Interface, 2 (1973) pp. 141-156.

A Minimum Description of the Linguistic Sign Repertoire (First Part)

WERNER KAEGI

I. INTRODUCTION

I am interested in the question as to whether the signs (LS) of the linguistic repertoire (LR) play a part in the construction of musical systems.

In order to be able to answer this question, the first requirement is a metrical description of the LS-generating classes of physical signals. This report deals with the first part of a minimum description developed by myself. I am of course perfectly aware of the vagueness of the term "LR". The question remains open as to whether there is a common LR at all for all natural languages spoken by human beings. At present, too, I am dispensing with the phonological problem of the LRs which are specifically orientated towards particular languages. However, I expect it to be possible, within the framework of my description, and by fixing in a series of experiments the argumentory values of the variables, to describe the LR of every conceivable natural language in such a way that it can be understood.

The project of describing an LR metrically consists of assigning an *external* representation (physical signals or their description) to an *internal* representation (generated LSs). Nowadays a spectrum, i.e. an external representation A(f), is generally assigned to the external representation A(t) of the amplitude in time of a signal (Fourier transformation: A(t):=A(f)). In this new representation A(f), the signal is finaly assigned to the generated LS, i.e. to the internal representation. (Such-and-such is the spectrum of a vowel *a*, such-and-such is the 1st formant of the phoneme /*a*/, etc.). In the following, I have dispensed with the Fourier transformation. This means that I assign the external representation A(t) directly to the internal representation. The basic principles of my description therefore come exclusively from the A(t) domain. The assignment external-internal must always be verified by simulations and tests with apparatus.

II. AN LS-GENERATING SIGNAL MODEL

Briefly, this is what caused me to select the basic principles of my description. Observations of LS-generating signals with respect to alterations in their shape by means of a square window resulted in the following, among others: the amplitude in time of signals which generate elements of one of the phonemes /u/, /o/, /a/, (1)

Received: 1 october 1973.

can be reduced, regardless of the duration of the period, to a time interval T without the generated LS being affected (Fig. 1 a and c). The reduced signal can even be idealized to the pulse form of a $\sin^2$ (Fig. 1d). Whether a vowel *u*, *o* or *a* is generated only depends in this case on the pulse width T of the $\sin^2$ pulse. ([2]) This provides the point of departure for my model.

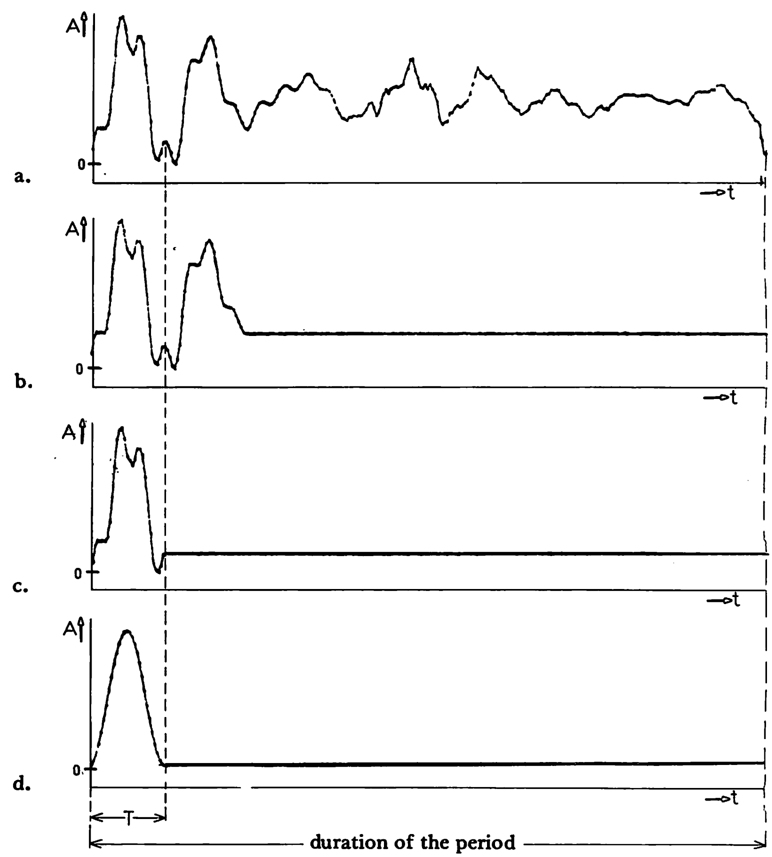


Fig. 1. Data reduction.

a. Amplitude curve in time of a very low spoken vowel ɑ̆ *a*. Duration of period = 270 samples (1 sample =40μs) =10.8ms =approx. 92Hz.

b. Reduction of the characteristic amplitude curve to 63 samples. Samples 64-270 are given the amplitude value of the 63rd sample.

c. Reduction of the characteristic amplitude curve to 30 samples. Samples 31-270 are given the amplitude value of the 30th sample.

d. Sin^2 pulse with pulse width T=30 samples. Duration of period = 270 samples. Signals b), c) and d) also generate the vowels *a*.

