Analysis of Varying Sounds, Part II, Practical Applications of the New Photographic Method.

By

Katumi IMAHORI

Abstract

The strobophotographic principle available for the analysis of varying sounds as described in the author's previous paper,[1] is applied to the study of various actual sound waves. Explanation of the apparatus used together with an improved process of preparing the analysing plates is given. Examples of the results obtained are shown in the following cases:

1. Speech melody curves,
2. Separation of various parts in musical performances,
3. Some acoustical characteristics of vowels and consonants, and
4. Various types of noises.

§ 1. Introduction

In the previous paper[1], which will be cited as Part I in the following, the author has described a new photographic method of sound analysis, which is specially adapted to the case of varying sounds such as speech, music and noises, and which has the merit of giving the result of analysis simultaneously as the sounds are produced. A theoretical formulation has also been made of the fundamental methods upon which studies on varying sounds are to be based, the object being to remove the ambiguity which has been left unexplored by so many authors and has often led them to false arguments. Examples to illustrate the typical cases of varying sound analysis were given there by using model waves. The present paper

1) K. Imahori: Journ. Fac. Sci., Hokkaidô Imp. Univ., Ser. II, Vol. III, No. 3, 1940.

Journ. Fac. Sci., Hokkaidô Imp. Univ., Ser. II, Vol. III, No. 5, 1941.

deals with practical applications which have not been undertaken in Part I.

The procedures by which an analysis is made of varying sounds such as speech sounds or transient vibrations of musical instruments have hitherto required so much trouble that a systematic investigation in this field seemed almost hopeless. Thus, for example, Fletcher[2] and his collaborators calculated the period by period spectra of a short sentence "Joe took his father's shoe bench out" spoken in various manners, and obtained the time variation of fundamental pitch, the speech melody curve, as well as its harmonics. Fletcher remarks however that in order "to get a typical cross-section of American speech, it would require at least 100 such sentences pronounced by at least 5 men and 5 women," and "so such a job for analysing only the steady-state part of speech would require about 210,000 hours, or 100 years working seven hours a day for 300 days per year."

In his investigation on the transient vibrations of musical instruments and speech sounds, Backhaus[3] adopted a similar method of procedure, in which time variation in amplitudes of various harmonics was obtained by calculation of the Fourier coefficients for each successive period. As the exact lengths of one period were difficult to determine in the transient part of the recorded oscillograms, he assumed quite arbitrarily the same period as that of the stationary part. Although he recognizes the ambiguity produced by this arbitrary assumption, nothing is shown as to the limit to which one is able to make safe assertions in interpreting the results obtained.

In case only the pitch variation is concerned, one may conveniently make use of the direct-reading pitch recorder developed by J. Obata and R. Kobayashi[4] and also by Grützmacher and Lottermoser[5]. This is an electronic frequency-meter preceded by necessary wave form adjusters, and calculates automatically the number of waves in a definite time interval. The sound must then be composed only of a single pitch, so that it becomes impossible to apply it where

2) H. Fletcher: Rev. Mod. Phys., 3, 258, 1931.
3) H. Backhaus: ZS. f. Techn. Phys., 13, 31. 1932.
4) J. Obata and R. Koyabashi: Proc. Imp. Acad., Tokyo, 13, 247. 1937; J.A.S.A., 9, 156. 1937; 10, 147. 1938; Proc. Phys.-Math. Soc. Japan, 21, 109. 1939.
5) M. Grützmacher and Lottermoser; Akustische ZS.. 2. 242. 1937; 3, 193. 1938. 5, 1. 1940.

polyphonic sounds are concerned, for example, a solo with accompaniments or a duet and so on.

Recently A. Gemelli[6] has made use of this pitch recorder in an attempt to analyse speech sounds automatically. The time variation of pitch and intensity is recorded by this apparatus on the one hand, and the tone-frequency spectrometer of Freystedt is used on the other hand to estimate the amplitudes of overtones. Thus in this case a semi-automatic analysis is accomplished, but it is impossible to ask for a universal application of this method to other cases where for example the pitch indicator fails, or the resolving power of the tone-frequency spectrometer is insufficient.

The photographic method developed in this paper will be found to solve all the difficulties described above. It gives, automatically and simultaneously as the sounds are produced, the analysed components of any sounds, which appear in a photograph as black and white stripes at different positions according to their frequencies. The dispersion and the resolving power of frequency and also the time constant of analysis, i.e. the rapidity in which the analysis is made, may be varied at will by a single interchange of analysing plates. These constitute the essential features of the present method which have been difficult to realize in other methods hitherto employed. Detailed accounts in this connection have been published in Part I.

§ 2. The Apparatus

The arrangement of the parts used in the present experiment is shown diagrammatically in fig. 1. The output of a condenser microphone designed by Prof. Obata[7] is amplified by a preliminary amplifier, then filtered by variable electric wave filters[8]. This is again amplified by a secondary three stages resistance-capacity coupled amplifier. Finally the amplified electric current is used to excite a glow lamp G. L. which transforms it into light intensity variation. A

6) A. Gemelli: Arch. Psicol., Neurol., Psichiat. e Psicoterap., a. I., f. 1–2. 1939.
7) J. Obata and Y. Ozawa: Proc. Phys.-Math. Soc. Japan, 12, 285. 1930.
8) These are used in order to bring out the vibrations in the desired frequency range with a reasonable amplitude, by eliminating other components which may happen to have large amplitudes and give undesirable effect in the amplifying stages. Another function of wave filters in the present method is described later.

modulation controller, the circuit diagram of which is shown in fig. 2, is sometimes inserted before the glow lamp automatically to control its bias current, so that the light intensity variation is always kept at as large as possible effective percent of modulation regardless of any change in its amplitude. The principle of this modulation controller is the same as that which is used in the noiseless system of recording for talkie films.

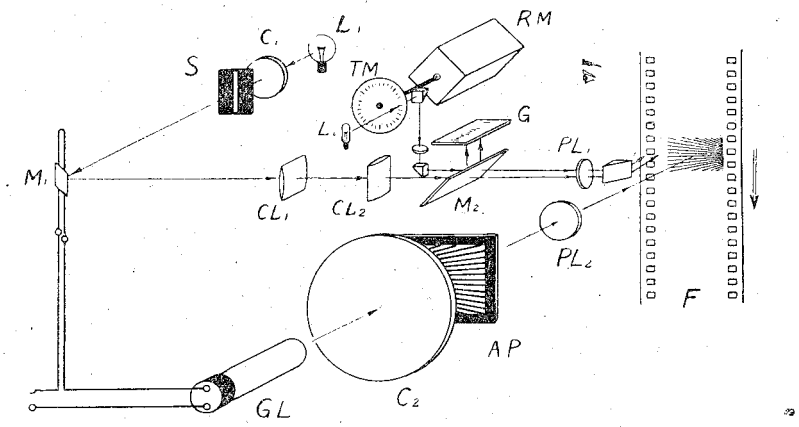


Fig. 1. Apparatus for Photographic Sound Analysis.

The electric current exciting the glow lamp is also fed into an oscillograph vibrator which is connected in series with it. The oscillograph used here is a commercial three elements portable oscillograph with a time marker. Only one of the available three elements is usually used, but in case of need the other two may be added. The light spot from the vibrator mirror M and that of time marks are focused on one side of a standard 35 mm. film F by means of a short focus photographic lens PL_1 of aperture f/1,8. A cylindrical lens CL_2 is inserted in order to form an image of the vibrating mirror M_1 at PL_1 so that the deflected light always passes through it irrespective of its amplitude. The recorded amplitude of vibration is of course reduced by this device, and at the same time the light intensity is increased. These are desirable features for making sharp reduced recording in a small space.

The principle of the analysing mechanism of the present method has been fully described in Part I. It is essentially the same as that

of a process by which an ordinary stroboscopic disc is used for the harmonic analysis of sound-frequency oscillation as described by T.d. Nemes[9]. In fig. 1, the analysing plate A. P. corresponds to the stroboscopic disc, and its optical image is formed by a photographic lense PL_2 on the moving film F. As compared to the stroboscopic method, two radically important changes are made here. While in the stroboscopic observation the disc moves with a constant angular velocity and the observer's eye remains stationary, the analysing plate stands still in this case, and the photographic film which may be compared to the eye moves relatively to the A. P. In both cases the

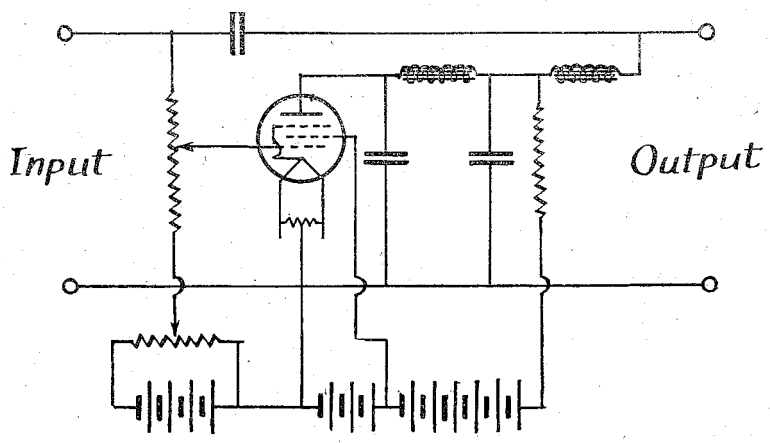


Fig. 2. Circuit Diagram of the Automatic Modulation Controller.

analysis is carried out by an integrating mechanism in the residual image of the eye or the photographic latent image. The superior characteristics brought about by this mere interchange of moving parts will be found in the case of varying sound analysis. The analysis may be made continuously both for frequency and for time, while in the usual stroboscopic method the analysed frequency is discontinuous. This has been the case with Seashore and other's pitch analysis and also with Nemes' harmonic analysis. Although a stroboscopic disc with a continuous frequency distribution may be designed, the analysis will then become discontinuous for time.

9) T.d. Nemes: Phil. Mag., 18, 303. 1934.

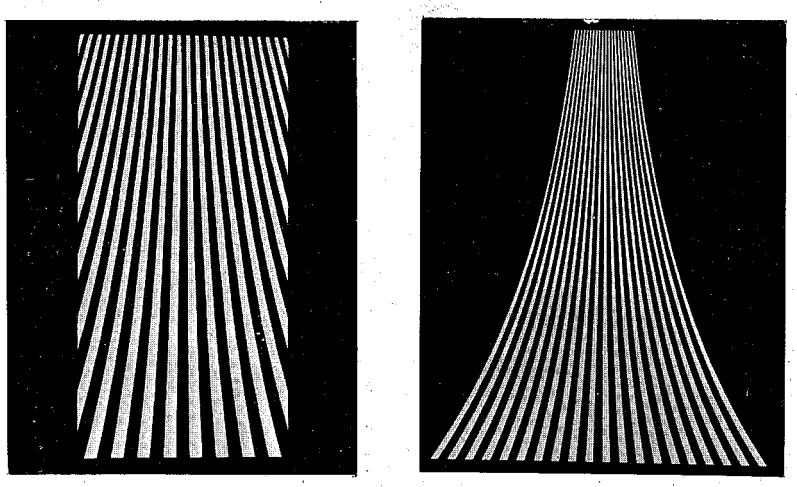
The second improvement made by the present method consists in the light transmitting characteristics of the analysing plate. The simple black and white stripes like those of a stroboscopic disc are used in special cases where, for example, the sound to be analysed contains only a limited range of frequencies extending not over 3 : 1 in frequency ratio. The reason is simply that if the black and white spacings are equal such a "square wave" will contain overtones of odd harmonics and "resonates" to the third, the fifth, etc. harmonics in addition to the fundamental. For general purposes therefore the light transmitting characteristics must be sinusoidal instead of square shaped. The analysing plates with this property are shown in Plate XXIII, where figs. 1–5 are those which are intended to cover five octaves with the same time constant, each covering a frequency range 3:1 in ratio. If a wider frequency range is required, one may attain that object by changing the film speed. Of course a change in the time constant is accompanied in this case, and so necessary precautions must be taken. By combining two, three, four or five of them, plates are also prepared which cover in a single analysis frequency ranges of 2–5 octaves respectively, the one with 5 octaves range being shown in fig. 6. As these frequency ranges should be contained in the limited width of 35 mm. film, the general dispersion of the spectra obtained diminishes inversely proportional to the range of frequencies involved. Analysing plates with the resolving power constant for all frequencies may be obtained simply by covering those plates described above with a black paper so that the same number of "waves" is contained for different "wave-lengths," an example of which was shown in Part I.

