BEITRÄGE ZUR ERKENNTNIS DER KLANGFARBESTRUKTUR BEI VOKALISCHEN KLANGBILDERN

Yoshiyuki OCHIAI und Teruo FUKUMURA

Elektrotechnische Abteilung

(Eingegangen am 31. Mai 1956)

Einleitung

Es wurde schon erwähnt unter der Überschrift "Timbre Study of Vocalic Voices" die Untersuchungen über die Veranschaulichung von Vokalklangfarbe in Bezug auf Phonemische Qualität (phoneme quality), ebenso wie auf stimmliche Qualität (voice quality). Das der phonemischen Qualität entsprechende Klangfarbemuster, bezeichnet man als phonemisches Klangfarbemuster, und das Klangfarbemuster, daß der stimmlichen Qualität entspricht, nennt man das stimmliche Klangfarbemuster. Das erstere steht in Verbindung mit der Verständlichkeit der Sprache, und letzteres läßt sich verbinden mit der Natürlichkeit der Sprache. Denn die phonemische Färbung ist ganz anders, als die individuelle stimmliche Färbung oder Stimmfärbung.

Was die Technik der Musterbesorgung anbelangt, so ist das phonemische bzw. stimmliche Klangmuster durch die Übereinanderlagerungsmethode (superposition method) bzw. Nebeneinanderlagerungsmethode (juxtaposition method) bestimmt. In oben erwähnter Studie, führten wir schon die Beispiele der fünf japanischen Vokale von mittlerer Stimmstärke (d.h. etwa in mezzo forte) bei verschiedenen Tonhöhen an, die von fünf jungen Japanern ausgesprochen wurden. Es war uns damals aber nicht möglich, eingehende Untersuchungen mit voller Genauigkeit auszuführen. Wir beschränkten uns nur auf eine allgemeine Grundlegung der Klangfarbeanalyse und versuchten eine ungefähre Schätzung derselben. Infolgedessen wollen wir hier unsere damaligen Beschreibungen ergänzen, indem wir davon nur ein Beispiel von einer Sprecherin auswählen, und außerdem mehrere eingehende Beobachtungen, die sich auf Einzelheiten beziehen, durchzuführen versuchen.

Allgemeine Beschreibung

Unter der Voraussetzung der Pegelanpassung (level matching), wird das phonemische Klangfarbemuster bestimmt durch die symbolischen Formel

$$T_1(V) = \sum_{p} V(p_1) + \sum_{p} V(p_2) + \sum_{p} V(p_3) + \dots$$

= $\sum_{p} \sum_{p} V(p)$

wo V den betreffenden Vokal angibt, und p_1, p_2, p_3, \ldots die beim Aussprechen des Vokals benutzten Tonhöhen zeigen. Ferner verdeutlicht \sum_{p} den Prozeß der Übereinanderlagerung der Klangbilder eines und desselben Vokals, betreffs verschiedener Tonhöhen. Andrerseits wird das stimmliche Klangfarbemuster festgestellt durch die Formel