Basic principles

Imagine two externally triggered tone-burst generators (single cycle $\sin^2$) with their outputs linked by the logical function OR. Both outputs are first characterized by the following five variables:

T pulse width of one $\sin^2$ pulse,
n number of $\sin^2$ pulses per cycle,
AP amplitude of $\sin^2$ pulse
M mean (i.e. $n*T+M=$ duration of one cycle),
D deviation of modulation of M (i.e. M(t) varies between M+D and M-D).

In order to distinguish the variables of the two outputs, they are followed by the numbers 1 and 2. The ordinal number k of the pulses is indicated by indices:
$P1_1, P1_2, P1_3, \ldots$ and $P2_1, P2_2, P2_3, \ldots$ (Fig. 2).
Variables are accordingly: T1, n1, M1, D1 and T2, n2, M2, D2, $AP1_1:AP2_1$ (with n1 and n2 $\in$ *IN*).
Constants for now are: single cycle $\sin^2$ ($n_i=1$), $AP1_1:AP1_2$ =const.=3:4, $AP2_k$ const., modulation of M = random modulation.

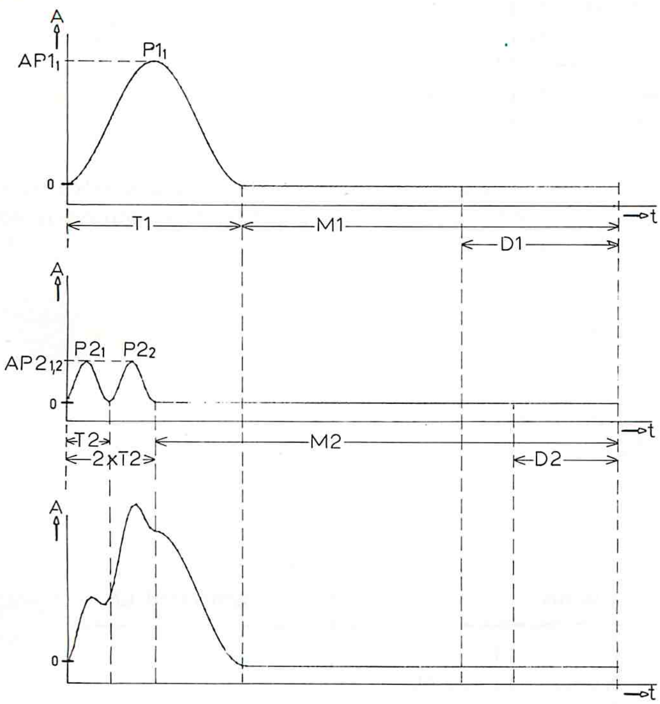


Fig. 2. *Top:* Output 1, single $\sin^2$ pulse. *Middle:* Output 2, double $\sin^2$ pulse with $AP2_k$const.<$AP1_1$. *Bottom:* Logical som. Further explanations in text.

The range of values of the variables

Since the following values were first obtained in an analog manner and then digitally too, (3) they are given in both milliseconds and as the number of samples (=S, one sample =40μs). The ms indication might make matters easier for the reader. The limits of the value-ranges of the described LSs are not sharply defined.

Table 1. With the assignments:

$n1 = 1$
$0 < M1 < \infty$ (4)
$D1 = 0$
$T2 = \infty$

T1 (5) in S	in ms	generated LS
$02 \leqslant T1 \leqslant 05$	$0.08 \leqslant T1 \leqslant 0.20$	*ä*-like 1
$05 \leqslant T1 \leqslant 10$	$0.20 \leqslant T1 \leqslant 0.40$	*ä*-like 2
$10 \leqslant T1 \leqslant 21$	$0.40 \leqslant T1 \leqslant 0.84$	$/a/^{br}$ *ä*-like (6)
$21 \leqslant T1 \leqslant 42$	$0.84 \leqslant T1 \leqslant 1.68$	$/a/^{br}$
$42 \leqslant T1 \leqslant 84$	$1.68 \leqslant T1 \leqslant 3.36$	$/o/^{br}$
$84 \leqslant T1 \leqslant 168$	$3.36 \leqslant T1 \leqslant 6.72$	$/u/^{br}$

Vowels are given the indication *br*= "brut" if they are generated by single $\sin^2$ pulses. They are typical of the phoneme */u/* (Fig. 4 *bottom*), but rarely occur in this form for */o/* and */a/*.(7).