Definite drawbacks in these analysing plates are that the dispersion of frequency becomes smaller for higher frequency in each of the analysing plates 1–5, and that discontinuity occurs in the combined analysing plates. The frequency dispersion δ_ν may be defined as $\nu dx/d\nu$, where x is the distance measured across the film and ν the frequency. This is inversely proportional to ν for plates 1–5, while in acoustics it is more desirable to have this value constant for all frequencies. In the latter case x is related to ν by a logarithmic function,

$$x = \frac{1}{a} \log \nu/\nu_0 ,$$

$$\text{or} \quad \nu = \nu_0 \, e^{ax} . \qquad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

The realization of an analysing plate with this property encounters many technical difficulties. But if the sinusoidal light transmitting characteristics of the analysing plate are abandoned, the above requirement for the dispersion may be satisfied simply by adopting the usual black and white, square-shaped transmission characteristics.



a b

Fig. 3. Analysing Plates with Logarithmic Frequency Scale and Square Transmission Characteristics.

Examples are shown in fig. 3, a and b, which were obtained by making a drawing in sufficiently large scale on a section paper and then taking its photograph. These analysing plates are, if used in conjunction with properly selected electric wave filters, as useful as those described above, and even the harmonics contained in the square shaped waves may sometimes be utilized to extend the frequency ranges by factors of 3, 5 or 7 provided that the sound wave to be analysed has a simple structure so that the intermixed spectra are easily interpretable.

As to the relation between the resolving power and the time constant of analysis, a full discussion has been presented in Part I, so that further description shall be omitted here.

For the film driving apparatus, a cinematograph film camera sold as an accessory to the oscillograph unit was employed in the earlier stage of the experiment. The constancy of the film speed was

not so good, mainly due to disturbances occurring irregularly in the frictional force. By loading the driving axis of rotation with a fly-wheel of large mass, they were eliminated altogether. But a small velocity fluctuation of frequency 20–30 per second still occurred occasionally, probably due to oscillations of the film excited by small irregularities in the spacings of perforations. Improvements in this connection are now being made by employing the mechanism used in talkie recording. Most of the photographs shown as examples of applications in the following are those taken by the apparatus explained above. In case the sound to be analysed is of short duration, a usual revolving camera may conveniently be used, where constancy of film speed is easily secured.

§ 3. Preparation of the Analysing Plates

A procedure by which the analysing plates are prepared has been described in Part I. Various other methods have been tried since, and the one which was found to give the most satisfactory result shall be described here.

The characteristic curve of the photographic emulsion which is to be used as the analysing plate is determined for a certain condition of development which is kept constant for all the following procedures. This is used to draw a curve which determines the light quantity to be exposed on the emulsion, such that on development sinusoidal variation in transparency with approximately 50% mean transmission is obtained. Such exposure curves are drawn on a section paper in black and white silhouettes one for each octave as shown in fig. 4, and then copies are made by photographing them. In fig. 5 P_1 is the photographic plate prepared in this manner, on which is formed a vertical image of the filament of a lamp E by means of a photographic lens L_1. A cylindrical lens L_2 is used to form on the plane of a wedge form slit S vertical stripes of light, the intensity variation of which is determined by the curve in P_1. The diverging rays of light from P_1 are in the horizontal section focused on the plane of S, while in the vertical section they cover the entire length of S, so that those from different points on P_1 but in the same vertical line are added in their intensity, and result in a total intensity proportional to the vertical length of the transparent portion in P_1. The plate P_1 and the rod R are fixed together and move with constant

speed as shown by arrows in the fig. The rod R pushes the bar B which is fixed to the holder of an unexposed photographic plate P_2 and gives rotation to the latter. The mechanism used in this device is the same as that previously described in Part I. A correction for

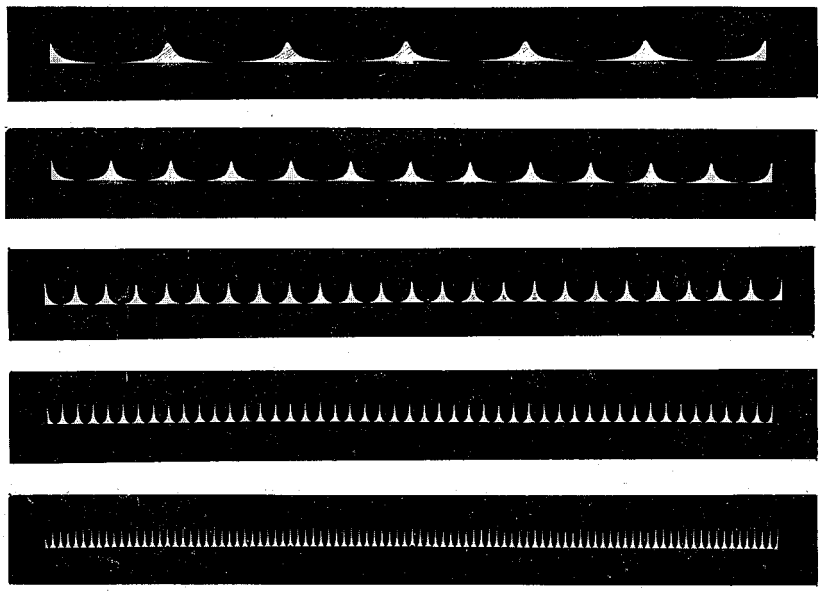


Fig. 4. Exposure Curves for Analysing Plates (positive).

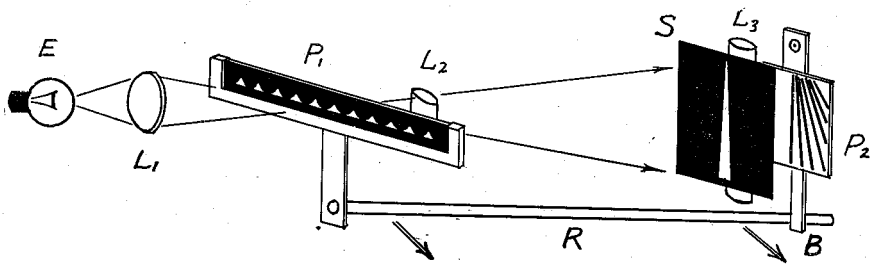


Fig. 5. Preparation of the Analysing Plates.

the light intensity to compensate the difference in exposure due to unequal rotation velocity which slows down as the inclination of the bar B from its vertical position increases, is made by another photographic plate which has proper density variation and is placed in contact with P_1.

In practice the analysing plates which have desired characteristics are obtained by trial and error method. The final test is made by taking microphotometric record and calculating the harmonic components contained in the curves by the usual method.

For many purposes it is frequently found useful to prepare a negative such that by printing on another photographic material the analysing plate is obtained which has the desired characteristics. For this purpose curves are determined which show the relation between the transparency of the positive and the exposure given to the negative for several cases of exposure rates in printing. From

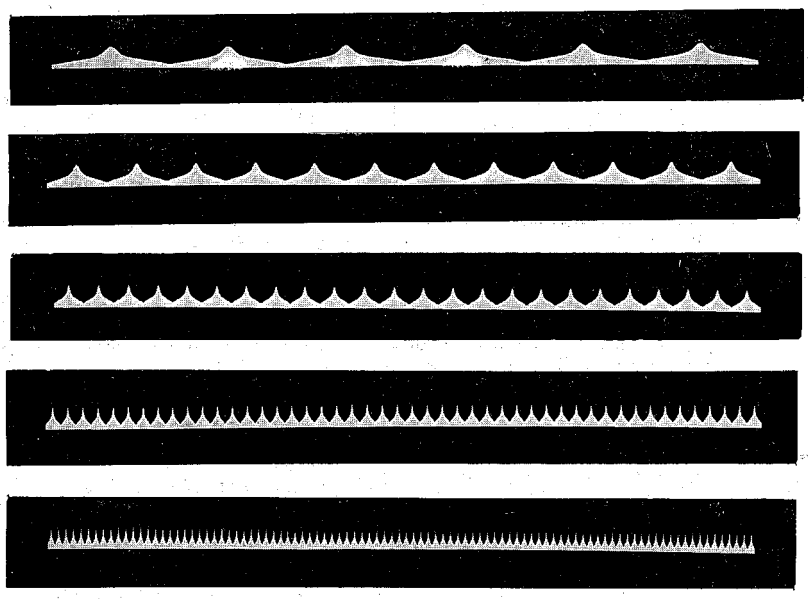


Fig. 6. Exposure Curves for Analysing Plates (negative).

these curves one is chosen which covers as wide a range as possible without unreasonable deviation from linearity, and then exposure curves for the negative corresponding to a sinusoidal change in transparency of the positive are drawn. Such exposure curves are shown in fig. 6, and the remaining procedure is the same as that described before.

In fig. 7 are shown photometric curves of analysing plates made by the above procedure for various degrees of deviation from sinusoidal transmission characteristics. Curves 1–2 and 4–5 indicate

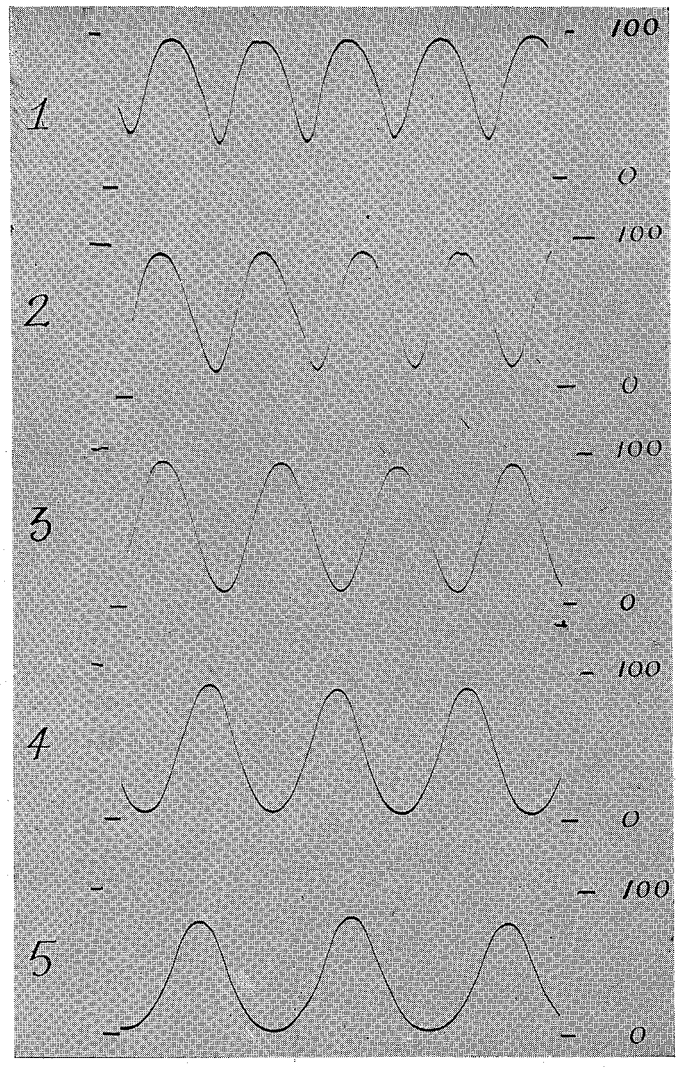


Fig. 7. Photometric Curves of Analysing Plates.

under- and over-exposures respectively. The levels marked as 0 and 100 correspond to both extremities of the light transmission. From these curves the harmonic components were calculated as follows,

Harm. Comp.	Proportionate Amplitudes for Curves				
	1	2	3	4	5
0	68.9	58.2	50.5	42.2	35.6
1	31.7	38.7	41.8	41.5	38.2
2	6.65	3.77	0.35	4.23	6.56
3	1.37	0.46	0.40	0.51	0.54
4	0.48	0.65	0.02	0.16	0.57
5	0.20	0.23	0.07	0.31	0.78
6	0.41	0.20	0.08	0.20	0.22
7	0.12	0.17	0.33	0.24	0.09
8	0.08	0.09	0.14	0.07	0.01
9	0.14	0.06	0.04	0.05	0.03
10	0.09	0.04	0.06	0.11	0.08
11	0.13	0.41	0.05	0.29	0.54
12	0.02	0.01	0.05	0.01	0.10
Mod. %	46.0	66.5	82.7	98.2	107.0

Total lengths of the curves excluding fractions of one period at both the ends were used in this calculation to obtain the n-, 2n, ..., and 12n-th harmonic components where n is the number of periods contained in the curves, so that the above values of proportionate amplitudes indicate the mean values of those for each period. The modulation percentage given in the last line is defined as the amplitude of light transmission for the fundamental component divided by its mean.

