

Die Nachahmung von Stimmen

Voice Imitation

Gino G. Sacerdote

Die Stimme jedes Menschen besitzt eine manchmal sehr ausgeprägte Individualität, die ihrerseits abhängt von der anatomischen Gestalt der Sprechorgane, von Erbfaktoren, von emotionalen Bedingungen, aber vor allem auch von der Erziehung in Kindesjahren, die ihr Merkmale verleiht, welche die Zeit kaum zu ändern vermag.

Eine Möglichkeit zur Änderung des Charakters einer Stimme bietet eine entsprechende Schulung, besonders das auditive Training, wie es von den modernen didaktischen Methoden der Orthophonie sowie zur Vermittlung von Fremdsprachen verwandt wird. Andererseits kann auch allmählich, wenn auch nicht in allen Fällen, eine phonetische Anpassung bewirkt werden, wenn jemand lange Zeit in einer Umgebung lebt, in der eine andere

Zusammenfassung · Es werden die ersten auf dem Wege der Analyse gewonnenen Ergebnisse über die Fähigkeit einiger Versuchspersonen, eine andere Stimme der gleichen Sprache nachzuahmen, dargestellt. Die Versuchsperson versucht, den gleichen Text nachzusprechen, der ständig wiederholt wird, so daß die Nachahmung unabhängig von der Verständlichkeit und dem Verständnis ist. Die Analyse der nachgeahmten Stimme erfolgt mit Hilfe von Sprachmelodie-Kurven, mit Hilfe des Sonographen und durch oszillographische Aufzeichnungen.

Summary · The first results of analysis are described, concerning the power of some subjects of imitating another voice of the same language. The subject tries to imitate the same message, continuously repeated, so that the imitation is not affected by intelligibility and understanding.

The analysis on the repeated voice is made by melody curves, by sonagrams and oscillographic records.

Résumé · On décrit les premiers résultats de la recherche sur la possibilité de certains sujets d'imiter un'autre voix qui parle le même langage. Les sujets cherchent à imiter le même message qui est répété en continuation, de façon que l'imitation n'est pas affectée par l'intelligibilité et la compréhension. Les analyses des voix répétés sont données par des courbes mélodiques, des sonagrammes et des enregistrements oscillographiques.

The voice of every person has an individuality sometimes very pronounced, depending on the anatomical conformation of the vocal organs, on hereditary factors, on the emotional conditions, and chiefly on the first education, which impresses characteristics that time can hardly alter.

The possibility to alter the characteristics of a formed voice can be product of an education, specially auditive, as the more modern didactic methods employ for orthophony as well as for foreign languages. On the other hand living for a long time in an environment where another language or dialect is spoken than one's mother tongue produces slowly, but not always, a phonetic conditioning. It is known that the characteristic Italian of the American immigrants,

Sprache oder ein anderer Dialekt gesprochen werden als seine Muttersprache. Bekannt ist, daß das charakteristische Italienisch der nach den USA Eingewanderten sowie das Französisch gebürtiger Deutscher stets von der Muttersprache beeinflußt bleiben.

Es soll nun untersucht werden, inwieweit die Möglichkeit besteht, daß jemand seine Stimme in sehr kurzer Zeit an eine andere angleicht. Bei dieser Untersuchung werden wir uns vielen Fragen phonetischer, audiometrischer und psychologischer Natur zuwenden müssen: Die Fähigkeit zur Assimilierung einer Sprache in manchen ihrer Aspekte ist viel größer bei Personen, die musikalisch sind und eine Ausbildung in Musik genossen haben.

Auch sollte im Zusammenhang mit der Möglichkeit der Nachahmung einer anderen Stimme die Bedeutung des Trainings berücksichtigt werden; so wäre es denkbar, daß man eine nachgeahmte Stimme auf ein Tonband aufnehmen und Verhöre und Aussagen von juristischer Bedeutung verfälschen könnte: deshalb sollte man der Fähigkeit der raschen Assimilierung bei der Beurteilung der Gültigkeit von auf Tonbändern aufgenommenen Zeugaussagen besondere Aufmerksamkeit schenken.

Die bei unseren Untersuchungen angewendete experimentelle Methode besteht darin, daß einem Zuhörer über Kopfhörer ein auf Tonband aufgenommener Text übermittelt wird. Die Versuchsperson wiederholt die gehörten Wörter, welche wiederum zum Vergleich mit dem Originaltext auf Tonband aufgenommen werden.