Table 2. With the assignments:

$n1 = 2$
$AP1_1 : AP1_2 = \text{const.} = 4{:}3$ (8)
$0 < M1 < \infty$ (4)
$D1 = 0$
$T2 = \infty$

T1 (5) in S	in ms	generated LS
$10 \leqslant T1 \leqslant 21$	$0.40 \leqslant T1 \leqslant 0.84$	$/a/^{cl}$ *ä*-like (6)
$21 \leqslant T1 \leqslant 42$	$0.84 \leqslant T1 \leqslant 1.68$	$/a/^{cl}$
$42 \leqslant T1 \leqslant 84$	$1.68 \leqslant T1 \leqslant 3.36$	$/o/^{cl}$
$84 \leqslant T1 \leqslant 168$	$3.36 \leqslant T1 \leqslant 6.72$	$/u/^{cl}$

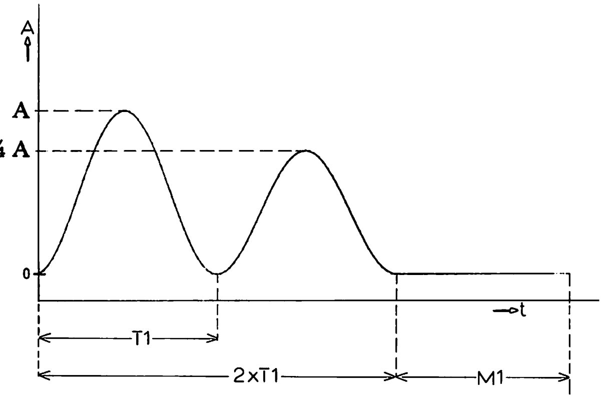


Fig. 3. Double $\sin^2$ pulse with $AP1_1 : AP1_2 = 4:3$.

Vowels are given the indication *cl*= "classic" if they are generated by double $\sin^2$ pulses (Fig. 3). Although they clearly belong to an LS, they also contain the sound of a neighbouring vowel. Since n1=2, the following applies, for example, to all signals of the phoneme $/a/^{cl}$ (as long as they are not borderline cases): 2∗T1 = pulse width of $/o/^{br}$. (According to the rules of bel canto "there is an *o*-vowel sound in the classic /*a*/.") In the LR the vowels *o* and *a* generally occur in the classic form (Figs. 4 *middle* and *top*, compare also Figs. 1a and b). However, this can only apply to the vowel *u* if its fundamental is low enough, since M1 cannot have a negative value.

Table 3. With the assignments:

$AP1_1 : AP1_2 = \text{const.} = 4{:}3$ iff $n1 = 2$
$0 < M1 < \infty$ (4)
$D1 = 0$

$$T2 = \frac{T1 * n1 * 2}{3}$$

$$\frac{AP1_1}{4} \leqslant AP2_1 \leqslant AP1_1$$

$n2 = 1$
$M2 = n1 * T1 + M1 - n2 * T2$
$D2 = 0$

T1 in S	in ms	n1	generated LS
42 ⩽ T1 ⩽ 84 (9)	1.68 ⩽ T1 ⩽ 3.36	2	/n/
84 ⩽ T1 ⩽ 168 (9)	3.36 ⩽ T1 ⩽ 6.72	1	/m/

Compare Fig. 5.