§ 4. Speech Melody Curves

The time variation of pitch in speech sounds or the so-called speech melody curve has been taken up by many physicists, psychologists and phoneticians as their subject of investigation. The characteristic types of pitch variation in speech for different nations, sexes, personalities and emotions etc., may only be studied by an objective method, for the ear is almost unable to estimate the rapidly and continuously varying pitch. The recent development of the direct-reading pitch recorder seems to contribute much in this field and in some respects it has unrivaled features for this purpose. A definite

drawback in this useful apparatus, however, is found in that it is sometimes imposisble to find the corresponding location in the curve for each spoken syllable, especially when voiced consonants and vowels appear successively, such as "Ame ga yamu"—"the rain stops." It is necessary to add some device to make these discriminations possible, and a definite improvement in this respect has been achieved in the present investigation. One is able to obtain rapidly varying spectra of speech sounds in which are shown changes in vocal characters in each syllable as well as changes in pitch and intensity.

Examples of the photographic sound analysis applied to this problem are shown in Plates XXIV, XXV and XXVI. The sounds were analysed of phonographic records which are sold for educational purpose. The analysed frequency regions are 115–345 cycles/sec. for male, and 200–600 or 260–780 cycles/sec. for female voices, so that only the pitch and intensity variations are shown in the photographs, the analysis of speech sounds for wider frequency region being given in a later section. The broad dark and bright bands in the photographs are due to rise and fall in the amplitude of vibrations which, by means of the automatic control of the bias current in the glow lamp, cause variations in the mean light intensity. Thus silent portions appear sometimes almost completely dark due to an increase in contrast by photographic printing. In the bright portions there may be observed striped bands appearing up and down as the melody rises and falls.

In Plate XXIV are shown three types of word pitch patterns in Japanese;

"Ha'na ni",
"Hana' ni", and
"Hana ni".

Dashes are used to indicate high pitched syllables. In Japanese, words written with the same letters mean different things when they are spoken in different pitch patterns. Japanese accent is a pitch accent which may be compared to the stress accent in English. Thus "Ha'na" means "the beginning", "Hana'" "a flower" and "Hana" "the nose". Often the pronunciations are entirely reversed in different districts, especially in Kantô and Kansai. These are characteristic features in spoken Japanese and have offered various pro-

blems to Japanese linguists. It is generally accepted that these word pitch patterns may be formulated by assuming two or three pitch levels such as high and low, or high, medium and low. The pitch patterns are characterised by various combinations of these levels, so that for the above examples,

"Ha'na", high-medium;
"Hana'", low-high ; and
"Hana", low-medium ;

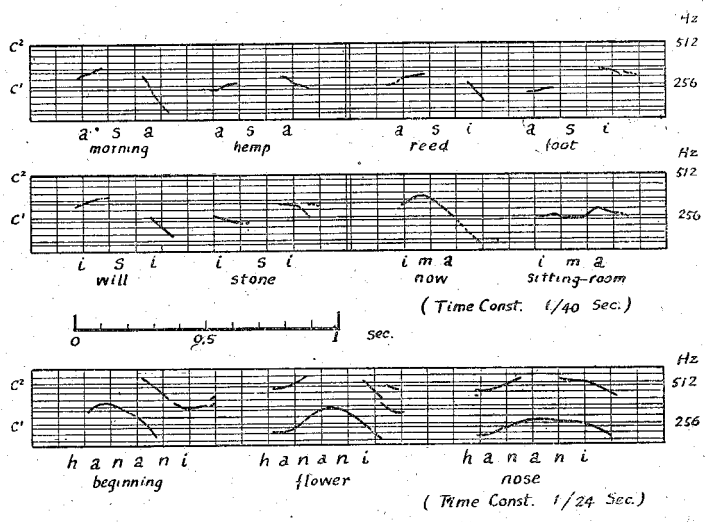


Fig. 8. Japanese Word Pitch Patterns.

are the types formulated by three-levels theory of Dr. Sakuma. The inspection of the pitch curves shown in fig. 8, which were obtained by measuring on the original film, reveals however that the correspondence between the theory and the actual phenomena is not so good. The curves are sometimes very smooth and have no stepwise change corresponding to a transition between diffeernt levels except for some fortunate cases. Judgment of pitch by the ears seems to be very flexible in cases of rapidly varying sounds such as spoken words, and it is easily affected by various surrounding conditions. Variation in amplitude, vocality and even letters[10] used to express

10) In Japanese letters, vowels preceded by a consonant are expressed by single letters, thus ha: ハ, na: ナ.

the word may take part in it, so that care must be taken not to draw any hasty conclusion in these cases.

In plates XXV and XXVI are shown examples of conversational speech by male and female subjects. They are

Plate XXV, "Kore desu ka?": "Is it this?"
"Ē sō desu." : "Yes, it is."

Plate XXVI, "Minasan otassya desu ka?":
"Are you all well?"

Two different patterns of the introgative "ka?" may be noted, the one showing a structure of low-high type and the other a more gradual and flatter one.

§ 5. Separation of Parts in Musical Performances

As an application of his direct-reading pitch and intensity recorder Prof. Obata[11] has recently reported a series of interesting works on musical performances of famous vocalists. A serious limitation seems however to prevent further development in this promising field of study. The instrument is applicable only to solo parts or those in which the accompaniments are not prominent. The relation of pitch variation in the solo part to its accompaniments, or mutual influences upon intonations in a duet etc. are some examples of the problems which should be included in the programme. Not only in vocal music, but also in instrumental there are various problems to be investigated. The tool needed for these purposes is a varying sound analyser of proper design, without which the ear has been the only measure for assessing musical performances, and so various dogmatic arguments have been prevailing in the musical circle. By laying the foundation on objective observations of phenomena, one may construct a "science" of musical performances, and it is the object here to show a possibility of throwing some light on this problem.

In Plate XXVII is shown a part of the photographic analysis made on a duet—La Travita—Un di felice, Columbia record—in which the soprano, Capsir and the tenor, Cecil are singing different melodies.

11) J. Obata and R. Kobayashi: J.A.S.A., 12, 188. 1940; Proc. Phys.-Math. Soc. Japan, 22, 691. 1940; 23, 239. 1941.

The frequency region is from 450 to 1350 cycles/sec., so that only the fundamental tone of the soprano and the first overtone of the tenor appear in the photograph. Owing to vibrato effect the wavy pitch variations are separated and entangled with one another in a complex manner. The result for a wider frequency region is shown by a graph in fig. 9, where the melodies are compared to the corresponding musical notes. The fundamental of the tenor voice c^1 hardly appeared on the lower frequency side of the original film and

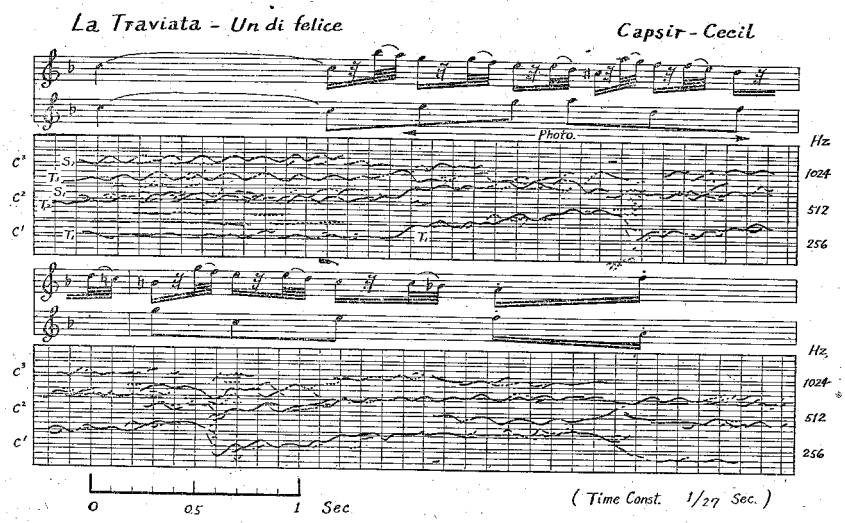


Fig. 9. Separation of Melodies in a Duet.

its first, second and third overtones were obtained. For the soprano only the first overtone beside its fundamental was measured. They are marked as T_1, T_2 and T_3 for the tenor, and S_1 and S_2 for the soprano respectively, and are shown in the figure as wavy curves which may be seen to have different forms of vibrato cycle for both voices. The portion where the original photograph was reproduced in Plate XXVII is marked "Photo." Some traces of orchestra accompaniment which is played just after the melody has begun are also found.

Another example is shown by a similar graph, fig. 10. This is a solo in Rigoletto—Caro nome sung by Capsir with orchestra accompaniment. The shaded portions are the predominant tones contained in the accompaniment, and owing to the lack of resolving power for

lower frequenices, they appear as broad bands extending over wide frequency regions. The melody also has its proper width in the original photograph, but in the figure it is shown by a curve the frequency of which corresponds to the centre of band.

Further examples are taken from a piece of Japanese music, Naga-uta. This is one of the most popular musical forms in Japan

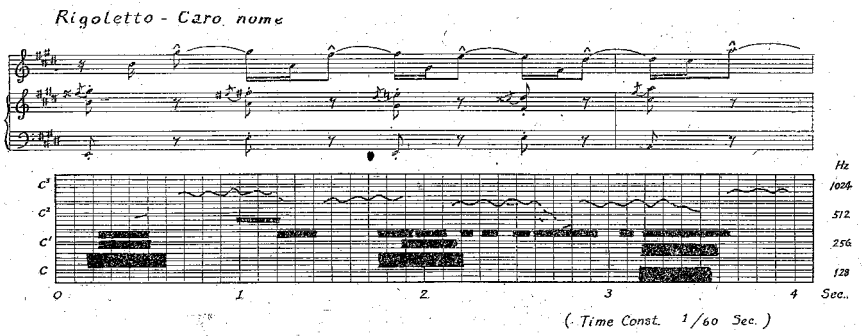


Fig. 10. A Vocal Solo with Orchestra Accompaniment.

widely spread among conservative women of all classes. A vocal solo with Syamisen[12] accompaniment is the usual manner of performance, but other instruments such as Taiko[13] or Fue[14] are also added. The melody played by Syamisen may easily be understood by ears trained in western music and is frequently transcribed into this apart from the solo part. The solo part has, however, much freedom in its execution and is differently sung by different artists. Sometimes it is very difficult to grasp the melody with the purpose to express it by musical notes.