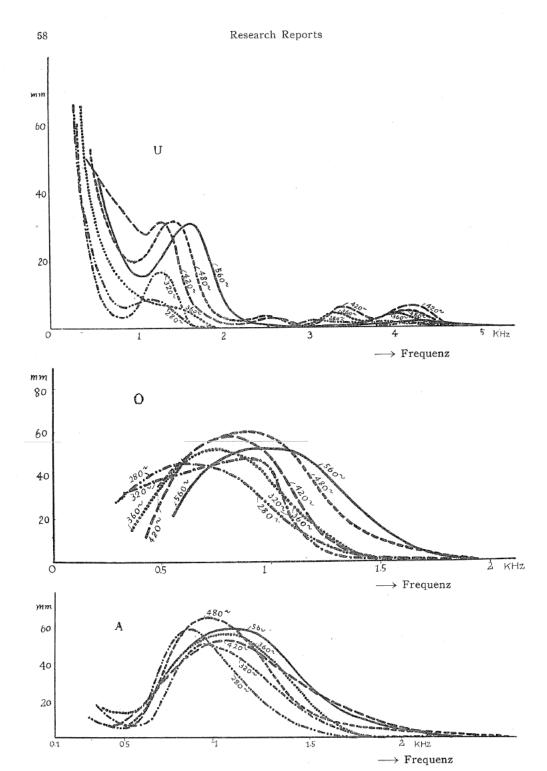
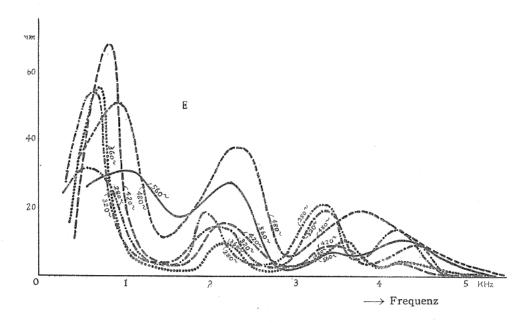


ABB. 1(a). Darstellung der Klangfarbestruktur der japanischen Vokale "U", "O" und "A", die ausschließlich von der Sprecherin H.H. mit verschiedenen Tonhöhen ausgesprochen wurden.



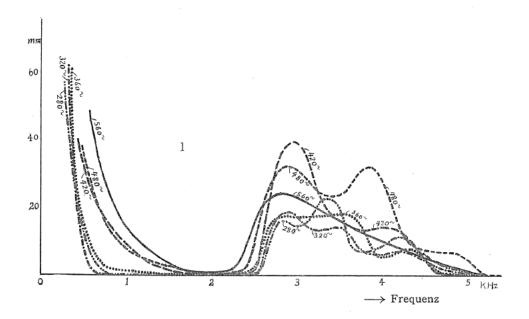


ABB. 1(b). Darstellung der Klangfarbestruktur der japanischen Vokale "E" und "I", die ausschließlich von der Sprecherin H.H. mit verschiedenen Tonhöhen ausgesprochen wurden. Jeder Kurve in Abb. 1(a), (b) ist das mittlere Resultat der fünf- od. sechsmaligen Messungen zugunsten der großen Schwankungen in Komponentendatum.

$$T_2(p) = \sum_{p} V_1(p) + \sum_{p} V_2(p) + \sum_{p} V_3(p) + \dots$$

= $\sum_{p} \sum_{p} V(p)$.

Hierbei deutet \sum_{r} den sog. entvokalisierenden Prozeß an, der sich durch das Nebeneinanderlagerungsverfahren verschiedener Vokale, aber mit denselben Tonhöhen, ermitteln läßt. Außerdem können wir uns den weißen Vokal vorstellen, der begrifflich bestimmt wird durch

$$W = T(V_1) + T(V_2) + T(V_3) + \dots$$

= $\sum_{v} T(V)$.

Der weiße Vokal mag für die Auslegung des Selbstlautes im Durchschnitt als bedeutungsvoll sich erweisen.

Besondere Eigenschaften der Stimme der Sprecherin H.H.

Die hier als Sprecherin dienende H.H. hat einen ziemlich großen Stimmumfang (wenigstens zwei Oktaven). Sie hat auch eine klare Stimme. Im Allgemeinen kann man sagen, daß manche Stimme von hoher Verständlichkeit, sehr oft eine geringere Qualität an Natürlichkeit hat. Um bei diesem Beispiel zu bleiben, ist es gekennzeichnet durch die Tatsache, daß ihre Stimme sowohl reich an Verständlichkeit, als auch an Natürlichkeit ist. Es ist außerdem nötig, die weibliche Stimme dieses charakteristischen Typuses in genügender Weise zu untersuchen. Denn wir haben schon eingehende Untersuchungen der männlichen Stimme beendet und auch wiederholt Berichte hierüber erstattet.