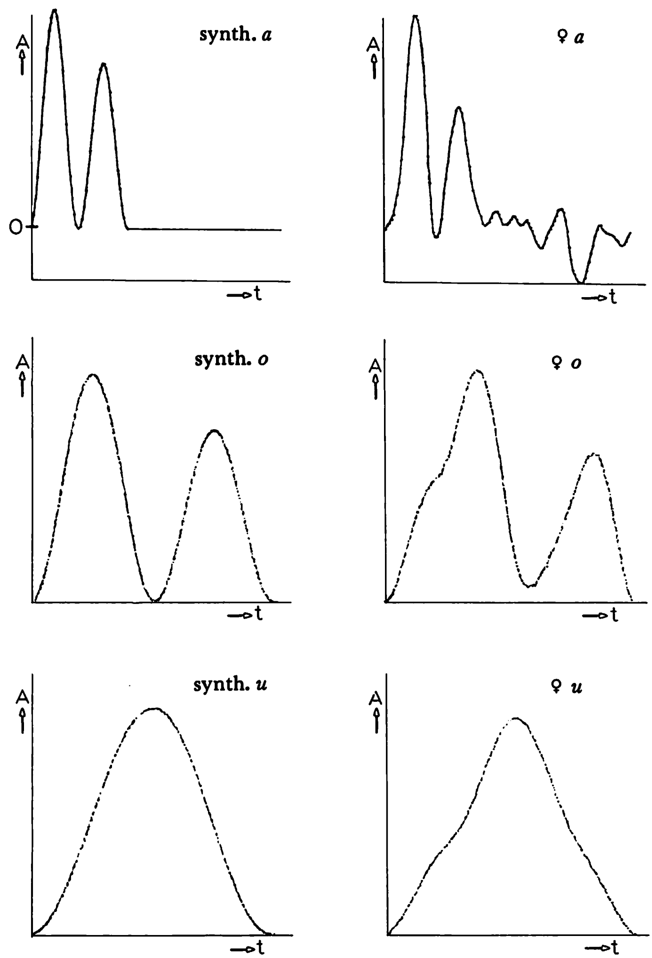


Fig. 4. Comparison of synthetically generated (*left*) and natural speech signals (*right*).
Top left: Synth. vowel *a* with T1=25 samples (S), M1=80 S.
Right: ♀ *a* with duration of period =130 S.
Middle left: Synth. vowel *o* with T1=67 S, M1=6 S.
Right: ♀ *o* with duration of period =140 S.
Bottom left: Synth. vowel *u* with T1=104 S, M1=6 S.
Right: ♀ *u* with duration of period =110 S. (Compare ([7])).

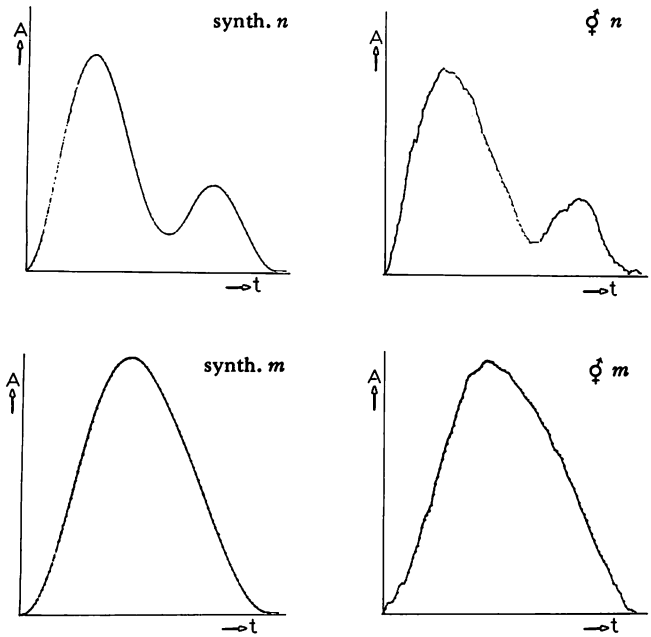


Fig. 5. Comparison of synthetically generated (*left*) and natural speech signals (*right*).
Top left: Synth. sonant *n* with T1=52 S, M1=4 S, T2=69 S, $AP2_1 = AP1_1$.
Right: ⚥ *n* with period duration =108 S. (⚥ = child.)
Bottom left: Synth. sonant *m* with T1=104 S, M1=4 S, T2=69 S, $4*AP2_1=AP1_1$.
Right: ⚥ *m* with period duration =108 S.

Table 4. With the assignments:

$AP1_1 : AP1_2 = \text{const.} = 4:3$ iff nl = 2
$0 < M1 < \infty$ (4)
$D1 = 0$

$$AP2_k \text{ const.} \leqslant \frac{AP1_1}{10}$$

$n2*T2 \leqslant nl*T1$
$M2 = nl*T1 + M1 - n2*T2$
$D2 = 0$

T1 in S	in ms	nl	T2 in S	in ms	generated LS
$21 \leqslant T1 \leqslant 45$	$0.84 \leqslant T1 \leqslant 1.80$	2	$6 \leqslant T2 \leqslant 10$	$0.24 \leqslant T2 \leqslant 0.40$	/ä/ (10)
$42 \leqslant T1 \leqslant 84$	$1.68 \leqslant T1 \leqslant 3.36$	2	$6 \leqslant T2 \leqslant 10$	$0.24 \leqslant T2 \leqslant 0.40$	/e/ (10)
$84 \leqslant T1 \leqslant 168$	$3.36 \leqslant T1 \leqslant 6.72$	1	$6 \leqslant T2 \leqslant 10$	$0.24 \leqslant T2 \leqslant 0.40$	/i/ (10)
$42 \leqslant T1 \leqslant 84$	$1.68 \leqslant T1 \leqslant 3.36$	2	$11 \leqslant T2 \leqslant 15$	$0.44 \leqslant T2 \leqslant 0.60$	/ö/ (11)
$84 \leqslant T1 \leqslant 168$	$3.36 \leqslant T1 \leqslant 6.72$	1	$11 \leqslant T2 \leqslant 15$	$0.44 \leqslant T2 \leqslant 0.60$	/ü/ (11)

Compare Figs. 6 and 7.

The other LSs can also be given within the same framework, but they will be described in a future number of INTERFACE. I shall restrict myself here to the description of the /*s*/-generating signals: $T1=\infty$, $1<T2\leqslant 3$ (in number of samples), $n2=1$, $T2+M2=4$ (in number of samples), $D2=1$ (in number of samples) for voiceless /*s*/ (Fig. 8). If instead of $T1\neq\infty$ there is the following statement: $42\leqslant T1\leqslant 168$, with $0<M1<\infty$ and $n1=1$, the voiced /*s*/ is generally described.