Viele Versuche sind von E. Colin Cherry durchgeführt worden, der die Fähigkeit des Verstehens und des Wiederholens von Texten untersucht hat, die in einem bestimmten Rhythmus auf das eine oder andere Ohr übertragen werden: Ziel dieser Untersuchungen ist es, die Verständlichkeit eines Textes zu prüfen, die vom Rhythmus der Alternationen abhängt.

Küpfmüller und Naujoks haben Untersuchungen durchgeführt, bei denen sie den übermittelten und den wiedergegebenen Text vom Standpunkt der Verzögerungszeit zwischen ersterem und letzterem verglichen haben. Nach Ansicht dieser Autoren ist die beobachtete Verzögerungszeit die Summe zweier Zeiten: der Reaktionszeit und der Rezeptionszeit.

Bei einem Text mit einer Dauer von 2 Minuten liegt die Verzögerungszeit zwischen 300 und 900 msec. und einem Durchschnitt von 550 msec.

Bei der Untersuchung der Ergebnisse stellten die Autoren eine Reaktionszeit von 240 msec und eine Rezeptionszeit von 310 msec. fest.

and the French of the German natives, were always affected by the mother tongue.

Our purpose is to examine the possibility to equalize one's voice to another, within a very short time; this research may involve many questions of phonetic, audiometric and psychological character: the power of assimilating a language for some of its aspects is much greater for subjects having musical ear and education.

The importance of training in the possibility of imitating another voice; it could be possible to record on a magnetic tape an imitated voice, should also be considered distorting trials and witnesses of juridical value: for this reason attention should be paid to this power of quick assimilation in evaluating the legitimacy of a witness attained through record.

The experimental technique employed for these studies induces a message recorded on a magnetic tape to a listener through ear-phones. The subject repeats the listened words, that are recorded for a comparison with the original message.

Many experiments are due to E. Colin Cherry, who investigates the power of understanding and repeating a message sent rhythmically to one or the other ear: the purpose of these studies is to investigate the intelligibility of the message, which depends on the rhythm of the alternations.

Küpfmüller and Naujoks have carried out some investigations to correlate the received and the repeated message from the point of view of the delay time of one when compared with the other. In the opinion of these authors the delay time observed is the sum of two times: a reaction time and an understanding time.

In a message of about 2 minutes the delay time is comprehended between 300 and 900 milliseconds, with an average of 550 milliseconds.

Examining the results the authors determine a reaction time of 240 milliseconds and an understanding time of 310 milliseconds. The total delay time is a function of the speed of the primary voice, while it seems not to be affected by the intensity of this voice. The research has been continued studying the repetition of single vowels: the reaction time has been found not to change very much around 180 milliseconds; while the understanding time goes from a minimum of 70 milliseconds for the *e* to a maximum of 100 milliseconds for the *o*.

Besides these, Tomatis has carried out some important investigations in the field of the conditioning of the voice through the listen-

Die Gesamtverzögerungszeit ist eine Funktion des Sprechtempos der Originalstimme, scheint jedoch unabhängig zu sein von der Lautstärke der Stimme. Die weitere Untersuchung beschäftigte sich mit der Wiedergabe von einzelnen Vokalen; es wurde festgestellt, daß die Reaktionszeit geringfügig um 180 msec. schwankt, während die Rezeptionszeit von einem Minimum von 70 msec. für das *e* bis zu einem Maximum von 100 msec. für das *o* reicht.

Außer diesen Autoren hat Tomatis wichtige Untersuchungen durchgeführt über den Zustand der Stimme durch bloßes Zuhören beim Erlernen von Sprache, Gesang und Fremdsprachen.

Anderer Art sind demgegenüber die Methoden und Ziele der zur Zeit in Gang befindlichen Untersuchungen des Laboratoriums für Akustik am *Istituto Elettrotecnico Nazionale*, Turin. Die „Rezeptionszeit“ wird ausgeklammert, da die Stimmnachahmung in der Weise erfolgt, daß immer der gleiche von derselben Person gesprochene und vom Tonband ständig wiederholte Satz wiedergegeben wird.

Folgende Methode wird angewendet: Jemand spricht in einem akustisch hergerichteten Raum einen kurzen Satz, der auf ein Tonband aufgenommen wird. Der gewählte Satz lautet:

le tue aiuole

(*le tooeh abyoo-ohleh: deine Blumenbeete*)

Der Satz wurde nach folgenden Gesichtspunkten gewählt: er sollte weniger als 3 Sekunden dauern, damit eine Analyse mit Hilfe des Sonographs möglich ist, und sollte alle 5 Vokale, einen dentalen Konsonanten und einige tonlose Pausen enthalten.