Phonemisches Klangfarbemuster

Abb. 1 veranschaulicht das phonemische Klangfarbemuster der fünf Vokale "I", "E", "A", "O", "U". Hier ist in deutlichster Weise die Verschiedenheit der Berandungskurvenformen der Vokale, die in Beziehung zum Auswählen der Tonhöhen stehen, zu ersehen. Man kann auch wahrnehmen, wie die Resonanzschärfe der Formanten und die Punkte maximaler Formantintensität sich verändern in Abhängigkeit von der Änderung der Tonhöhen der Stimmgebung. Zu Gunsten der besseren Darstellung der Genauigkeit, betreffs der Verteilung der Spektrkomponenten, geben die Abszissen die Frequenzen in linearer, anstatt in logarithmischer Skala an. Es bedeuten also die Ordinaten die Komponentamplituden (mm). Beachtenswert ist, daß in diesem Phenomen von gesprochenen Vokalen, "A" und "O" annähernd als Einziger-Formant Vokal erscheinen, während die Vokale "I", "E" "U" (*u* in japanischer Aussprache) bestehen fast aus Doppelt-Formanten Konstruk-Nur beim Aussprechen höherer Tonhöhe, können die Formantkomponenten im höheren Frequenzgebiete ziemlich stark erscheinen. Es ist nötig davon Notiz zu nehmen, daß die Feststellung der richtigen Lage der Formantspitzen in voller Abhängigkeit von dem Prozeß der Pegelanpassung steht. Wir benutzten hier hauptsächlich VU-(Volume Unit) Anpassung. Die Methode der Lautstärke, anstatt der Pegelanpassung, und zwar das Anpassen gleicher Vokalklänge verschiedener Tonhöhen, ebenso wie das gleicher Vokalklänge verschiedener Stimmen, mögen vieleicht von großer Bedeutung sein. Besonders wichtig und ganz unausweichlich ist das im Nebeneinanderlagerungsprozeß.