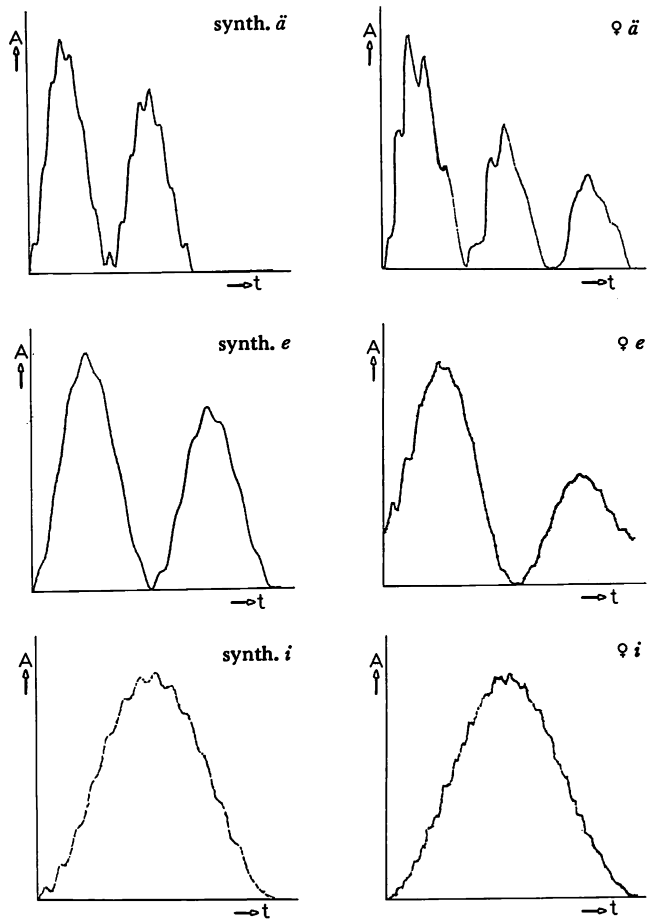


Fig. 6. Comparison of synthetically generated (*left*) and natural speech signals (*right*).
Top left: Synth. vowel *ä* with T1=45 S, M1=44 S, T1=6 S, $10*AP2_k$const.$=AP1_1$.
Right: ♀ *ä* with duration of period =134 S.
Middle left: Synth. vowel *e* with T1=64 S, M1=4 S, T2=9 S, $20*AP2_k$const.$=API_1$.
Right: ♀ *e* with duration of period =132 S.
Bottom left: Synth. vowel *i* with T1=116 S, M1=6 S, T2=8 S, $20*AP2_k$const.$=AP1_1$.
Right: ♀ *i* with duration of period =122 S.

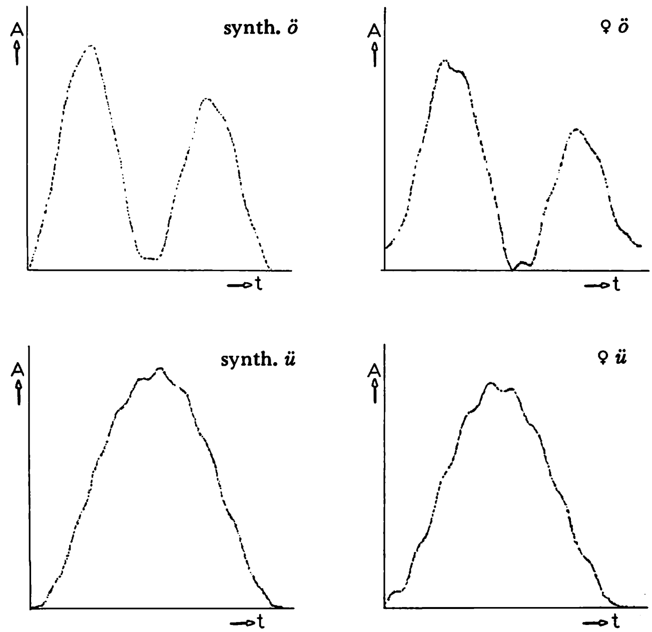


Fig. 7. Comparison of synthetically generated (*left*) and natural speech signals (*right*).
Top left: Synth. vowel ö with T1=63 S, M1=6 S, T2=9 S, 20*AP2$_k$const.=AP1$_1$.
Right: ♀ ö with duration of period =132 S.
Bottom left: Synth. vowel ü with T1=116 S, M1=6 S, T2=9 S, 20*AP2$_k$const.=AP1$_1$.
Right: ♀ ü with duration of period =122 S.