Der Satz wird auf ein geschlossenes Tonband aufgenommen, wird ständig wiederholt und einer Versuchsperson mittels Kopfhörern übermittelt. Die Versuchsperson wiederholt den Satz ständig und versucht dabei, die Originalstimme so gut wie möglich nachzuahmen. Der Versuch dauert drei Minuten. Sowohl die Stimme der Versuchsperson als auch die Originalstimme werden gleichzeitig auf ein zweispuriges Tonband aufgenommen, und zwar zu Beginn des Experiments, zu verschiedenen Zeitpunkten während der Dauer des Experiments und bei dessen Beendigung.

Außerdem wird die Stimme der Versuchsperson vor Beginn des Experiments, d. h. wenn sie noch unbeeinflusst ist vom Abhören der Originalstimme, auf Tonband aufgenommen.

Die Untersuchung der Tonbandaufnahmen und ihr Vergleich mit Hilfe von verschiedenen Analysen gibt Aufschluß über die Fähigkeit verschiedener Versuchspersonen zur Nach-

ing in learning, speech sing and foreign language.

The research in course at the acoustical laboratory of the *Istituto Elettrotecnico Nazionale* of Turin has methods and purposes rather different: the "understanding time" is eliminated, as the voice imitation is performed repeating always the same phrase pronounced by the same subject and continuously repeated by magnetic recording.

The following technique is used: a subject pronounces in an acoustically treated room a short phrase, that is recorded on a magnetic tape. The chosen phrase is

le tue aiuole

(*leh tooeh abyoo-ohleh: your flower beds*).

The choice is determined by the following elements: the phrase should last less than 3 seconds, to allow an analysis with the sonograph, it should contain all the 5 vowels, a dental consonant and some silent intervals. The phrase is recorded on a closed magnetic band continuously repeated and listened to by a subject through earphones. The subject repeats continuously the phrase, trying to imitate the original voice as well as possible. The experiment takes 3 minutes; the voice of the subject and the primary voice are recorded simultaneously on a two-track magnetic tape at the beginning, for a given number of intermediate times and at the end of the experiment.

Before the experiment the voice of the subject is recorded, when it is not yet affected by listening to the primary voice.

Examining the record, comparing them by different kinds of analyses, the power of imitating the primary voice is investigated for different subjects: the investigation concerns above all the following elements of the language: melody, rhythm, timbre.

The imitating power is correlated with the attitudes and the musical education of the subject. The imitating attitude is investigated for similar voices and for very different voices, too.

Beyond the physical analysis of the results, listening directly may help very much in judging the attitude and the power of imitating.

This research is continued examining the tuning of speech or singing voice with pure sounds of constant frequency, not corresponding to a musical scale. In this case also the waiting time, the quickness of intonation, the tuning error, in some cases the impossibility of tuning are elements that can indicate the psychophysical characteristics of the subject.

ahmung einer Originalstimme: die Untersuchung bezieht sich insbesondere auf folgende Sprachelemente: Melodie, Rhythmus, Timbre. Die Fähigkeit zur Nachahmung wird in Beziehung gesetzt zu den Attitüden der Versuchsperson und ihrer Ausbildung in Musik. Die Imitationsattitüde wird sowohl für ähnliche als auch für ganz verschiedene Stimmen untersucht.

Abgesehen von der physikalischen Analyse der Ergebnisse kann sich auch das direkte Zuhören für die Beurteilung der Imitationsattitüde und -fähigkeit als sehr hilfreich erweisen.

Die weitere Untersuchung erstreckt sich auf die Anpassung an die Tonlage der gesprochenen oder gesungenen Stimme, und zwar für reine Töne mit konstanter Frequenz, die keine Beziehung zur Tonleiter haben. In diesem Falle gehören auch die Verzögerungszeit, die Schnelligkeit der Intonation, Fehler im Treffen des Tones, in einigen Fällen auch die völlige Unmöglichkeit den Ton zu treffen, zu den Elementen, die Aufschluß über die psychophysikalischen Merkmale der Versuchsperson geben.

Die Wiederholung des gleichen Textes über eine kurze Zeit kann Lernphänomene zutage treten lassen; es sollte berücksichtigt werden, daß in beschränktem Maße die Möglichkeit besteht, die eigene Stimme zu hören.