A small part of photographic analysis made on the Etigojisi, a well known Naga-uta, sung by two famous virtuosos, Kosaburo Yosizumi and Wahû Matunaga is shown in Plates XXVIII and XXIX. In Plate XXVIII is shown the characteristic mode of Matunaga's performance which, in spite of the presence of very strong sounds of Taiko, is clearly separable. The voice starts with a very low pitch, then rises gradually into a steady note and lasts with remarkable

12) A stringed instrument. See J. Obata and Y. Ozawa: Proc. Phys.-Math. Soc. Japan, 13, 1. 1931.
13) A sort of drum.
14) A sort of flute.

constancy in pitch. On the contrary, Yosizumi's performance is characterised by a rapid and complex variation in pitch as shown in Plate XXIX, so that the melody is hardly expressible in a western form of musical notation.

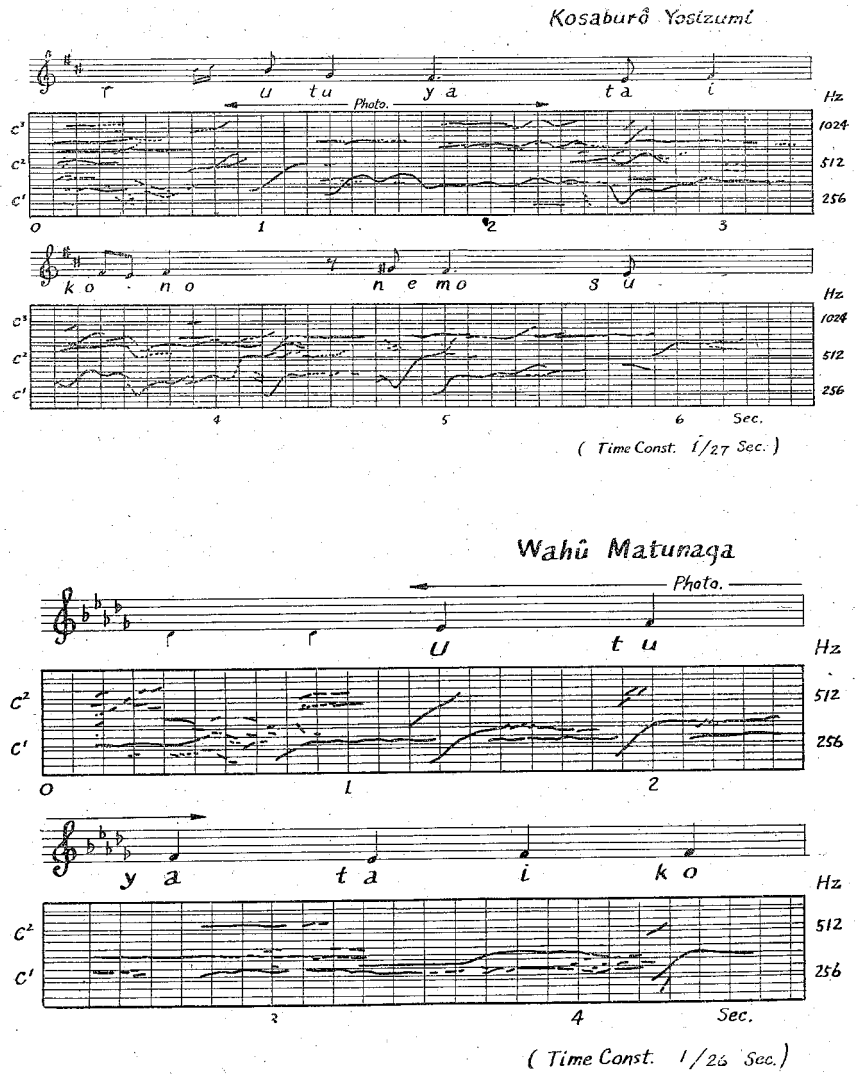


Fig. 11. Comparison of Performances in Naga-uta, a Japanese Vocal Solo.

For clearer understanding of this difference in performances the melody curves are compared in fig. 11, together with musical notes ascribed to them by the author. The expression for Yosizumi's melody is however an arbitrary one, which was forcibly and provisionally decided by listening again and again to its phonographic reproduction.

§ 6. Acoustical Characteristics of Vowels and Consonants

Extensive studies have been made by various authorities on the acoustical characteristics of vowels of different nations, and the theory on the nature of vowels may be recognized to have been established in a definite form, which, originating from Helmholts's idea, is usually called by the name formant- or resonance-theory. The fact, however, that the experimental materials used in these investigations were chosen samples specially prepared for study and so more or less different from living speech, proves that the real characteristics of vowels actually spoken by a particular nation or a particular person are not yet clearly known. The formant of a vowel is apt to vary in accordance with various circumstances, and so the statistical method must be employed for a complete solution of the problem. One may immediately recognise that the photographic method of sound analysis will be most conveniently applied for this purpose with least labour possible, in which is only included a procedure just like that of talkie recording. Not only the vowels—voiced or unvoiced—but at the same time consonants of all sorts may be analysed by this method. A systematic investigation will be included in the author's future works, so that a brief description of only a few examples relating to these topics shall be made in the following.

The analysing plate used in the following examples is that shown in fig. 6, Plate XXIII, which includes five octaves from about 200 to 6400 cycles/sec. In Plate XXX is shown an analysis of "Oh" spoken in decreasing pitch. As the pitch decreases and its overtones cross over the characteristic frequency regions, the intensities are enhanced, showing the resonance positions which one may easily find in the photograph at frequencies of about 850 and 600 cycles/sec. The same principle may be applied to other vowels[15] and will be found best

15) Recently D. Lewis and T. Tuthill have studied by a similar principle the resonance frequencies and damping constants involved in the production of sustained vowells "O" and "Ah" in which the vowels are produced with vibrato effect. D. Lewis and T. Tuthill: J.A.S.A., 11, 451. 1940.

suited to determine formants in lower frequency regions. For example the lower formant of "ii" spoken by the author was determined to be at 300 cycles/sec. This is shown by the Plate XXXI in which the faint higher formant will perhaps be lost in the reproduction.

Another example relating to the vocal resonance is shown in Plate XXXII which is a part of an analysis made on a continuous production of five Japanese vowels in the order "i-e-a-o-u". The sifting of formant is perfectly continuous and it is impossible to draw a line between "i" and "e" or "e" and "a" and so on. The photograph

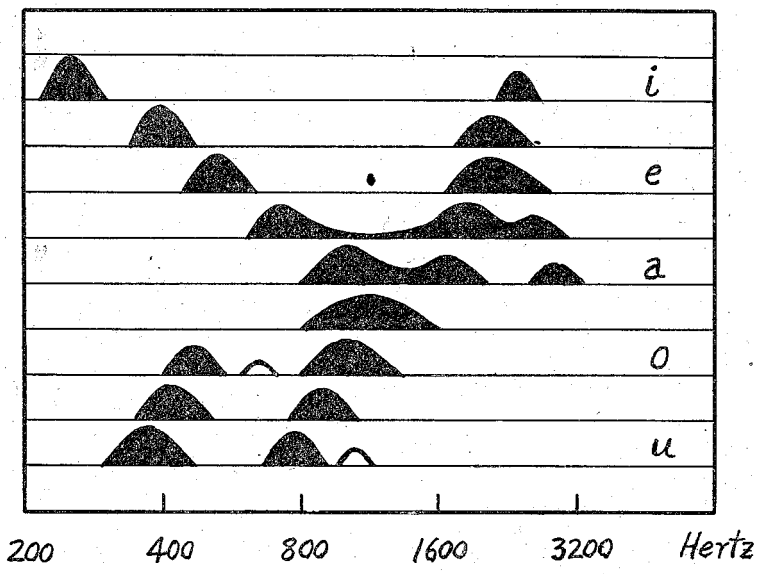


Fig. 12. Continuous Shift of Vowel Formants.

reproduced here is that part where transition between "e" and "a" occurs. The higher formant of "e" which extends over 1600–3000 cycles/sec. splits up into two, of which the one at the higher frequency side goes over into "a" with practically the same frequency, while the one at the lower frequency side goes down to the 1000–1500 band of "a". The lower formant of "e" at 400–500 cycles/sec. changes into the most prominent formant of "a" at about 900 cycles/sec. A clearer view of these transitions may be obtained by fig. 12 in which

is shown a sketch drawn by inspection of the original photograph to indicate a rough estimation of the resonance intensities. The shaded portions indicate those formants which were observed in the photograph described here, and the unshaded formants shown in the "o" and "u" spectra are those obtained by the preceding method of decreasing pitch voice production and not observed in the former. Further systematic inquiry into variants of vowel formants in actual spoken Japanese and their relation to the above formulation would require a series of independent investigations.

As to the nature of consonants, the investigations made up to the present by various authors have been insufficient for establishing a reasonable theory. The reason is two fold, the one, the lack of suitable tools, and the other, that the basic principles governing the perception of varying sounds are as yet unknown. Some clues to this problem may be obtained by the present method as the following brief description would show.

The characteristics of consonants may be described by two properties. The first is the usual frequency spectrum the structure of which discriminates between different consonants such as "s" and "sh", so that a similar idea like that used in the case of vowels may be applied also. The second property is found in the transient phenomena in the rapidly varying vibrations. Semi-vowels such as "y" and "w" in "ya" and "wa" are marked examples of this. Plate XXXIII shows that "ya" is a rapidly spoken "i-e-a—" in which "i" and "e" last about 1/10 sec. respectively. Similarly "wa" is a contraction of "u-o-a—".

Explosive consonants such as "p", "t" and "k" seem to be characterised by this property of transient state, for the frequency spectra of them are greatly influenced by the subsequent vowels, and entirely different for the same, for example, "k" in"ka" and in "ki", the difference being of the same nature as that between "a" and "i". In Plate XXXIV are compared the spectra of "pa" and "ta" spoken in a whisper. The resonant frequencies of unvoiced "a" appear most conspicuously in the region of 1000–1600 cycles/sec. At the beginning of sound vibration these resonant vibrations are set up abruptly in the case of "ta", while they grow up somewhat gradually in "pa". "Ka" was found more like "ta" than "pa", but seemed somewhat different in its mode of settting up. Of course a more detailed study

is necessary for any definite conclusion to be made, but the object here is to show how the present method might be utilised to solve the problem of consonants.

§ 7. Noises

Most of the investigations on noise have hitherto dealt principally with the problem of its measurement, its reduction or its psychological influences. The acoustical characteristics of a particular kind of noise, the tone colour in relation to its mode of vibration on the one hand, and the mechanism of noise production relating to the mechanical structure of its source on the other, have been as yet far from any systematisation. Here also is a wide field of exploration to which the present method may be successfully applied.

Noises may be classified roughly into four classes which have the following idealised properties:

1) Anharmonic sound with sufficiently long duration. The tone colour is determined by its spectrum which has line structure. In some cases where beats occur between adjacent frequency components, the beat frequency must be sufficiently high, so that uniform impressions are received by the ear.
2) Fricative sound. The tone impression is of uniform duration and its spectum continuous. Fricative consonants, thermal noise produced by Brownian motion in a resistance wire or the so-called carbon noise of a microphone are representative examples. Unvoiced vowels may also be classified into this type.
3) Explosive sound. Sounds produced by a percussion or other instantaneous shocks. The amplitude decreases rapidly and its tone character is effected by its mean frequency and the mode of grow and decay in vibration.
4) Oscillating sound. The amplitude or the frequency is altered periodically with sujcient rapidity, such as the sound of an electric bell or rapid trillers executed on a musical instrument.