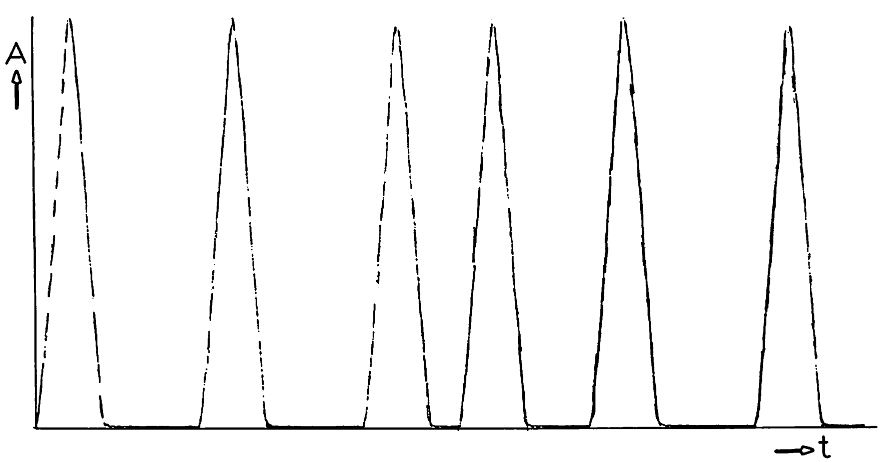


Fig. 8. Synth. consonant *s* voiceless, with T1 =∞, T2=2S, M2=2S, D2=1S.

Transformations

Transformations can be obtained from the given description. These are indications for deriving from certain LSs certain other LSs. All the transformations together finally result in a general indication for deriving from any LS any other LS. Here are five transformations:

I $\quad n1*T1 \; := n1*T1*2.$

II $\quad n1 = 1 \; := n1 = 2$, with $AP1_1 : AP1_2 = \text{const.} = 4:3.$

III $\quad T2 = \infty \; := T2 = \frac{T1*n1*2}{3},$ with: $\frac{AP1_1}{4} \leqslant AP2_1 \leqslant AP1_1,$
$n2 = 1,$
$M2 = n1*T1 + M1 - n2*T2,$
$D2 = 0.$

IV $\quad T2 = \infty \; := 11 \leqslant T2 \leqslant 15,$ ([12]) with: $AP2_k \text{ const.} \leqslant \frac{AP1_1}{10},$
$n2*T2 \leqslant n1*T1,$
$M2 = n1*T1 + M1 - n2*T2,$
$D2 = 0.$

V $\quad T2 = \infty \; := 6 \leqslant T2 \leqslant 10,$ ([12]) with: $AP2_k \text{ const.} \leqslant \frac{AP1_1}{10},$
$n2*T2 \leqslant n1*T1,$
$M2 = n1*T1 + M1 - n2*T2,$
$D2 = 0.$

The applications for transformations I – V are:

for I : *ä*-like 1 : = *ä*-like 2 : = $/a/^{br}$ *ä*-like : = $/a/^{br}$: = $/o/^{br}$: = $/u/^{br}$,
$/a/^{cl}$ *ä*-like : = $/a/^{cl}$: = $/o/^{cl}$,

for II : $/a/^{br}$ *ä*-like : = $/a/^{cl}$ *ä*-like, $/a/^{br}$: = $/a/^{cl}$, $/o/^{br}$: = $/o/^{cl}$,

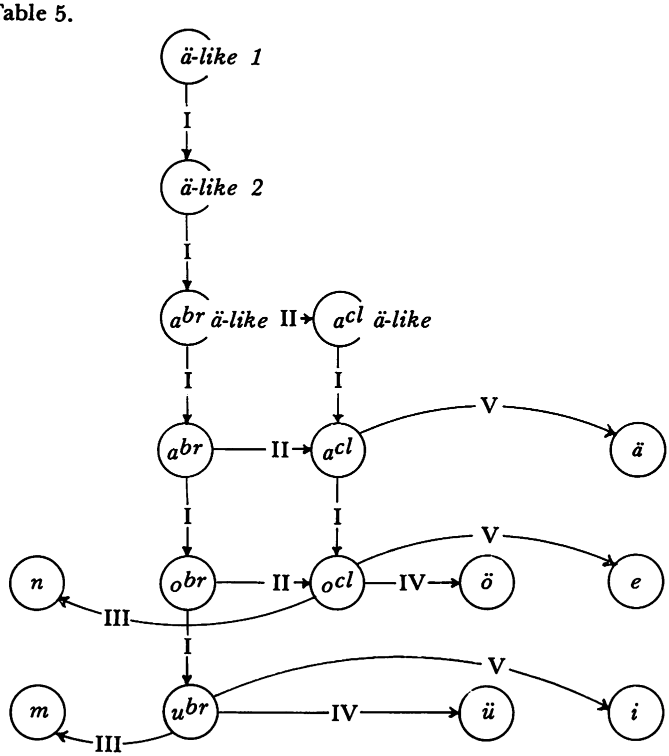
for III : $/o/^{cl}$: = */n/*, $/u/^{br}$: = */m/*,

for IV : $/o/^{cl}$: = */ö/*, $/u/^{br}$: = */ü/*,

for V : $/a/^{cl}$: = */ä/*, $/o/^{cl}$: = */e/*, $/u/^{br}$: = */i/*.