Für den Satz »le tue aiuole« wurde zunächst nach der Methode von Grützmaker die Sprachmelodie untersucht. Bei diesen Messungen können Schwierigkeiten auftreten, da die Oberschwingungen der Stimme die gleiche Amplitude oder manchmal eine größere Amplitude haben als die Grundschwingung (folglich muß, um die Grundschwingungen herauszufiltern, ein Filter benutzt werden, dessen Durchlaßfrequenzband schmal ist); darüber hinaus kann sich die Grundschwingung über mehr als eine Oktave erstrecken, und der Filter sollte deshalb ziemlich breit sein. Mit Hilfe eines automatischen Frequenz-Bandschalters, der an anderer Stelle beschrieben wird [1], kann man ohne Schwierigkeiten die Grundschwingung erkennen: einige Ergebnisse werden in Abb. 1 dargestellt.

Die Angaben in dieser Abbildung beziehen sich auf drei Personen:

A – eine männliche Stimme, mit tiefem Timbre, Versuchsperson mit guter musikalischer Bildung, 36 Jahre;

D – eine männliche Stimme, mit mitteltiefem Timbre, Versuchsperson ohne besondere musikalische Bildung, 21 Jahre;

B – eine männliche Stimme, mit hohem Timbre, Versuchsperson ohne musikalische Bildung, 19 Jahre.

Repeating for a short time the same message phenomena of learning may be put into evidence; it should be taken into account that there is a limited possibility of hearing one's voice.

For the phrase "le tue aiuole" the melody was first examined, employing the method of Grützmaker. In this measurement some difficulties may arise, due to the fact that the harmonics of the voice have the same amplitude, or sometimes a greater amplitude than the fundamental (as a consequence it is necessary to employ a narrow filter to extract the fundamental); besides, the fundamental may have an extension larger than an octave band, as a consequence the filter should be rather large. An automatic band switch [1] allows to discriminate easily the fundamental: some results are reported in fig. 1.

The data of this figure concern three subjects: A – a male voice, with a low timbre, relative to a subject with good musical education, 36 years old;

D – a male voice, with a mean-low timbre, relative to a subject without particular musical education, 21 years old.

B – a male voice, with high timbre, relative to a subject with no musical education, 19 years old.

In the different diagrams the higher part represents the melody curve, with frequencies varying between 100 and 200 cps. Below, the oscillogram of the emitted voice is reported. Beginning from the upper part of the figure, for every subject the diagrams report the melody of the original voice, and of the imitating voice after 15 seconds, after 1, 2 and 3 minutes.

The diagrams marked E concern the voice to be imitated, emitted by a male voice, with high timbre, relative to a subject 24 years old.

These diagrams show that A, modifying his voice very much, can imitate pretty well, also after a rather short time.

D has a melody curve rather irregular with the time and some tendency to the imitation. B, in spite of his lack of musical education, succeeds sufficiently in imitating the melody, due to the fact that the primary voice and his own have a pretty similar timbre, as can be observed from the first and the last diagram of the column.

The analysis by the sonograph can give other elements on the way of imitating the voice: 5 groups of diagrams of the sonograph are reported in fig. 2: the first, number 1, concerns the voice of the original subject, number 2 the imitating voice after 15 seconds,

