Acta Paediatrica Hungarica, 24 (4), pp. 315—321 (1983)

A new acoustic method for the discrimination of monozygotic and dizygotic twins

G FORRAI, G GORDOS

District 14 Policlinic for Children, and Institute for Telecommunication Electronics, Polytechnical University, Budapest, Hungary

The close similarity of the voices of consanguineous persons has suggested that there must exist hereditary phonetic parameters. To test this hypothesis, a method for characterizing numerically the similarity or dissimilarity of voices of twins with zygosity previously established by anthropological methods was elaborated. By speech processing and appropriate classification strategy the method has attained the discriminating power of blood-group determination for zygosity. The hereditability of certain physical parameters of the human voice has also been suggested.

Comparison of concordance and discordance for a trait or disease between monozygotic resp. dizygotic twins is a widespread method in human genetic research aiming at the degree of heritability of the trait in question. If concordance within monozygotic twins significantly exceeds that observed in dizygotic twins of the same gender, it may be anticipated that the condition is more under genetic than environmental control.

Therefore, discrimination between monozygous and dizygous twins is of great importance. It can be carried out by morphological examination of the placenta at birth; this method is, however, not practicable with twins of adult age or when data concerning the placenta are no more available. Beside a series of anthropological tests, extensive blood-group determinations are most wide-spread. In zygosity studies, a like-sexed pair of twins is a priori regarded as monozygous as long as any difference in an inherited blood-group trait has not been found. The discriminative power of the method can be increased by using blood-groups as many as possible. Absolute safety in exclusion of dizygosity cannot be achieved. This and the large quantity of blood necessary for extensive blood-group studies lend importance to research into painless methods that can replace blood-group determinations or further increase their discriminating power.

It is well-known that the voice of closely related persons is often confounded, especially over the telephone where the individual nuances are partly lost. This has prompted us to conceive the idea that some parameters of human voice are genetically determined and, if so, monozygous twins have a higher intra-pair similarity of voice than dizygous like-sexed twins. If this similarity can be quantified and made measurable, a new acoustic method can be elaborated. The hypothesis can be tested by applying the voice-based method to twins whose zygosity has been established by detailed blood-group analysis.

The blood-group studies utilized three erythrocyte membrane systems (ABO system: A_1, A_2, B, O, A_1B, A_2B phenotypes; MN system: M, MN, N phenotypes; Rh system: C, C^w, c, D, E, e factors) and two serum protein systems (H_p system: H_p 1–1, 2–1 and 2–2 types; Gm system: Gm/1/factor).

Materials and Methods

Some years ago an extensive twin study was carried out jointly by ourselves and the Institute of Anthropology, Eötvös Loránd University, Budapest. Within this work the voices of both members of 117 like-sexed adult twin-pairs reading the same text of about one minute duration were recorded (by B. Lubi) on tape.

For zygosity discrimination the methods of speech recognition and speaker identification by machine had to be further developed [2]. All these methods consist of two steps: feature extraction followed by classification. In feature extraction appropriately selected details of the spoken text are characterized numerically. Pitch is one of such numerically characterizable features.

Before describing the traits examined in this study some general comments have to be offered. Feature extraction of speech characteristics is performed all over the world by computerized electronic signal processing. For our present study, however, characteristics that can be measured by autonomous devices of low intelligence and the measurement of which can readily be automated had to be chosen.

Here we give a list of characteristics along with their numerical parameters considered in our study.

An obvious physical trait is pitch (fundamental frequency). Since it is not an absolute characteristic feature of the speaking person but is modulated by the content of speech, word stress, phase of respiration, etc., the mean value of the pitch and its mathematical standard deviation were used as parameters. The degree of "signing", reflected by the standard deviation of the pitch, characterizes a person rather reliably.

The so-called formant frequencies are characteristics widely used in phonetics. Formants arise when tenable voiced sounds (all vocals and the consonants l, m and n) are formed. The air current leaving the lungs and interrupted by the vocal cords at the rate of the fundamental frequency passes the pharyngeal, oral and eventually the nasal cavities and the elements separating them and its spectrum shows characteristic maxima when entering the environmental air. The formant frequencies are defined by these maxima. In our study, each three formants of the vocals "a" and "e" were measured and averaged over their occurrences during the one minute reading.