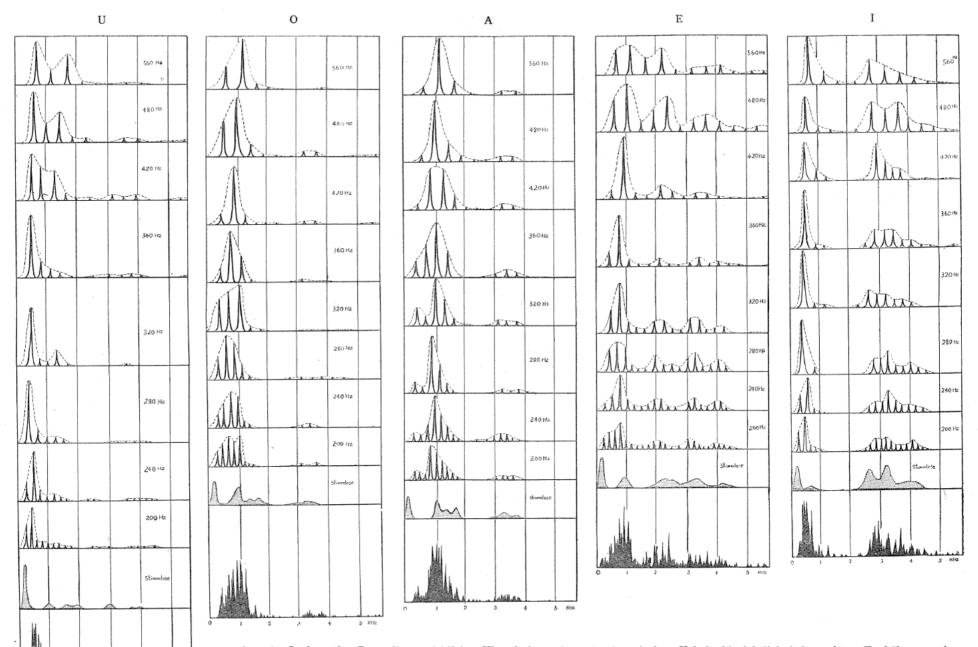


ABB. 2. Stufenartige Darstellung wirklicher Klangfarbestruktur der japanischen Vokale hinsichtlich jeder achten Tonhöhe, zwecks deutlicher Veranschaulichung der allmählichen oder plötzlichen Musteränderungen mit veränderter Tonhöhe. Jedes Bild in jder Zelle (Tonhöhe) ist das Durchschnittsresultat der fünf- od. sechsmaligen Meßungen, da es dabei beträchtliche Schwankungen in Bilder gibt, ungeachtet der Aussprache gleichen Vokals bei gleicher Tonhöhe,

Besondere Beobachtungen

Mit Rücksicht darauf, daß dieser Formantenbegriff bei der Fernsprechübertragung und Aufzeichnung der Vokale eine besondere Rolle spielt, haben wir einigen Verfahren für die Darstellung der Vokalformanten ausfindig zu machen. Es ist nicht ohne Interesse und auch nicht ohne Grund, die individuellen Klangfarbestrukturen in Bezug auf eine wirkliche Stimme mit etwas Tonhöhe andeuten zu versuchen. Denn man kann bekanntlich höher oder tiefer sprechen, ohne daß die Vokale ihre phonemischen Klangfärbungen stark ändern. Daraus folgt, daß dabei die Vokale aber ihre stimmlichen Klangfärbungen ziemlich ändern mögen. Hier sind wir bei einer interessanten Frage angekommen. Für die Beobachtungen dieser Art, können wir auf Abb. 2 hinweisen. Diese Abbildung zeigt die individuellen Vokalklangbilder bei 8 Tonhöhen, innerhalb des Intervalls zweier Oktaven (200,

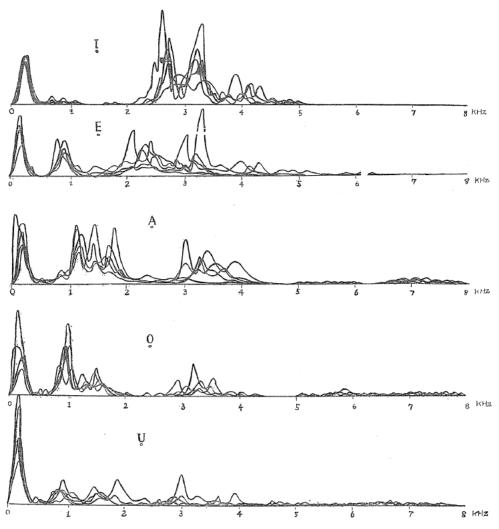
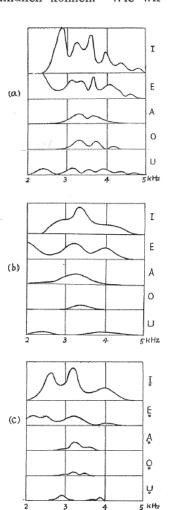


ABB. 3. Zusammenfassenderweise dargestellte Klangfarbebilder der fünf Flüstervokale.

240, 280, 320, 360, 420, 480 und 560 Hz); jedes ist nebeneinander in vertikale Richtung gestellt, und es wird zusammenfassend in jeder letzteren Zelle die sog. Formantenbauart jedes Vokals veranschaulicht. Diese Abbildung verdeutlicht uns auch, wie sich die Klangfarbestruktur ändert hinsichtlich der Änderung der Tonhöhe. Wir fügen noch die Klangbilder der sog. Flüstervokale hinzu, die in Verbindung mit dem stimmhaften Vokale treten. Es ist sehr wichtig zu untersuchen, ob dieser Flüstervokal dem Klangbild bei höherer Tonhöhe ähnlich ist, oder ob er nicht vielmehr dem bei niederer Tonhöhe ähnelt. Abb. 3 gibt das Flüstervokalmuster derselben Sprecherin an. Hierin ist jedes Vokalmuster beim Flüstern in solcher Weise zweckmäßig dargestellt, daß sich die im ganzen fünfmaligen Meßergebnisse für jeden einzelnen Vokal zusammenfassen lassen, da es dabei beträchtliche Abweichungen gibt, hinsichtlich der Formantenlage einerseits und der Formanten-Als Formantenform der Flüstervokale erhalten wir die intensität andrerseits. Berandungskurven, die die Kurvenscharen, die in Abb. 3 sich geben, ganzlich einhüllen können. Wie wir schon erörtert haben, können wir auch hier, um bei