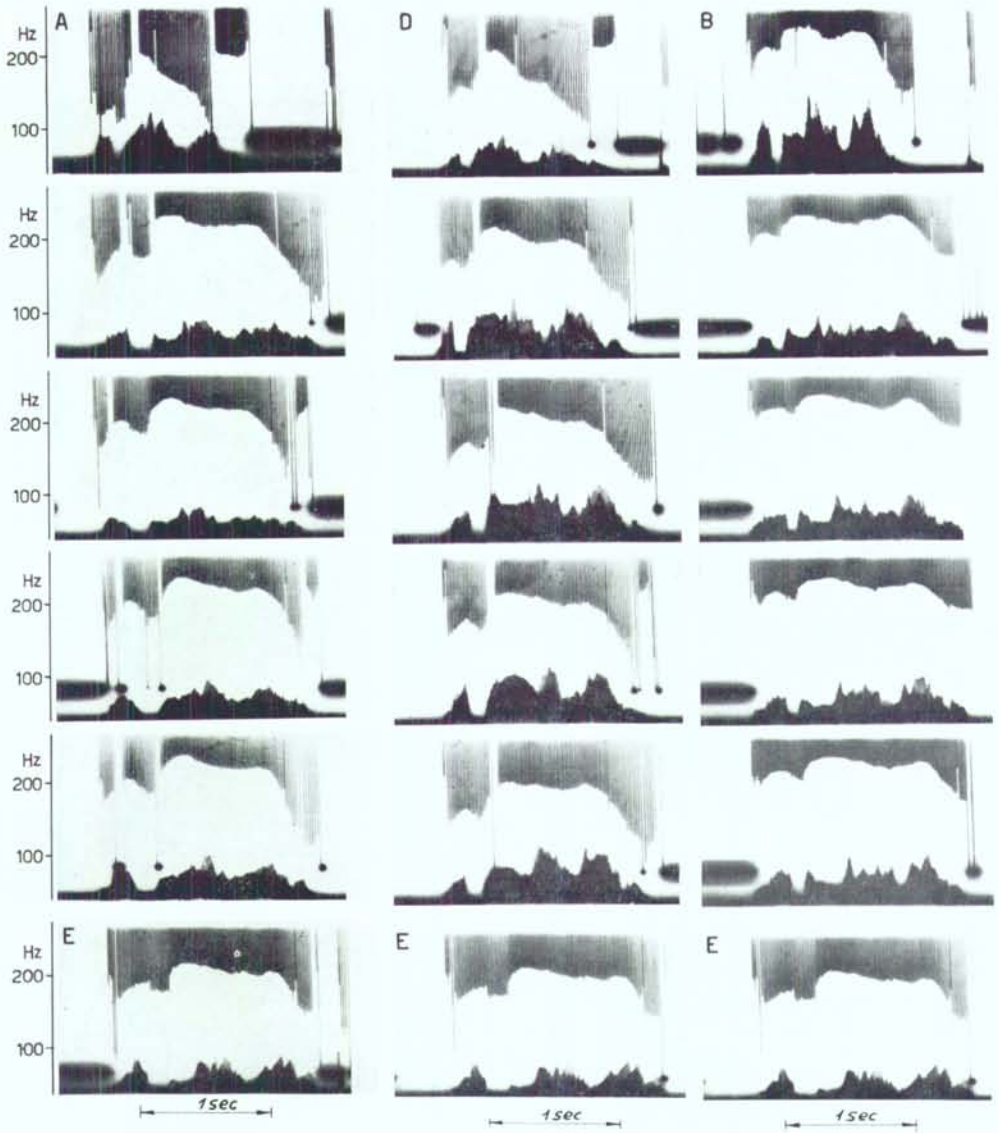


Abb. 1 Sprachmelodie-Kurven von drei Versuchspersonen (A, D und B) im Vergleich zu der Melodiekurve der nachzuhörenden Stimme

Fig. 1 Melody curves for three subjects (A, D and B) compared with the melody curve of the voice to be imitated

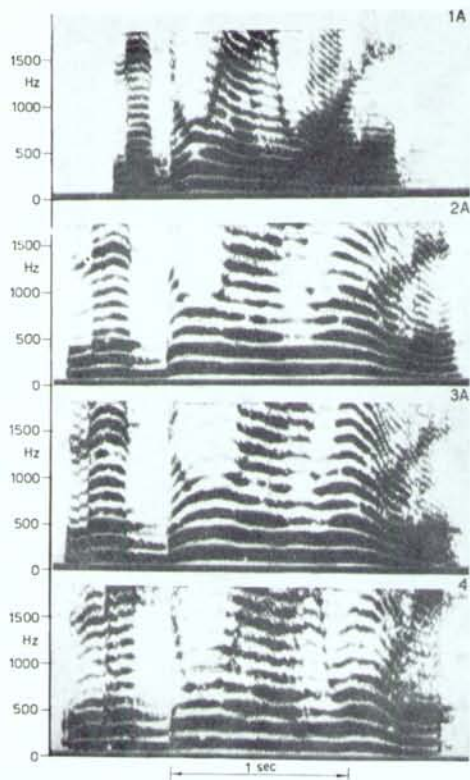


Abb. 2 Sonogramme von fünf Versuchspersonen (A, B, C, D und E)