In the above description of the classification the words fast or slow, gradual or abrupt in expressing the rapidity of change of sound are used in their comparison to the time difference by which the

ear can discriminate sounds coming in succession. Similarly the words continuous or discontinuous used in describing structure of a spectrum are related to the resolving power of the ear for frequency difference. In actual cases the sounds may occur in a more complicated manner than those described above, and may belong at the same time to two or more classes in the above classification. It is interesting to note that the consonants may be classified in a similar manner, and that various onomatopoeias so frequently used in Japanese contain one or more corresponding consonants in them. The vowels also take part in these expressions to indicate the approximate pitch corresponding to formants of the vowels used. Thus the same method of study may be applied to the case of noises as that described for vowels and consonants in the preceding section.

Examples showing the results obtained by the present method are given in Plates XXXV and XXXVI. The click of a metronome was analysed and shown in Plate XXXV, fig. 1. This is a sound with very short duration and its spectrum consists of a series of broad bands. Fig. 2 is another example of explosive sound which was produced by tapping a wooden drawing-board. Partial tones of various degrees of duration may be observed. It will be noticed in both examples that some of the partials seem to vary in frequency with time. This effect occurs in two cases where the one is a case of true frequency change and the other is due to the existence of two or more adjacent components which differ in their time constants. When sufficient resolving power is not used to separate them, it occurs apparently that the frequency shifts from the component with smaller time constant to that with the larger one.

In Plate XXXVI an interesting example of oscillating sound is given. This is the sound of a school whistle. The frequency is oscillating up and down rapidly in an irregular manner. The explanation is simply as follows: a small piece of cork contained in the hollow space of the whistle is set into irregular motion so that the narrow opening of the hollow is partly closed from time to time and so the reasonating frequency of the hollow space is altered.

§ 8. Conclusion and Acknowledgments.

The examples given in the preceding sections are merely examples, not intended to draw any definite conclusion concerning a special

subject, for each one of them will require a systematic investigation which will constitute a large part of the author's future programme.

In concluding this paper the author wishes to express his hearty thanks to Prof. T. Hori under whose kind guidance and incessant encouragement the present study has been carried out. The cost of this research was defrayed from the Scientific Research Expenditure of the Department of Education. Also a part of it was met by the Hattori Hôkôkai. The author wishes to express his sincere thanks to the authorities concerned.

Analysing Plates.

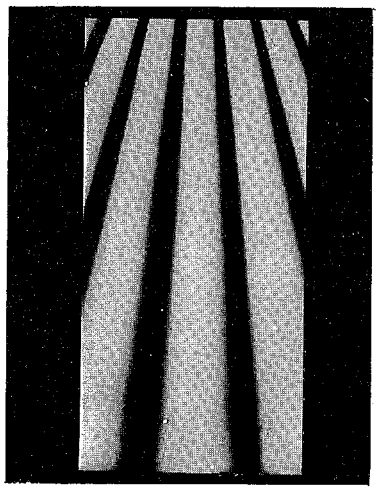


Fig. 1.

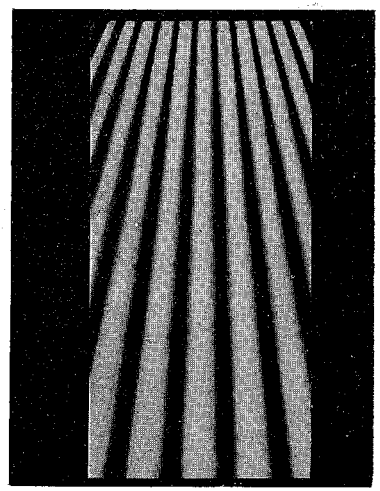


Fig. 2.

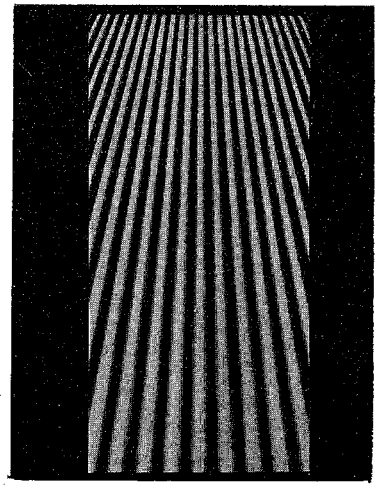
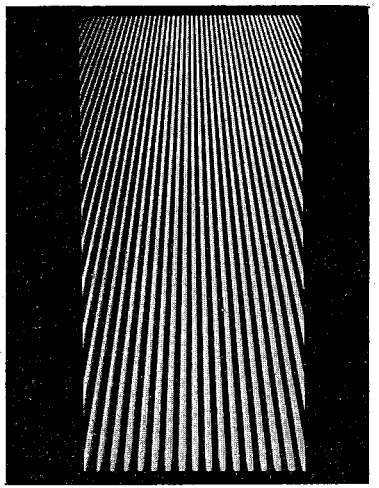
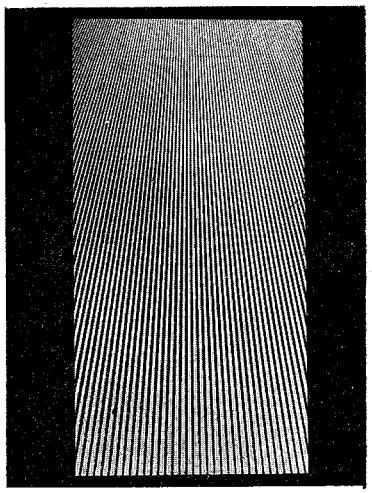
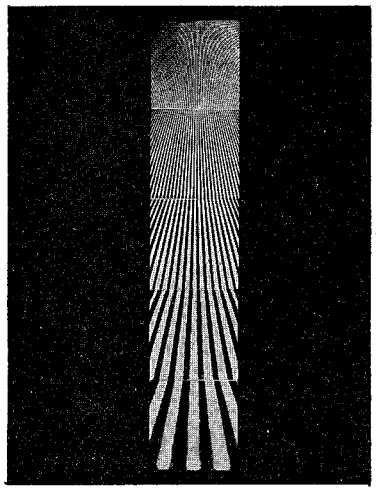


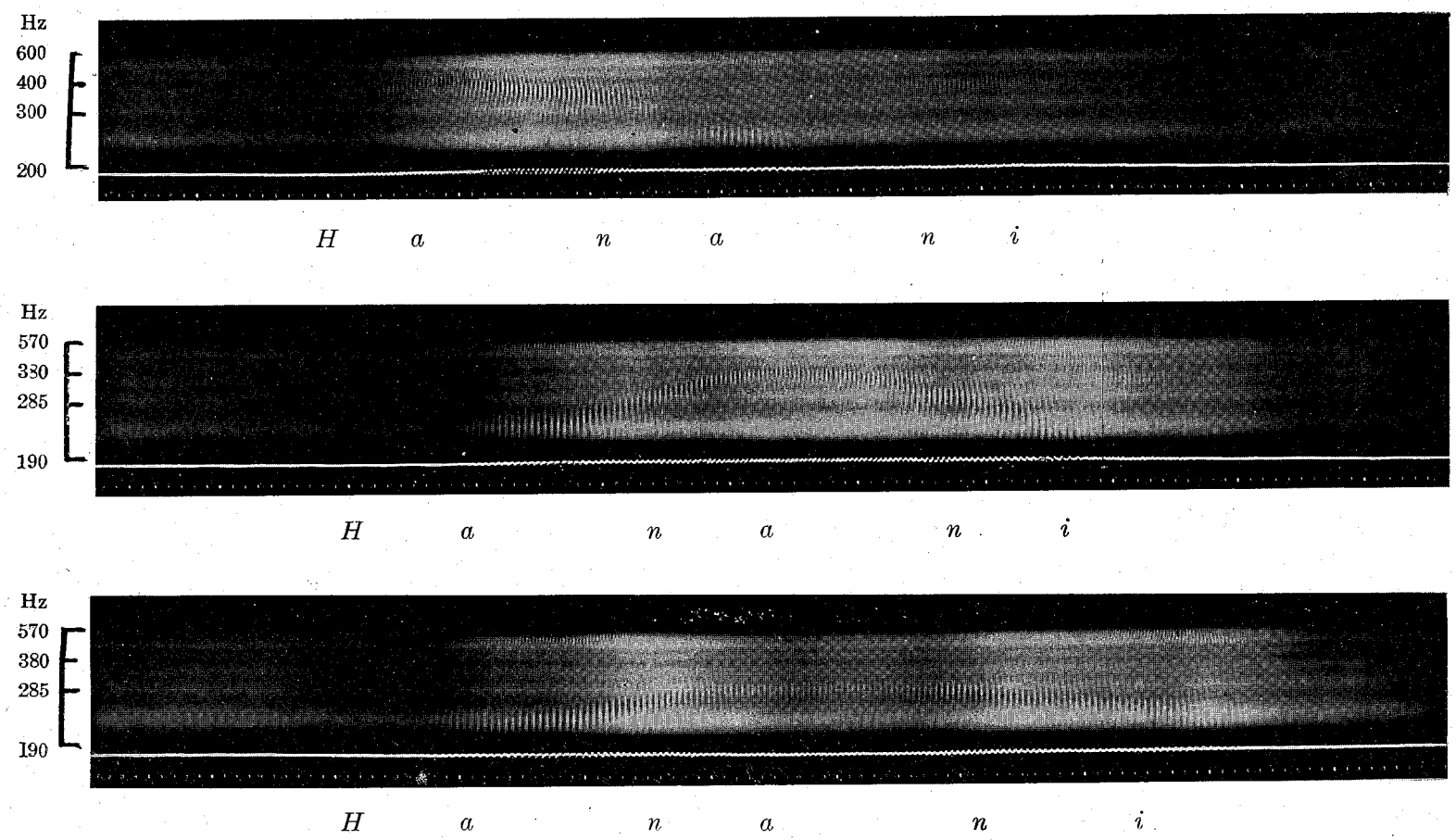
Fig. 3.



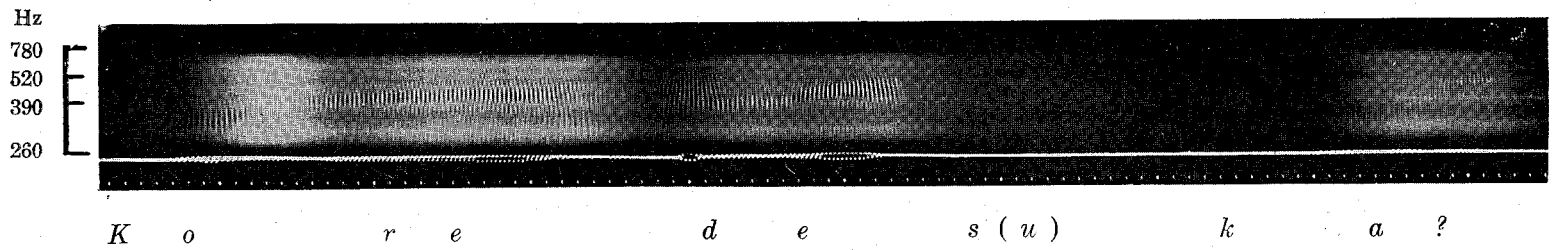


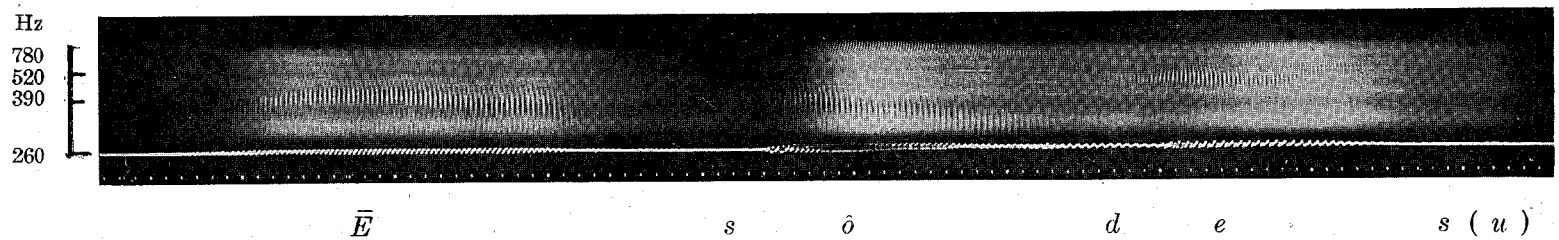


Word Pitch Patterns.