The so-called plosive consonant transition times [3] were also investigated in the study. The phonation of voiceless burst consonants (e.g. "k") is always followed by a silent interval lasting for some 10 msec only thereafter can the subsequent sound be formed. Research aimed at the identification of anonymous phone callers revealed that this interval, the plosive consonant transition time, is characteristic of the individual. It is determined by the anatomical and physiological properties of the individual by which he or she restores the air pressure having dropped at bursting. For our study, investigation and measurement of the transition time following all word-initial "k" sounds were chosen. Their mean value was calculated for each individual.

Further five parameters are based on the spectral analysis of the so-called monochorus [4, 7]. The monochorus is the electronic addition of a large number of speach segments of the same length but of different content, of the same speaker. In our study, about 5000 segments, each of 50 msec duration, were copied one on top of the other, by a computer. The concept of monochorus was initially based on the subsequently confirmed hypothesis that in the monochorus accidental events (faltering of voice, malphonations, etc.) are mutually extinguished but the stable characteristics are reinforced. The monochorus was subjected to computerized spectral analysis. In the spectrum we detected two hardly distinct peaks, we named these double peaks. It was also observed that at frequencies lower than that of the double peak there is only one further peak while at higher frequencies than that of the double peak there are several. From these latter, only the peak adjacent to the double peak was taken into consideration. We found that the frequencies of the first peak, of both peaks of the double peak, of the peak of the envelope fitted over the double peak and of the third peak are fairly characteristic of the speaking individual.

Thus, a maximum of fourteen parameters were taken into consideration to characterize each individual. It has to be stressed that devices directly measuring these parameters are not available on the market. The most used device in speech analysis is the computer supplemented by an electronic gear capable of storing the speech to be analysed, of carrying out operations and mathematical transformations and of visualizing curves or diagrams derived from the analysed data. The desired operations and mathematical transformations have to be fed into the computer as programmes. Most of our investigations and programme developments were carried out in a computer type PDP 11/40 of Digital Equipment Co., USA [2].

No automatic methods are known which would yield all the parameters listed above. In other words, the computer sometimes stops "hesitantly" and one has to have the programme proceeded by making a decision. Most parameters were determined by such interactive operations. The success of our fully automatic procedure to determine the pitch and the plosive consonant transition time, developed at the Institute of Telecommunication Electronics, Polytechnical University, Budapest, and experience described in the literature give hope that appropriately constructed small devices with a microprocessor small as a heart may make larger computers dispensible, at least in zygosity determination of twins.

Since the measurements of the fundamental frequency and the explosive consonant transition could fully be automated, these parameters were determined for all the 177 twin-pairs participating in the study. The total of all fourteen parameters has been calculated up to now for 16 twin-pairs.

It has to be noted that in addition to these parameters put forward by the general practice of computer processing of speech and by our own experience gathered during this study, several other parameters could be investigated with equal chance for success. The number of parameters is the resultant of two opposed points of view: too many parameters lead to superfluous prolongation of the analytical work. On the other hand, too few parameters may make the result uncertain, by being exposed too much to intra-individual variability caused by incidental events like acute airway disease of the individual or other causes. One of our intentions was to determine the minimum number of parameters necessary for a safe zygosity discrimination.

Having obtained the parameters,we could turn to the classification of the twin-paiıs. For this purpose, discriminance analysis and learning algorhythm based on the nearest neighbour principle seemed most appropriate. The procedure needs a complicated mathematical apparatus but it is generally applied and has been described [5, 6]. Therefore, here we restrict ourselves to the most essential steps. Initially all twin-pairs are represented by a set of figures (in mathematical terms: a vector) representing intra-pair differences in terms of the differences of the two individual values of each parameter, e.g. the difference in fundamental frequencies, the difference of standard derivations of the fundamental frequencies, etc. Thereafter each member of this set is multiplied by a member of a set of so-called weighing figures. If the sum of the resulting set of figures is lower than a certain threshold value, this speaks for monozygosity while the opposite case for dizygosity. The elements of the weighing set and the threshold value are determined automatically by the procedure itself. To obtain this, for a sufficient number of twin-pairs not only the characteristic parameters but also the zygosity classification attained by any independent method have to be fed into the computer. From these data the computer so to say learns the weighing set and the threshold value necessary for the right decision, hence the term learning algorhythm.

Finally, the expenses of the acoustic method involving voice recording, computer feeding and calculation time are considerably lower than those of blood-grouping if the expenses of blood sampling and storing, the price of chemicals and devices are taken into consideration.