demselben Beispiel zu bleiben, das Frequenzgebiet von 2 kHz mit den wenigsten Klangkomponenten bemerken. Auf dem Gebiete, daß wir vokalisches Tal oder Schlucht nennen wollen, treten die Vokalkomponenten nur gelegentlich auf. Infolgedessen können wir mit Hilfe des vokalischen Tals, die Klangbilder bestimmt trennen in zwei Teile: obere und untere Strukturen. Die erstere ist die Struktur auf dem Gebiete oberhalb des Frequenzpunktes 2 kHz, während die letztere die auf dem unterhalb des Frequenzpunktes 2 kHz ist. Insoweit es sich um die Feinstruktur des Klangbildes handelt, müssen wir auch unsere Betrachtungen einteilen. In Abb. 4 zeigen wir nur die Oberstrukturen, um besser ihre Feinstruktur beobachten zu können. Das obere Teilbild veranschaulicht die aus den phonemischen Klangfarbemuster erhaltenen Feinstrukturen, während das mittlere Teilbild dagegen die zeigt, die man aus den individuellen Klangbildern bei Tonhöhe 240 Hz erhält, und der untere Teil des Bildes gibt die gleichen wieder, die wir aus den Flüstervokalbildern erlangten. Vergleichende Beobachtungen dieser Teilbilder ermöglicht uns das Dasein der sog. Invarianten oder den

ABB. 4. Oberfeinstrukturen der Vokalklänge. (a) Das durch Nebeneinanderlagerungsverfahren ermittelte phonemische Klangfarbemuster der stimmhaften Vokale. (b) Wirkliche Klangfarbebilder der stimmhaften Vokale bei Tonhöhe 240 Hz. (c) Klangbilder der Flüstervokale.

unveränderlichen Formantteil (*invariant formant*) im oberen Frequenzbereiche zu bemerken, die bei dieser Sprecherin annähernd dem Frequenzgebiete 3.3 kHz entsprechen. Genauere Beobachtungen offenbaren ferner die interessante Tatsache, daß der unveränderliche Formant deutlich nur bei vier Vokalen ist, nämlich bei "I", "E", "A", "O" und das bei Vokal "U" derselbe in etwas verschobener Stellung erschien. Diese Vorgänge lassen sich deutlich bemerken nur im mittleren und unteren Bild. Im oberen Teilbild sind die oben angeführten Erscheinungen ziemlich selten. Schließlich ist zu erwähnen, daß bei derselben Sprecherin dieser veränderliche Formantteil am stärksten ist beim Aussprechen der Tonhöhe 240 Hz.

Diskussionsbemerkungen

In Bezug auf die Bedeutung der oben besprochenen bemerkenswerten Vorgänge, müssen wir weiter hier einen zusammenfassenden Einblick versuchen, da dort die einzelnen Vorgänge, jeder für sich, zusammenhangslos heraustritt. Daß wir hier