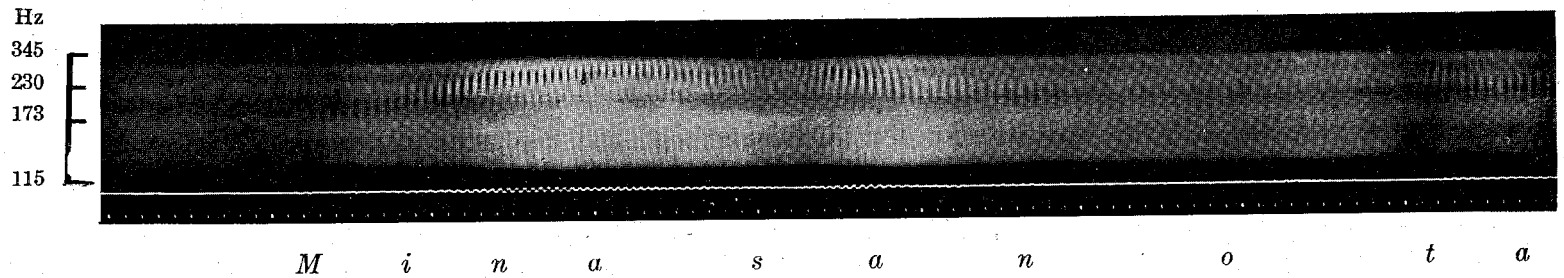


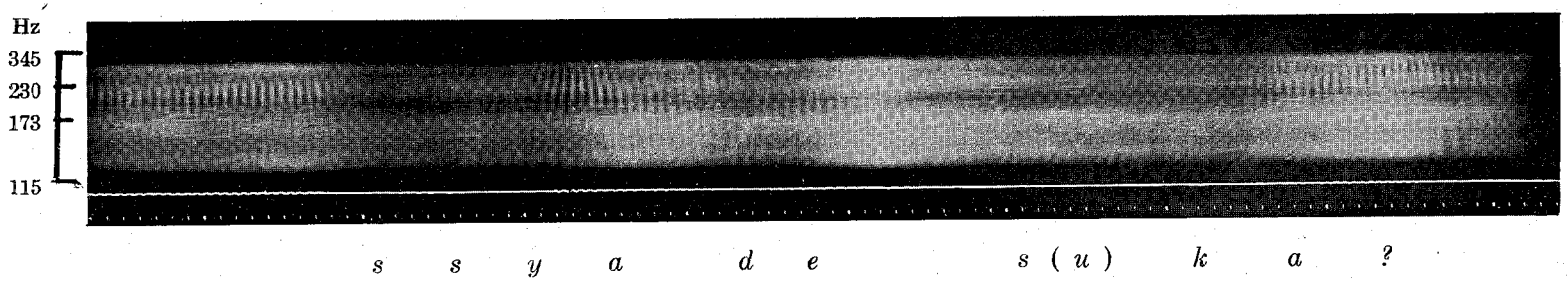
Conversational Speech. (Females)



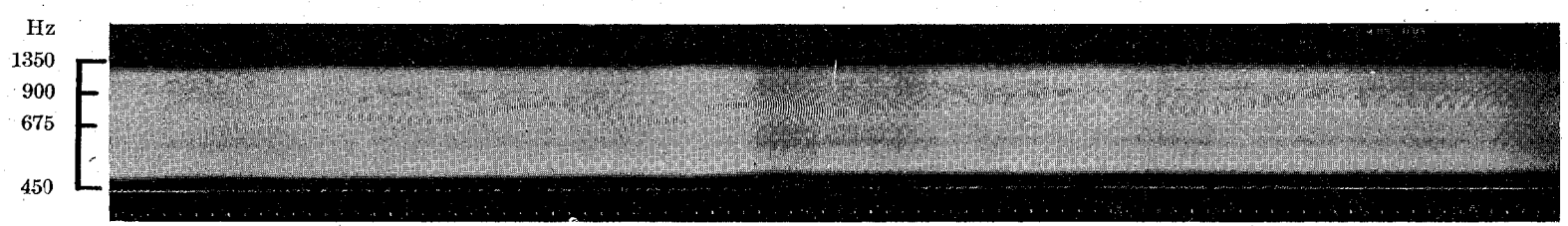


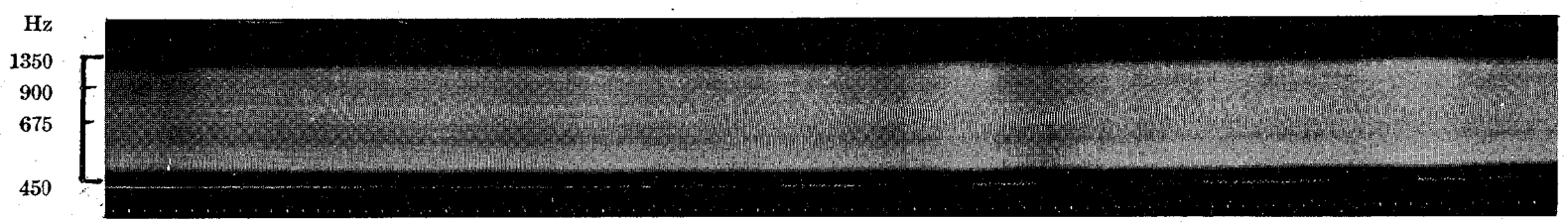
Conversational Speech. (Male)



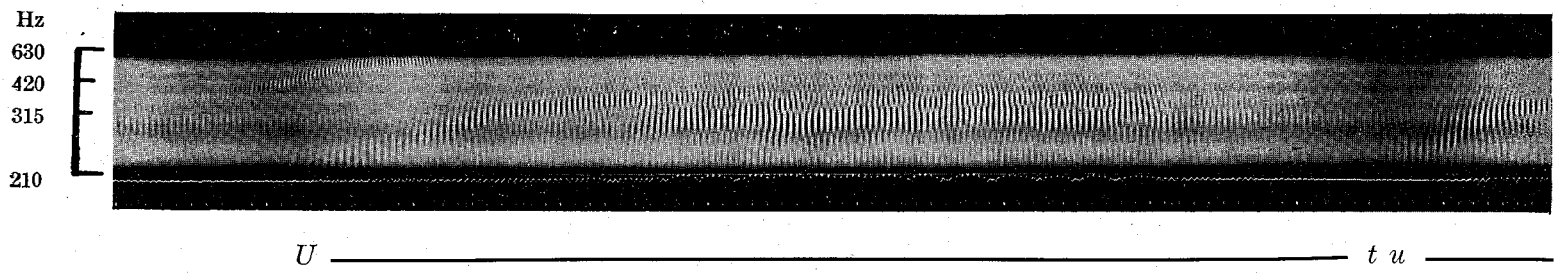


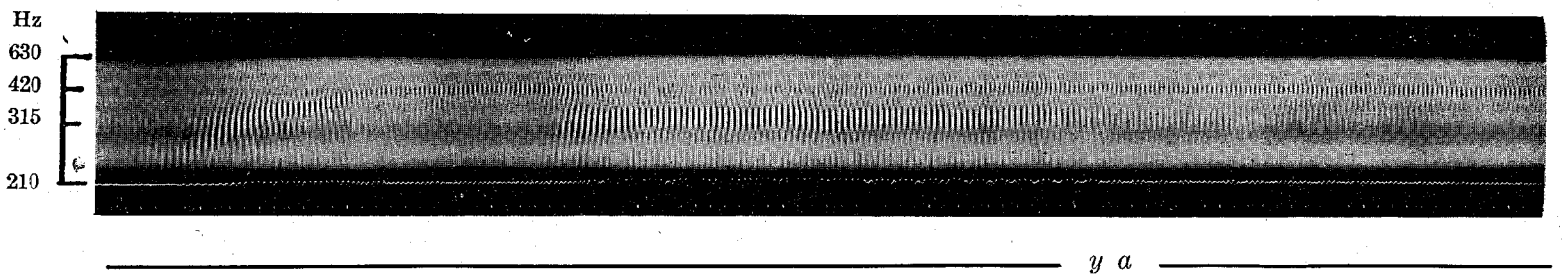
Duet by Capsir and Cecil.



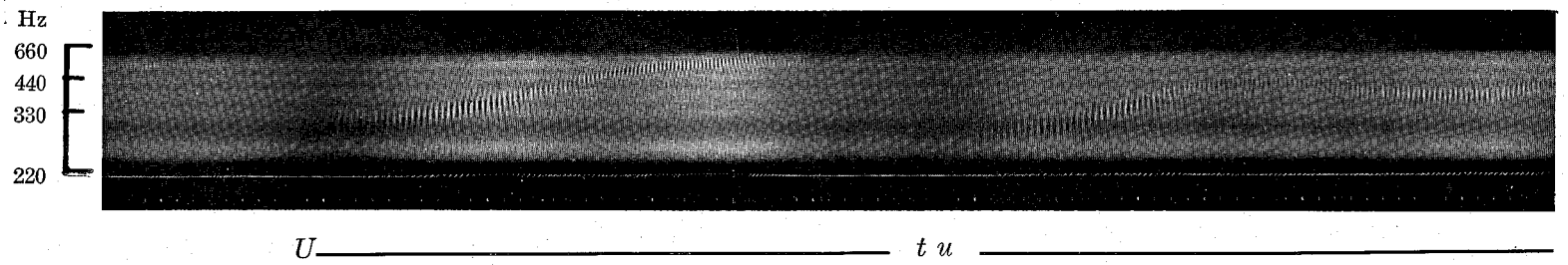


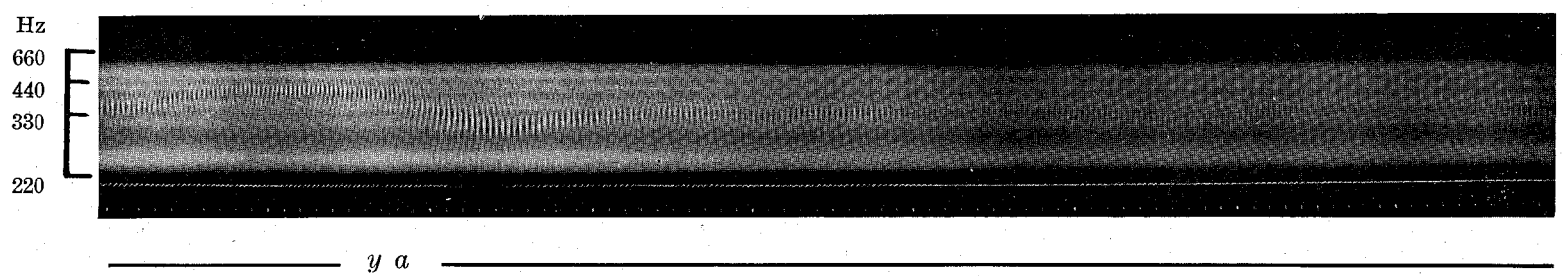
Naga-uta by W. Matunaga.