This can be summarized in the following diagram (Table 5):

Table 5.



If we start from an LS selected at random, the only thing required for deriving the description of any other desired LS is the description of the first LS (in accordance with (Tables 1-4) and the corresponding transformation(s) (according to Table 5). (At present, of course, the method is restricted to the LSs already described.)

An example:
Given: /*a*/^{br}, desired: /*i*/.
Description of *a* ∈/*a*/^{br} (according to Table 1, interpreted):
T1=24[12]
n1=1
M1=176[12]
D1=0
T2=∞
Transformations (according to Table 5):
/*a*/^{br}→I,→I,→V :=/*i*/.
Execution (both transformations I are executed in one step):

$$\begin{bmatrix} T1 = 24 \\ n1 = 1 \\ M1 = 176 \\ D1 = 0 \\ T2 = \infty \end{bmatrix} := \begin{bmatrix} T1 = 24*2*2 \\ n1 = 1 \\ M1 = 104 \\ D1 = 0 \\ T2 = \infty \end{bmatrix} := \left[\begin{array}{l} T1 = 96 \\ n1 = 1 \\ M1 = 104 \\ D1 = 0 \\ 6 \leqslant T2 \leqslant 10^{(12)} \\ AP2_k \text{ const.} \leqslant \dfrac{AP1_1}{10} \\ n2*T2 \leqslant n1*T1 \\ M2 = n1*T1 + M1 - n2*T2 \\ D2 = 0 \end{array}\right. \quad \begin{array}{|l} \text{interpreted}^{(13)}: \\ T1 = 96 \\ n1 = 1 \\ M1 = 104 \\ D1 = 0 \\ T2 = 8 \\ AP2_k \text{ const.} = \dfrac{AP1_1}{20} \\ n2 = 12 \\ M2 = 104^{(12)} \\ D2 = 0 \end{array}$$

This can also be done the other way round (against the arrow's direction). The transformation is then read from the expression at the right towards the expression to the left of the sign ":=". The previous example backwards may serve as an example: given /*i*/, desired /*a*/^{br}. The chain of transformations is in this case: /*i*/←V,←I,←I := /*a*/^{br}. Mixed chains are also possible, e.g.: /*m*/←III,←I,→II,→IV :=/*ö*/.

III. DISCUSSION

Obviously, the smaller the number of required dimensions, the greater is the value of a metrical minimum description of the LR. Within the framework of my description, the number of dimensions for the phonemes /*u*, *o*, *a*/ is one (T1), for /*m*, *n*/ two (T1, n1) and for the phonemes /*ä*, *e*, *i*, *ö*, *ü*/ three (T1, n1, T2). Should the assumption: $M1 \to \infty$ be rejected, the number is increased by one dimension (M1) in each case. This description only fulfills its purpose, of course, if it can be applied to all other LSs, not just the ones referred to here. At the present stage of my investigations, this appears to be the case. The number of required dimensions still seems to be astoundingly small here too; for instance, three dimensions are sufficient to describe the voiceless /*s*/ (T2, M2, D2), and one and two respectively are added for the voiced /*s*/ (T1 and possible M1) as demonstrated above. The simplicity of this description is definitely very promising. As experiments have shown, it enables an extremely simple technique for speech synthesis([14]), on the basis of which the description of the LR can not only be verified by means of simulation, but can also be continually refined in terms of a feedback circuit. One of the advantages is that M1 always occurs in this description as a free variable for all voiced phonemes, so that the question of speech melody (fundamental frequency contours) can be studied independently of the dimensions which characterize the LSs. The value-range of M1 can be extended beyond the range of the human voice (e.g. $f_0^{-1} < n1*T1+M1 < \infty$, with f_0=70Hz for /*u*, *o*, *a*, *ö*, *ü*, *ä*, *e*, *i*, *m*, *n*/, $0 < n1*T1+M1 < f_0^{-1}$, with f_0 = 1kHz for /*s*/ voiced, etc.), and f_0 can be replaced by musical pitch-time-duration structures whose metrical description was, of course, developed by music a long time ago. Something similar is conceivable for D1. Whatever the sounding results of such operations and ones like them may be, one thing is always the same in this case: *they all have the conceptual property of being elements of an LS.* (It becomes clear here that my description can also be given within the framework of predicative logic.)[15] In this way the world of sound can be experimentally explored in a systematic fashion, and ultimately the question as to the relationship of the linguistic and "musical" repertoire of signs can be answered. It will then be time to decide whether the LSs are really the optimal operators for an exhaustive classification of sounds, as I already suspect today.