Results

Tables I and II illustrate the results obtained in 49 male and 68 female twin-pairs for a single parameter, the intrapair difference of the averages of the fundamental frequencies. Next to these data the result of zygosity

Table I

Intra-pair differences of average fundamental frequencies (ΔF) and zygosity diagnosis based on blood-group determination (Z) in 49 male twin-pairs

ΔF (Hertz)	Z	ΔF (Hertz)	Z
0.8	MZ	7.9	MZ
1.1	MZ	8	DZ
1.1	DZ	8.2	DZ
1.2	MZ	8.3	MZ
1.5	MZ	8.6	DZ
1.6	MZ	11	DZ
1.9	MZ	11.9	DZ
1.9	MZ	12	DZ
2	MZ	12	MZ
2.1	MZ	14.5	DZ
2.1	MZ	16.6	DZ
3.1	MZ	18	MZ
3.2	DZ	21.6	DZ
3.4	MZ	23.4	DZ
4.1	MZ	23.7	MZ
4.1	MZ	23.8	DZ
4.5	MZ	27.9	MZ
4.6	MZ	30.5	DZ
5.1	MZ	33.5	MZ
5.7	MZ	33.7	MZ
5.7	DZ	38.9	DZ
5.9	DZ	39.1	DZ
6	DZ	52.3	MZ
6.5	MZ	59	MZ
6.7	DZ		

Table II

Intra-pair differences of average fundamental frequencies (ΔF) and zygosity diagnosis based on blodd-group determination (Z) in 68 female twin-pairs

ΔF (Hertz)	Z	ΔF (Hertz)	Z
0	MZ	12.3	MZ
0	MZ	13	DZ
0	MZ	13.1	DZ
1.3	MZ	14.3	MZ
2.1	MZ	14.7	DZ
3.6	MZ	16.6	DZ
3.7	MZ	16.7	MZ
3.9	MZ	17.4	DZ
4.1	MZ	17.9	MZ
4.4	MZ	18	DZ
4.7	MZ	18.9	DZ
4.8	MZ	19.1	DZ
4.9	MZ	19.2	DZ
6.1	MZ	19.4	MZ
6.7	MZ	19.5	DZ
7.6	MZ	19.6	MZ
8.1	MZ	19.6	DZ
8.1	MZ	19.7	MZ
9.1	MZ	20.5	MZ
9.5	MZ	20.5	DZ
10	MZ	23.1	DZ
10	MZ	25	MZ
10.1	DZ	30.7	DZ
10.2	MZ	30.9	DZ
10.2	DZ	31.2	DZ
10.9	MZ	31.9	DZ
11.4	DZ	32.4	MZ
11.5	DZ	34.9	DZ
11.6	DZ	35.6	DZ
11.9	DZ	38.2	MZ
12	MZ	45.6	MZ
12	MZ	50	MZ
12	MZ	51.3	DZ
12.2	MZ	58.6	DZ

classification (monozygotic: MZ, dizygotic: DZ) obtained by previous bloodgrouping is indicated. In case of complete concordance of the two methods the designation MZ should be replaced by DZ at a certain point if the data are placed in the order of increasing intra-pair differences. Since this is not the case and only a trend like that can be seen, it may be concluded that this single parameter is not capable of complete discrimination.

If all 14 parameters are available as in the case of the twin-pairs whose results are illustrated in Table III, there is perfect fit between the zygosity diagnoses obtained independently by the phonetic and the bloodgrouping method. These data formed the input programme fed to the computer. One of the characteristics of the internal operation of this programme is that by a learning algorithm the parameters exert their effect through different weights. Thereby, the jumble of data, perfectly confusing for the human eye, becomes classifiable for the computer.

The next tasks in the development of the study are increasing the number of fully evaluated twin-pairs, from sixteen on one side, and decreasing the number of parameters from fourteen as well as further automatization of the determination of parameters on the other side. Further, there is hope that refined definitions of measurable parameters characteristic of the heritability of human voice may clarify various medical and biological problems.

TABLE III

Parameters of individuals of 16 twin-pairs investigated in terms of fourteen parameters.
a twin-pair. Parameter No. IX is