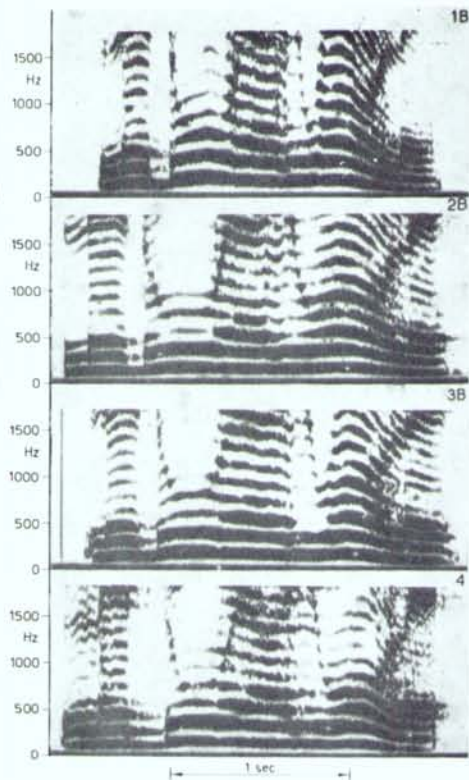


Fig. 2 Sonograms of five subjects (A, B, C, D and E)

Der obere Teil der einzelnen Diagramme stellt die Kurve der Sprachmelodie dar, deren Frequenz zwischen 100 und 200 Hz schwankt. Darunter befindet sich das Oszillogramm der emittierten Stimme. Die Diagramme beginnen am oberen Ende der Abbildung und zeigen für jede Versuchsperson die Melodie der ursprünglichen Stimme sowie die der nachgeahmten Stimme nach 15 Sekunden, nach 1, 2 und 3 Minuten.

Die mit dem Buchstaben E gekennzeichneten Diagramme zeigen die nachzuahmende emittierte Stimme, und zwar eine männliche Stimme, mit hohem Timbre; der Sprecher ist 24 Jahre alt.

Die Diagramme zeigen, daß A, der seine Stimme sehr stark modifiziert, eine gute Nachahmungsfähigkeit besitzt, und zwar schon nach verhältnismäßig kurzer Zeit.

D zeigt eine Melodiekurve ohne deutliche Korrelation zur Versuchszeit, besitzt aber einige Imitationsneigung.

B kann trotz fehlender musikalischer Bildung

number 3 the imitating voice after 3 minutes, and last number 4 the voice to be imitated.

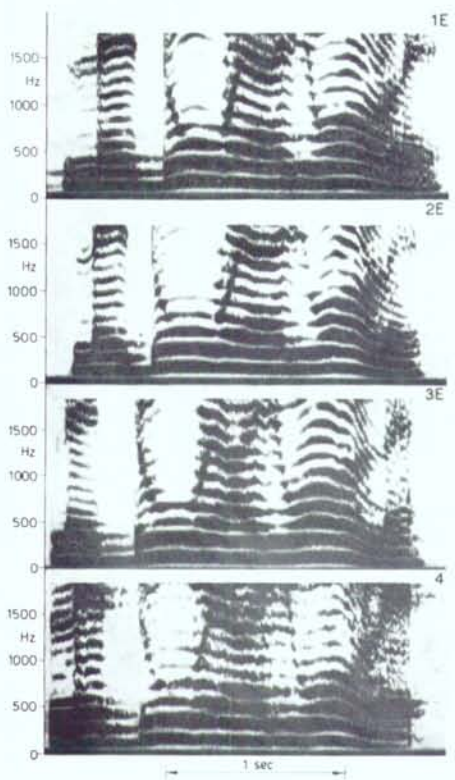
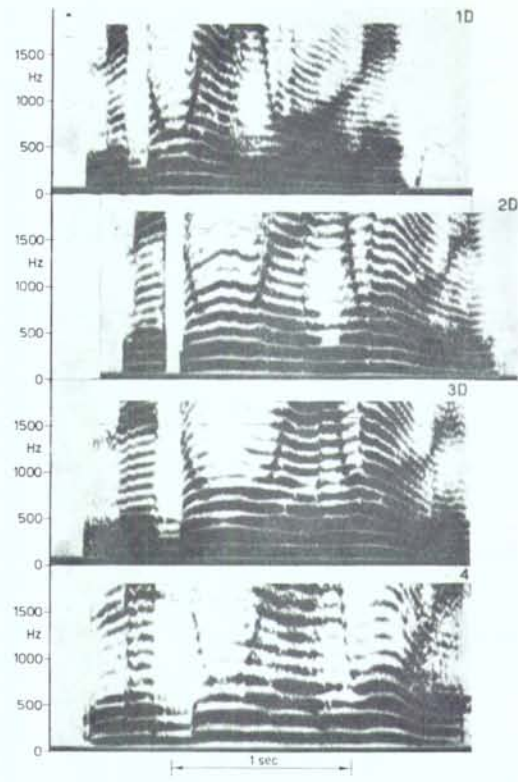
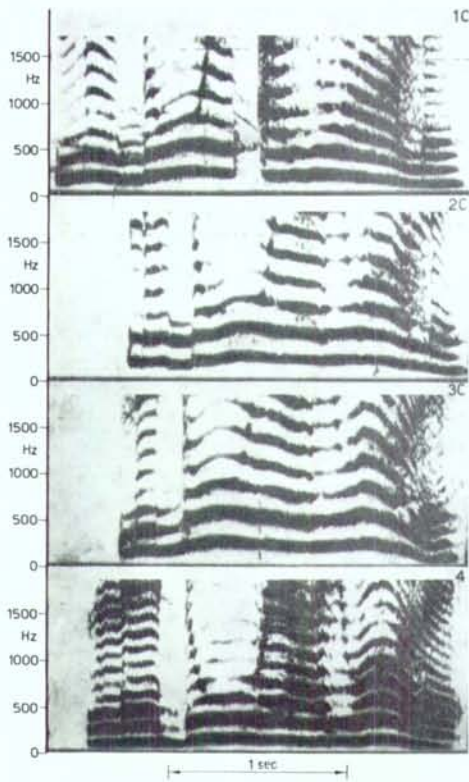
The normal "visible speed" has been modified so that its frequency scale is expanded, to show better some details.

