenuntersuchungen am Corpus geniculatum medial der wachen Katze, Habilitationsschrift. Universitat Erlangen*1974
21.Keidel W D, Recent advances in information processing within the auditory system
*In: W D Keidel, W Handler & M Spreng (ed.) Cybernetic and Bionics. R. Oldenburg, Munchen-Wien*1974
22.Aitkin L M & Dunlop C W, Interplay of excitation and inhibition in the cat medical geniculate body
Journal of Neurophysiology Page: 31: 44-61, 1968
23.Watanbe T & Simida Z, Auditory temporal masking: an electrophysiological study of single neurons in the cat cochlear nucleus and inferior colliculus
Japanese Journal of Physiology Page: 21: 537-550, 1971
24.Ades A E, Adapting the property detectors for speech perception
*In: R J Wales & E Walker (Ed.) New Approaches to Language Mechanisms, North Holland, Amsterdam*1976
25.Sawusch J R & Pisoni D B, Category boundaries for linguistic and neolinguisitc dimensions of the same stimuli
*Paper presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California*1973
26.Eimas P D, Developmental aspects of speech perception
*In: R Held, H W Leibowitz & H L Teuber (Ed.) Perception: Handbook of Sensory Physiology Vol. 8 Springer, Berlin Heidelberg-New York*1978
27.Darwin C J & Baddeley A D, Acoustic memory and the perception of speech
Cognitive Psychology Page: 6: 41-60, 1974
28.Dorman M F, Kewley-Port D, Brady S & Turvey M T, Two processes in vowel recognition: Interferences from studies of backward masking
Haskins Lab. Stat. Rep. Speech Res Page: SR-37: 233-253, 1974
29.Spreng M, AER-review emphasizing the temporal component (AII) and stimulus induced training of primary inhibited elements
In: M Hoke & E de Boer (Ed.) Models of the Auditory System and Related Signal Processing Techniq
Page: Suppl 9, 1979
30.Poppel E, Time perception
*In: R Held, H W Leibowitz & H L Teuber (Ed.) Perception: Handbook of Sensory Physiology, Vol. VIII, Springer, Berlin-Heidelberg-New York*1978
31.Albert M L & Bear D, Time to understand a case study of word deafness with reference to the role of time in auditory comprehension
Brain Page: 97: 373-384, 1974
32.Schultz-Coulon H J, Tanhohen-und Lautstarkeanderungen de Sprech-und Singstimme bei Storung der audiphonatorischen Kontrolle
*In: M. Spreng (Ed.) Interaktion Zwischen Artikulation and akustischer Perzeption. Thieme, Stuttgart-New York*1980
33.Spreng M, Bemerkungen zum Reglerverhalten bei akustischer Ruckkoppelung unter Berlarmung und mit Sprachstorungen
*In: M. Spreng (Ed.). Interaktion zwischen Artikulation and akustischer Perzeption. Thieme, Stuttgart-New York*1980
34.Linke D, Vorprogrammierung and Ruckkoppelung bei der Sparche
*In: M Spreng (Ed.) Interaktion zwischen Artikulation and akustischer Perzeption. Thieme, Stuttgart-New York*1980

35.Kittel G,   Stottern bei Horstorungen
*In: M Spreng (Ed.). Interaktion zwischen Artikulation and akustischer Perzeption. Thieme, Stuttgart-New York*1980
36.Fodor J A, Garrett M & Bril S. Pi ka pu,   the perception of speech sounds by prelinguistic infants
Perception and Psychophysics   Page: 18: 74-78, 1975