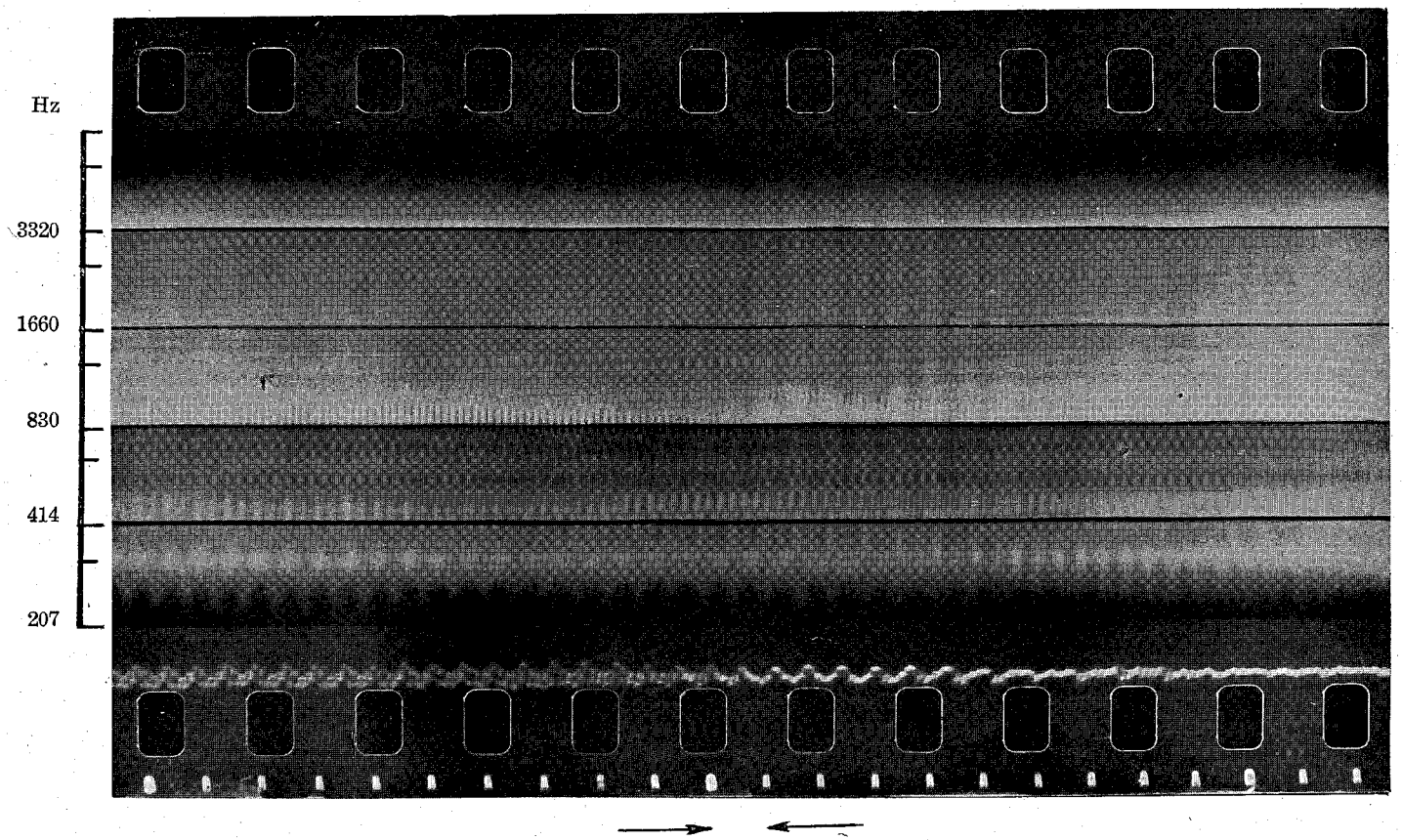


Naga-uta by K. Yosizumi.

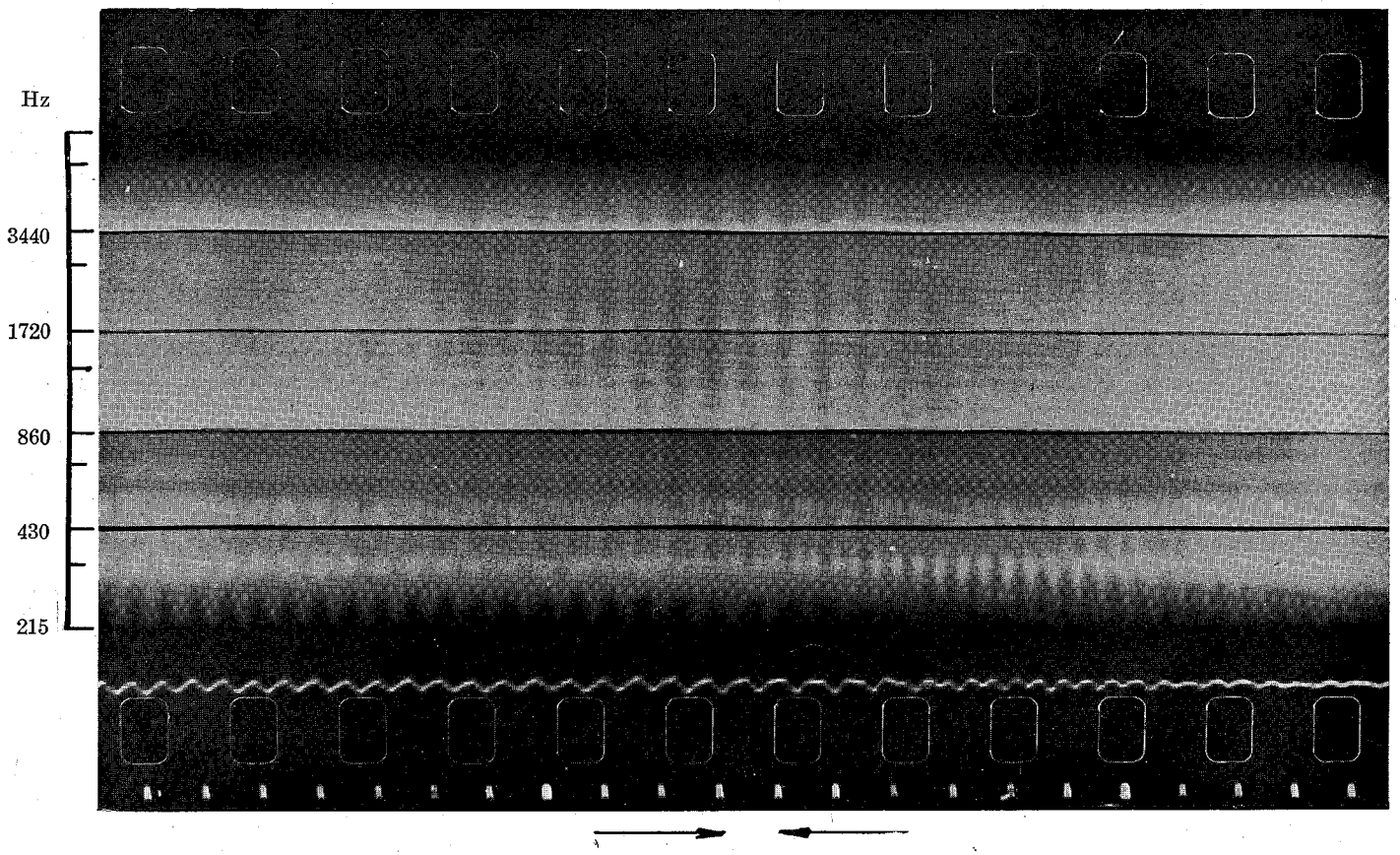




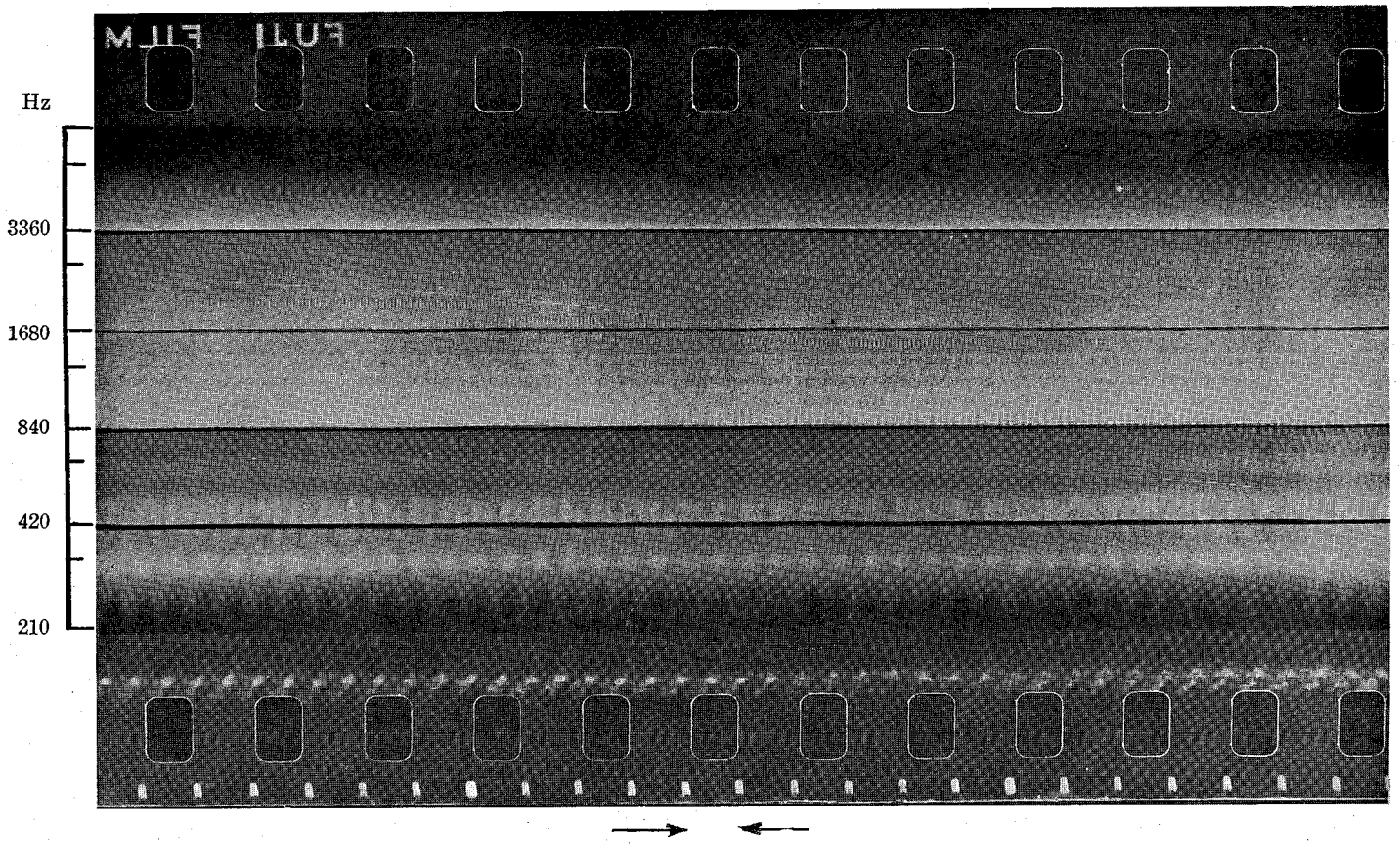
Vowel "Oh".