Beyond the three subjects already described, the data concerning C are reported: a female voice, with high timbre, musically educated, 37 years old.

These figures show that A imitates well, D pretty well, C well, but with a higher timbre, and the limits of repeatability of E, with no doubt better than the other cases.

A third group of diagrams (fig. 3) reports the oscillograms of the recorded voice from a cathode ray tube. Oscillograms 1 concern the original voice of subjects A, B, C, D and E; the second oscillograms the imitating voice after 15 seconds, the third the imitating voice after 3 minutes.

Oscillograms 2 and 3 report, in their higher part, the voice to be imitated, on a reduced scale; oscillograms E reports the same voice on the same scale as the other subject.



die Sprachmelodie hinlänglich gut nachahmen, weil das Timbre der Originalstimme dem seiner eigenen Stimme sehr ähnlich ist, wie dies das erste und letzte Diagramm der Spalte zeigen.

Andere Elemente der Stimmnachmung erschließt die Analyse mit Hilfe des Sonographs. Abb. 2 zeigt fünf mit dem Sonograph aufgestellte Diagramme: das erste, Nr. 1, stellt die ursprüngliche Stimme der Versuchsperson dar, Nr. 2 die Stimme der Versuchsperson nach 15 Sekunden, Nr. 3 die Stimme der Versuchsperson nach 3 Minuten und das letzte, Nr. 4, die nachzuahmende Stimme.

Die normale »sichtbar gemachte Sprache« wurde modifiziert, so daß ihr Frequenzbereich erweitert wurde, um einige Einzelheiten besser hervortreten zu lassen. Zusätzlich zu den drei bereits beschriebenen Versuchspersonen liegen die Daten für C vor: eine weibliche Stimme mit hohem Timbre, musikalisch gebildet, 37 Jahre.