Acknowledgments: My thanks are due to S. Tempelaars who wrote the necessary computer programmes for me. The research was supported by the Fonds National Suisse de la Recherche Scientifique.

Bibliography

1. Rohlfs G. Grammatica storica della lingua italiana e dei suoi dialetti. Einaudi, Torino 1966, p. 5 ff.
2. Kaegi W. A New Approach to a Theory of Sound Classification. Interface 1 (1972) p. 104.
3. Carnap R. Der logische Aufbau der Welt. Berlin 1928, p. 13 and p. 100 ff.

Notes

1. */u/* is for the time being the set of open *and* closed *u* vowels; the same applies to */o/* and */a/*. This division corresponds to the vowel pattern in classical Latin. [1]. Examples for */a/*: (Italian) It*a*li*a*; for */o/*open: (German) R*o*ss, closed: (French) *eau*; for */u/*open: (German) M*u*tter, closed: (French) T*ou*l*ou*se.

2. In [2] I posed the hypothesis according to which every period of a vowel *V*, offered in the form of a one-shot, is an element of the set */V/*. If we now also admit the assumption: $M \rightarrow \infty$ (with M= period duration minus n*T), things are tightened up, since the shape of the signal is reduced to a single cycle $\sin^2$. Compare test results on page 157.

3. Analog with the tone-burst generators HP 3000A phase lock 3302A, and Wavetek VCG 116; digital with the PDP-15 computer of the Institute of Sonology, Utrecht State University. See page 155 for the programme VOSIM 2 which was used.

4. Compare note (2)

5. Compare test results page 157.

6. A sound between */a/* and */ä/*, as in Eastern New England pronunciation of: *a*sk.

7. The vowel *u* is often found with: 3*T2=T1, $10 * AP2_k$const.=API_1 instead of $T2=\infty$. This makes the vowel somewhat brighter. Compare *fig. 4, bottom right.*

8. Idealized decay constant.

9. The range may be extended up to $21 \leqslant T1 \leqslant 168$. We intend to test with the Kruskal method, but have not yet done so.

10. */e/* and */i/* are the sets of open *and* closed *e* and *i* vowel sounds. This completes the vowel pattern of classical Latin. Compare [1]. Examples for */e/*open: (German) B*e*tt,closed: (French) n*é*, *épée*; for */i/*open: (German) F*i*sch, freundl*i*ch, closed: (French) m*i*d*i*,

11. */ä/* is the set of intermediate sounds between open */a/* and open */e/*: (German) Tr*ä*ne. */ö/* and */ü/*are the sets of open *and* closed *ö* and *ü* vowel sounds. Examples for */ö/*open: (German) H*ö*lle, closed: (French) n*eveu*; for */ü/*open: (German) f*ü*ttern, closed: (French) m*u*r.

12. Number of samples.

13. */i/*implies *i*.

14. It is also an interesting point of departure for pattern recognition.

15. In this connection I should like to remark that the Kruskal test (see page 157) is constructed on the undefined basic principle of "simularity". One could however try to establish this principle in the framework of logic of relations, as Carnap did in [3], i.e. in our case: Simularity :=if $\{x,y\} \in K$, then $xRx \wedge yRy$, and also $xRy \wedge yRx$ (with K:=set of pulses under consideration, R:=Relation).

APPENDIX

S. TEMPELAARS

In order to generate the signals digitally which are described in this article, the programme VOSIM was written, the second version, VOSIM 2, being used. With this programme the PDP-15 produces a signal (via DA conversion) consisting of a repeated pulse which is interrupted after a stated number of pulses by a pause, the duration of which can also be stated.

All data connected with the time structure are given in the form of the *number* of increments of the real time clock used for synchronization (Hewlett Packard Generator 4204A set at 25kHz, a clock increment thus corresponding to 40 μs).

Although at first experiments with various pulse shapes were made, and although the possibility of an off set was also built in, the final decision was to use two pulse shapes: a single $\sin^2$ pulse and a double $\sin^2$ pulse (amplitude ratio 4:3), 0 volt being kept as off set.

A multiple $\sin^2$ pulse can be added to this signal with the understanding that (in contrast to the general formalism) this series begins together with the main pulse and that it can not be longer than the main pulse. The main pulses do not follow in direct succession. The interval between them can be constant or variable. In the latter case we have modulation which can be according to a sine function or random. In both cases the average value (M) must be stated, also the deviation (D) and, in the case of sine modulation, the number of periods within which a complete modulation cycle has to be finished. In the case of random modulation a random number of clock increments between M+D and M-D is selected for the interval between the pulses.

At present VOSIM 3 is being developed, in which both pulse signals can be treated autonomously. The source programmes can be obtained from the Institute of Sonology, Utrecht.

Dr. W. Kaegi
c/o Institute of Sonology
Utrecht State University
14-16 Plompetorengracht
Utrecht, Netherlands