besonders das Schiefproblem der Unterstruktur behandelten, beruht auf dem Folgenden: Das Problem der Zuverlässigkeit der Verzerrung RSD (Revolutional Synchronous Distortion), worüber wir schon eingehend berichteten, zeigte die engste Verknüpfung mit der Klangfarbestruktur des durch die Verzerrung zu übertragenden Sprachlautes. Dort teilten wir die fünf Vokale, die von vier jungen Japanern ausgesprochen wurden, in drei Gruppen ein: Gruppe "A", "O", "I", die unsymmetrische Kennlinien in Zuverlässigkeit zeigt, d.h. mehr tief sinkende Charakteristik im Uals im D-Bereiche; einen anderen Typus gibt Vokal "U" an, der mehr tief sinkende Kennlinien im D- als im U-Bereiche aufweist; "E" ist ein einziger symmetrischer Typus. Bei Untersuchungen der Unterstruktur, können wir also die fünf Vokale in drei Gruppen einteilen: Gruppe, "A", "O", die nach unten hin und Gruppe "I", "U", die nach oben hin sinkt; "E" ist ein symmetrischer Typus. Es gibt kleine Unterschiede zwischen beiden Gruppierrungen, denn die Umstände bleiben sich nicht ganz gleich. Es ist daher nötig, weitere Untersuchungen durchzuführen, um sichere Resultate ermitteln zu können.

Aufmerksame Betrachtung der Abb. 2, ergeben ohne weiteres folgende Tatsachen:

 Hinsichtlich der sog. oberen Invarianten, die nur in der Oberstruktur zu be-

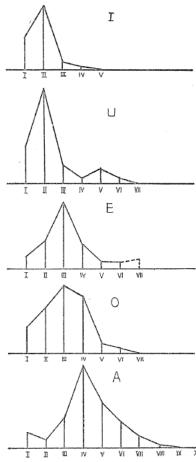


ABB. 5. Unterfeinstrukturen der fünf Vokale bei Tonhöhe 240 Hz.

merken sind, können wir wahrnehmen, daß dieselben am stärkste auftreten nur in der Nähe der Tonhöhe von etwa 240 Hz, bei der die Vokale ausgesprochen wurden, und ferner, daß diese Erscheinung fast unabhängig von dem Unterschiede der Vokale ist.

2. Andrerseits, gibt es auch einen beachtenswerten Vorgang betreffs der Resonanzschärfe in der Unterstruktur. Wir können beobachten, daß die Resonanz der Unterstrukturkurven schärfer oder stumpfer wird, in ganz enger Verknüpfung mit der Tonhöhe stehend. Es gibt auch einen Bereich der Tonhöhe, der eine verhältnismäßig höhere Resonanzschärfe in der Unterstruktur veranlasst, während es

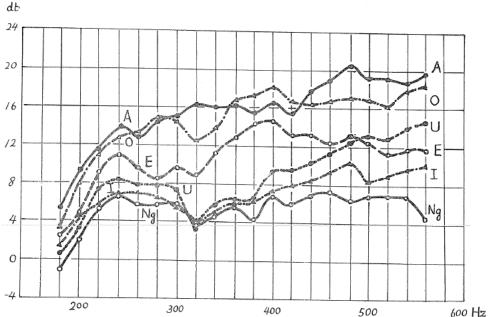


ABB. 6 (a). Abhängigkeit von Ausgangspegel der fünf mündlichen Vokale, ebenso wie eines Nasallautes. Stimmstärke möglichst konstant (mezzo forte) behaltend. Tonhöhe ist gewählt mit Abstand von 20 Hz im Tonbereiche von 180 Hz zu 560 Hz. Auswahl der Tonhöhe (betreffs ihrer Ordnung) ist ganz randomisiert, Beobachtungen unter einer Bedingung (Tonhöhe) und für jeden Sprachlaut beruhen im ganzen auf achtmaligen Messungen.

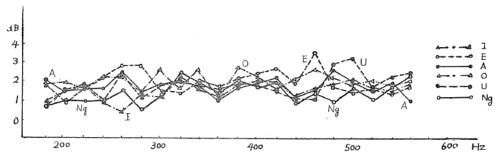


ABB. $6\,(b)$. Streuungen oder mittlere Schwankungen des Ausgangspegels bei denselben Sprachlauten, wie in Abb. $6\,(a)$.

auch einige Tonhöhen gibt, die eine etwas dumpfe Resonanzschärfe verursachen. Die Lage dieser besonderen Tonhöhen, die eine höhere bzw. niedere Resonanzschärfe zur Folge haben, mag wahrscheinlich in Abhängigkeit von der Vokalgattung stehen.