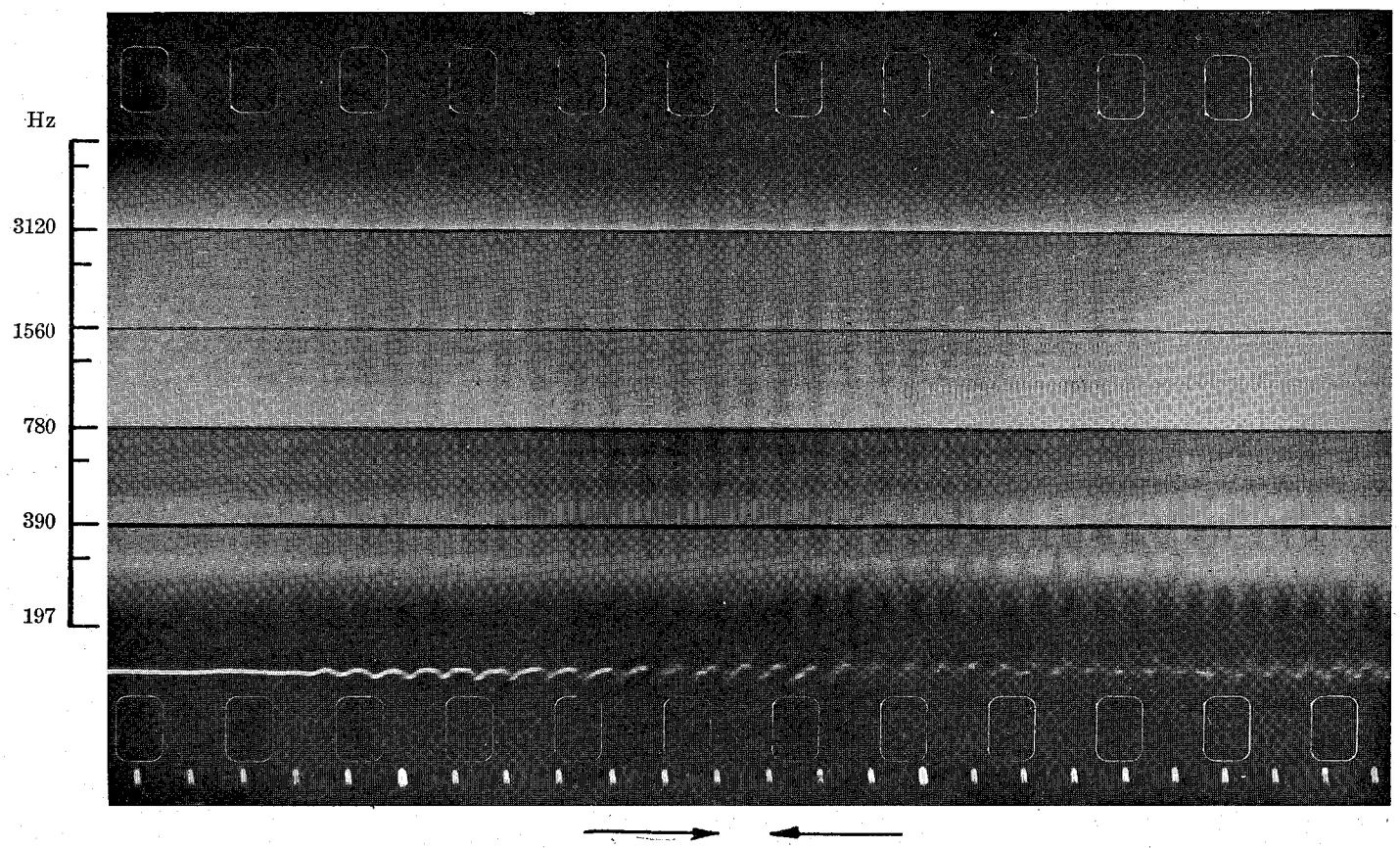
Vowel "Ii".



Formant Shift in "e-a".



Semi-vowel "y" in "ya".



Explosive Consonants "t" and "p".

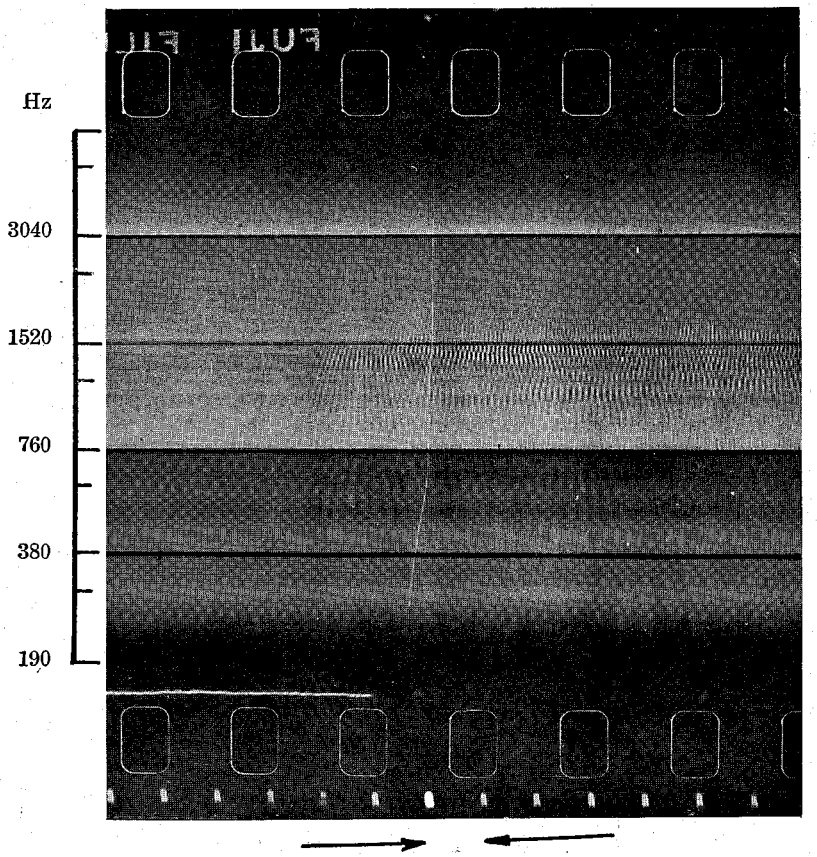


Fig. 1. "Ta" (unvoiced)

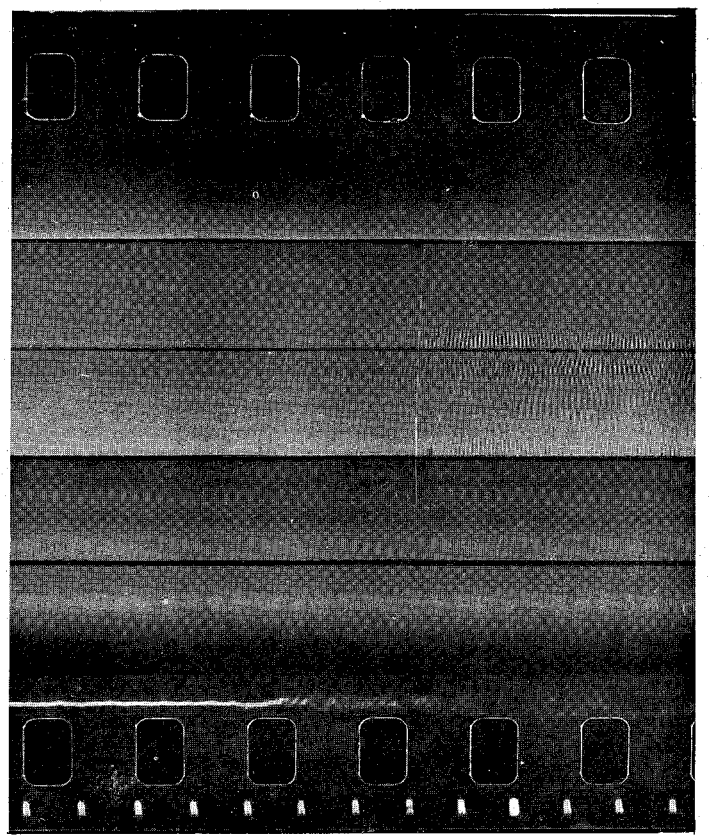


Fig. 2. "Pa" (unvoiced)

Percussion Sounds.

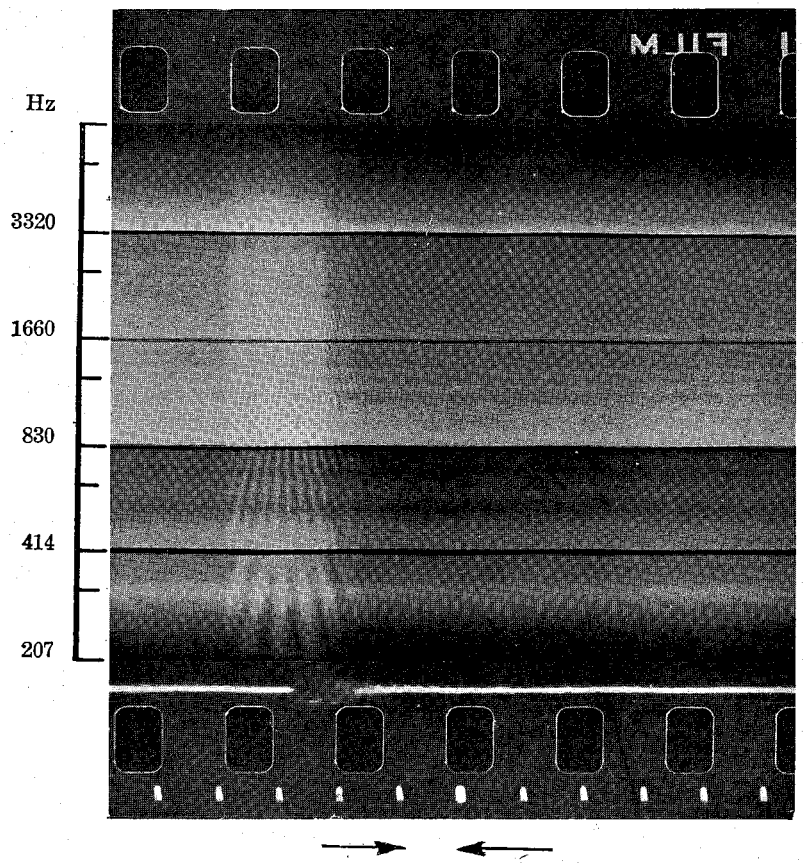


Fig. 1. Metronome Click.

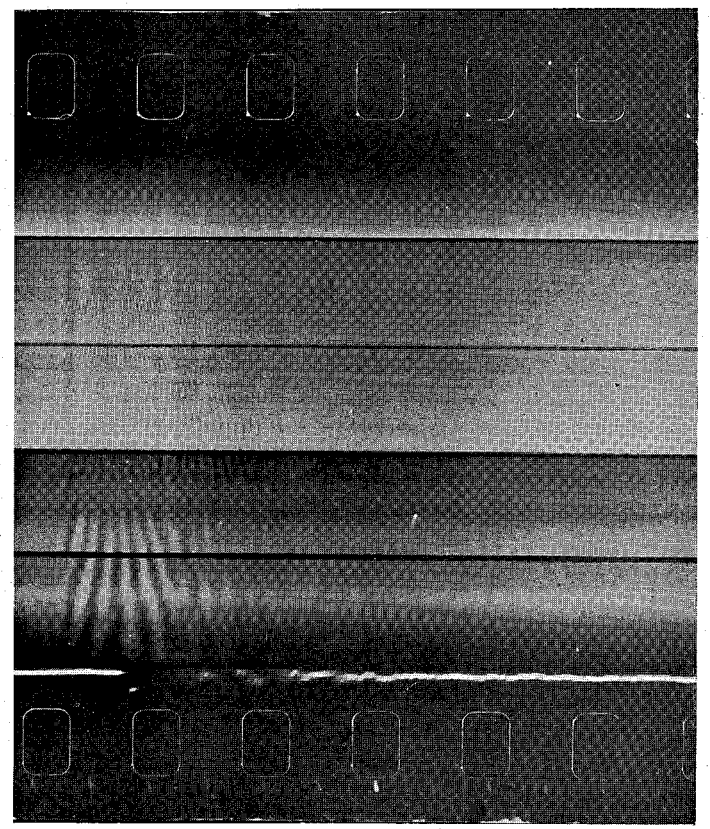


Fig. 2. Wooden Board,

School Whistle.