Serial number of twin-pairs	1	2	3	4	5	6
Fundamental frequency, expected value, I	243.9 227.3	180 198	227 196	222 235	229 179	177 187
Fundamental frequency, standard derivation, II	27.1 26.2	47 15	17 24	25 43	54 12	19 12
First formant of "a", expected value, III	747.5 767.0	989 929	703 814	754 787	772 707	797 736
Second formant of "a", expected value, IV	1495.0 1319.5	1339 1486	1476 1436	1514 1495	1467 1414	1363 1315
Third formant of "a", expected value, V	2002 2223	2217 2352	2179 2030	2002 2204	2007 2070	2115 2104
First formant of "e", expected value, VI	253,5 526.5	301 365	355 275	292 326	331 451	276 268
Second formant of "e", expected value, VII	2463.5 2099.5	2054 1901	2062 1983	2008 2040	1984 2193	2116 2106
Third formant of "e", expected value, VIII	2892 2626	2806 2778	2915 2623	2528 2611	2810 2580	2990 2758
Mean of transition time in "k", IX	28 21.7	24.0 21.3	25.0 29.7	29 37	38.0 28.3	26.7 23.3
First peak of monochorus, X	255 245	210 290	245 220	230 245	220 185	200 200
First peak of the double peak of monochorus, XI	420 440	325 470	485 400	445 465	385 335	415 390
Peak of the envelope of the double peak of monochorus, XII	445 475	355 480	510 440	480 480	430 390	515 510
Second peak of the double peak of monochorus, XIII	465 490	370 520	565 490	495 515	450 415	650 595
Third peak of monochorus, XIV	710 745	580 640	715 645	740 730	640 565	740 760
Sex	♀	♀	♀	♀	♀	♀
Zygosity, based on blood-group determinations	DZ	DZ	DZ	DZ	DZ	MZ

Upper and lower entries in a box represent parameters of one and the other member of measured in msec, the rest in Hertz

7	8	9	10	11	12	13	14	15	16
202	263	131	141	102	138	126	112	117	131
206	256	101	153	114	123	175	110	115	133
23	17	28	55	11.3	14.7	5.1	19	21	4.4
14	8.2	7	15	11.3	25.6	7.2	33	38	6.1
889	924	1001	720	715	702	728	1124	721	806
809	745	1079	1092	910	1079	734	741	1118	897
1368	1493	1495	1441	1573	1495	1482	1306	1307	1385
1306	1330	1313	1306	1365	1345	1475	1534	1391	1560
2120	2062	2002	2423	2483	2288	2327	2106	2190	2496
2301	2234	2015	2398	2483	2418	2392	2626	2112	2340
354	369	286	422	357	312	416	409	520	293
308	355	539	370	442	286	455	267	429	351
1951	1954	2047	2039	2353	2164	2210	1930	1904	1989
2032	2217	1904	2080	1995	1989	1976	2230	2392	2333
3015	2640	2632	2531	2912	2652	2762	2743	2912	2737
2772	2749	2528	2821	2730	2600	2522	2782	2525	2990
18.4	19.0	22.4	5.2	22.3	19.2	16	14	8.7	34.5
21.3	19.7	35.6	11.2	34.0	26.4	17	24	13.5	36.0
240	245	145	155	205	270	240	235	225	275
245	260	220	145	225	145	210	225	245	270
415	495	270	410	315	395	350	345	395	395
420	495	325	295	320	245	330	320	415	345
480	510	305	435	370	420	400	390	445	490
470	510	385	325	390	290	410	395	445	380
565	615	340	470	395	465	465	415	510	540
565	535	470	345	445	370	490	460	510	440
810	730	490	638	630	640	590	540	610	640
685	740	570	490	540	535	590	535	585	615
♀	♀	♂	♂	♂	♂	♂	♂	♂	♂
MZ	MZ	DZ	DZ	DZ	DZ	MZ	MZ	MZ	MZ

Acknowledgements

In the elaboration of some details of feature extraction and classification much help was furnished by several students of the Polytechnical University, Budapest, especially by Mr. Z. Kormány and Mr. S. Dankó.

References

1. Andrews HC: Introduction to mathematical techniques in pattern recognition, John Wiley, New York 1972, p 242
2. Dunn HK, White SP: Statistical measurements on conversational speech. J Acous Soc Am 11:278, 1940
3. Gordos G: Recent results in speech processing. Period Polytechn, 24:243, 1980
4. Gordos G: An interactive speech analysis system and its application to the diagnosis of twin-zygosity. Proc Symposium on Speech Acoustics, Budapest 1980, Pp 45
5. Nilsson NJ: Learning machines, McGraw Hill, New York 1965, p 137
6. Takács Gy: Personal communication
7. Tarnóczy T, Radnai J: Eine Möglichkeit automatischer Erkennung von Vokalen. Proc VII. International Congress on Acoustics, Vol 3, Budapest 1971. Pp 61—64

Received 9 February 1983

G Forrai MD
Alig u 5
H-1132 Budapest, Hungary