Wir zeigen z.B. in Abb. 5 die Unterstruktur der fünf Vokale bei Tonhöhe 240 Hz, dabei können wir bemerken, wie die Strukturformen abhängig von der Vokalgattung sind: z.B. aus dem Gesichtspunkte der Symmetrie der Unterstrukturformen, lassen sich die fünf Vokale einteilen in drei Gruppen: Der ganz besonders symmetrische Vokal "E", das ist so in fast allen Fällen, ohne Rücksicht auf den Unterschied der Tonhöhe; als unsymmetrische Typen die Vokale "I" und "U" in einer und die Vokale "A" und "O" in einer anderen Gruppe, hierbei steht die erstere im Gegensatz zur letzteren. Das Problem der Resonanzschärfe in den Unterstrukturen scheint von besonderer Wichtigkeit zu sein für Untersuchungen der Klangfarbequalität, und ferner die Vorstellung des Mechanismus der Vokal- und Stimmgebung in artikulatorischen Organen.

Abb. 6 zeigt uns die Abhängigkeit der Ausgangsstärke (dB Angabe des Lautstärkemessers) der fünf mündlichen Vokale, ebenso wie eines Nasallautes von den Tonhöhen, die von der Sprecherin H.H. beim Aussprechen benutzt wurden. Ohne Unterschied der Tonhöhe, war sie eifrig bemüht, ihre Stimmstärke doch fortwährend fast konstant einzuhalten, obwohl die Kontrolle nur durch die subjektive Beurteilung der Sprecherin selbst ausgeübt wurde. Aus dieser Kurve, die irgend eine Schwelle der Phonation dieser Sprecherin angibt, können wir einige interessante Tatsachen entnehmen. Auch in Bezug auf einen bestimmten Vokal, stimmt die die schärfste Resonanz in der Unterstruktur angebende Tonhöhe überein, mit der, bei der der Ausgangspegel des betreffenden Vokals einen Aufschwung erlebt, und die Tonhöhe, die die dumpfeste Resonanz angibt, entspricht der, die den stärksten unterdrückten Pegel zeigt (Siehe Abb. 2 und Abb. 6). Das scheint sowohl von Interesse, als auch von Wichtigkeit zu sein.

Zusammenfassung

Es handelt sich hier um die Feinstruktur bei der Darstellung der Klangfarbebauart. Eingehende Untersuchungen und einige Einzelheiten betreffende Beobachtungen wurden durchgeführt in Bezug auf eine weibliche Stimme und ihre fünf Vokale. Den Ablauf in der Änderung der Klangfarbestruktur beim Aussprechen mit veränderlichen Tonhöhen studierten wir in genauer Weise. Mit Hilfe des Begriffes "vokalisches Tal" könnten wir die Struktur trennen in zwei Teile: in Ober- und Unterstruktur. Die sogenannten Invarianten in der Oberstruktur und die "Schiefe" in der Bauart der Unterstruktur sind hier zum erstenmal in Frage gestellt.

Beim Abschluß dieser Studie, gebührt unser Dank dem Japanischen Unterrichtsministerium für die uns gewährte geldliche Unterstützung. Ferner danken wir herzlich Fräulein Hanako Higuchi für ihre freundliche Mitwirkung als ausgezeichnete Sprecherin, und dem Forschungsinstitut für Telekommunikation der Japanischen Telegraphen- und Fernsprech-Korporation, daß uns die Meßgeräte zur Verfügung stellte, sind wir zu Dank verpflichtet.

Literaturverzeichnis

1) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices. Memoirs of the Fuculty of Engineering (M.F.E.), Nagoya Univ., Vol. 5, No. 2, September, 1953.

- Y. Ochiai and T. Yamashita: On Timbre Quality (I). M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- 3) Y. Ochiai: Transmission of Quality. M.F.E., Nagoya Univ., Vol. 6, No. 2, November, 1954.
- T. Fukumura and K. Hara: Phoneme Figures of Sustained Oral Vowels by Two-Dimensional Representation (I and II). M.F.E., Nagoya Univ., Vol. 6, No. 2, November, 1954.
- Y. Ochiai, S. Saito, M. Watanabe and T. Fukumura: Quality Studies of Speech Transmission System Especially with Reference to Characteristic Forms in Frequency Response. M.F.E., Nagoya Univ., Vol. 6, No. 2, November, 1954.
- Y. Ochiai and T. Fukumura: Studies on Qualities of Speech and Voice by Timbre Distortion (or Carrier Synchronous Distortion). M.F.E., Nagoya Univ., Vol. 4, No. 2, November, 1952.
- Y. Ochiai: General Consideration on Studies of Speech Qualities in Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- 8) Y. Ochiai, S. Saito and Y. Sakai: Articulation Study of Speech Qualities in Rotational Synchronous Distortion. M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- Y. Ochiai and N. Izumitachi: Timbre Study on Mishearing Phenomena of Speech Phones in Rotational Synchronous Distortion (Report I: Mishearing of Vowelic Timbre). M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- 10) Y. Ochiai, S. Saito and Y. Watanabe: Allowance Problem in Rotational Synchronous Distortion as a Study on Timbre Discrimination by Infinitesimal Position-Shift in the So-Called Timbre-Space. M.F.E., Nagoya Univ., Vol. 7, No. 2, November, 1955.
- 11) Y. Ochiai and T. Fukumura: Study on Fundamental Qualities of Vocalic Timbre by Rotational Synchronous Distortion. (M.F.E., Nagoya Univ., in press).
- T. Fukumura and Y. Ochiai: Loudness Comparison between Vowels within Voices. M.F.E., Nagoya Univ., Vol. 7, No. 1, May, 1955.
- 13) Y. Ochiai and T. Fukumura: Timbre Study of Vocalic Voices Viewed from Subjective Phonal Aspect. Part I. Preliminary Studies on Naturalness and Articulation Qualities Actually and Directly Measured with Respect to Band-Eliminating Distortion. (M.F.E., Nagoya Univ., in press).
- 14) Y. Ochiai, T. Yamashita und T. Fukumura: Über den Grundbegiff der Übertragungsgüte. M.F.E., Nagoya Univ., III-1, 1950.
- Y. Ochiai und T. Yamashita: Logik der Treue als Grundproblem der Nachrichtentechnik. M.F.E., Nagoya Univ., III-2, 1950.
- Y. Ochiai: Über die Natürlichkeit der Sprachübertragungen. M.F.E., Nagoya Univ., IV-1, 1951.
- 17) Y. Ochiai: Mémoire sur les Sons des Voix Humaines. M.F.E., Nagoya Univ., Vol. 4, No. 1, July, 1952.
- 18) T. Yamashita and M. Tsuda: Some Observation of "Voicing Area" in Vocalization Phenomena. M.F.E., Nagoya Univ., Vol. 6, No. 1, May, 1954.
- 19) Y. Ochiai and T. Yamashita: Characterization of Transmissional Distortions in Speech Communication. M.F.E., Nagoya Univ., Vol. 6, No. 2, November, 1954.
- Y. Ochiai et S. Saito: Sur la Représentation de la Structure Phonétique des Langages. M.F.E., Nagoya Univ., Vol. 4, No. 2, November, 1952.
- T. Yamashita: Some Data on Voices Observed in the Wave Modes of Body-Surface Vibration. M.F.E., Nagoya Univ., Vol. 4, No. 1, July, 1952.