Diese Abbildungen zeigen, daß A gut nach-

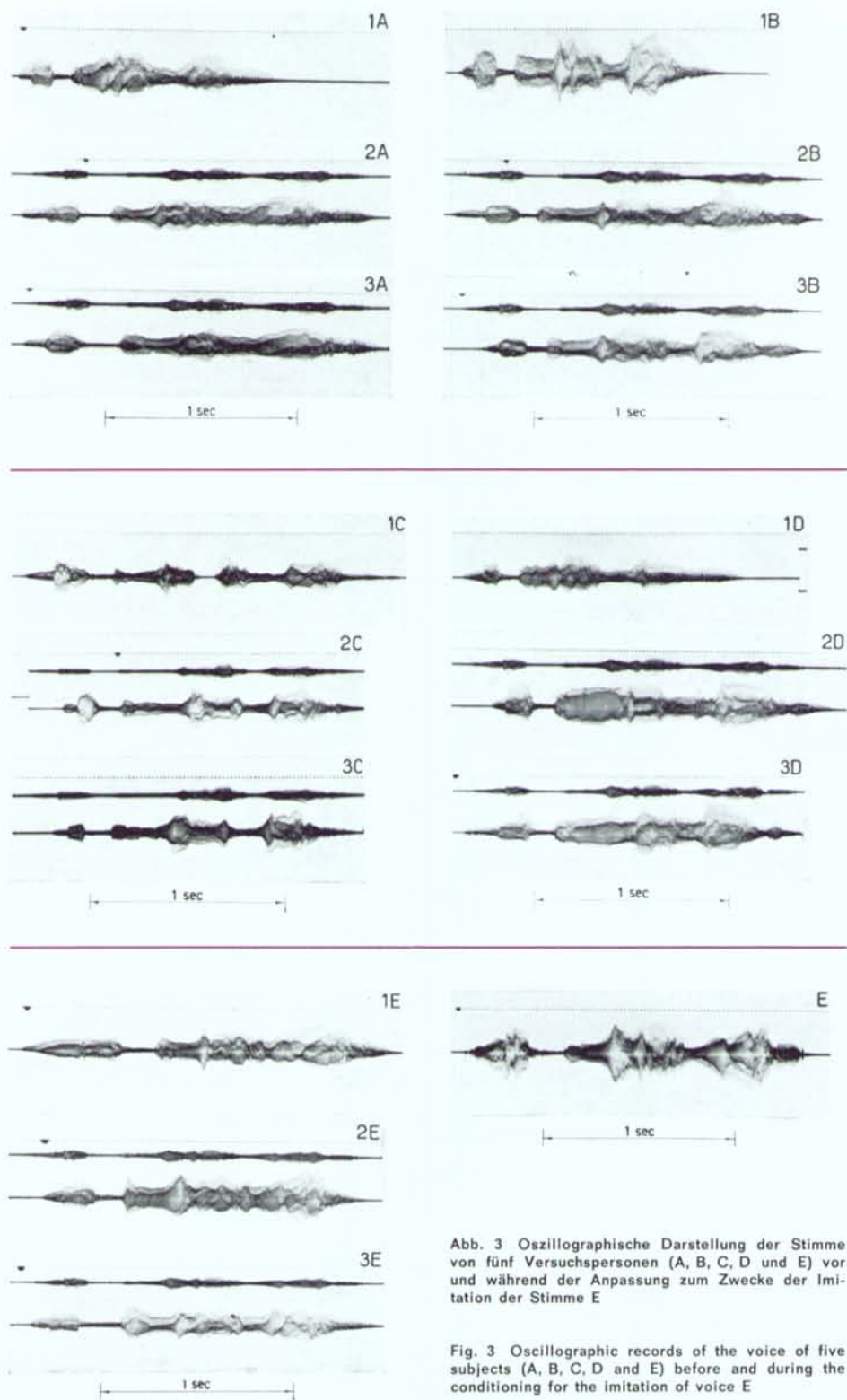


Abb. 3 Oszillographische Darstellung der Stimme von fünf Versuchspersonen (A, B, C, D und E) vor und während der Anpassung zum Zwecke der Imitation der Stimme E

Fig. 3 Oscillographic records of the voice of five subjects (A, B, C, D and E) before and during the conditioning for the imitation of voice E

ahmt, D recht gut und C ebenfalls gut, allerdings mit höherem Timbre, und daß E ohne Zweifel eine bessere Nachahmungsfähigkeit besitzt als die anderen.

Eine dritte Gruppe von Diagrammen (Abb. 3) zeigt die mit Hilfe einer Kathodenröhre ermittelten Oszillogramme der aufgenommenen Stimmen. Gruppe 1 dieser Oszillogramme zeigt die ursprüngliche Stimme der Versuchspersonen A, B, C, D und E, Gruppe 2 die jeweils imitierende Stimme nach 15 Sekunden und Gruppe 3 die jeweils imitierende Stimme nach 3 Minuten.

Die Oszillogramme 2 und 3 zeigen in ihrem oberen Teil die zu imitierende Stimme in einem reduzierten Bereich; die Oszillogramme E zeigen dieselbe Stimme im gleichen Bereich wie die der anderen Personen.

Betrachtet man die Oszillogramme, so vermitteln sie nur ein Gesamtbild des Imitationsphänomens, aber man kann sehr klar die Abweichungen einiger Elemente, z. B. des Rhythmus und des Timbre, von der Originalstimme beobachten. So kann man beispielsweise erkennen, daß die Endung von »le tue« sich nach einer gewissen Zeit mit der Endung der nachzunehmenden Stimme deckt.

Diese Beispiele wurden hier dargestellt, um die für die Untersuchung dieser Probleme angewandten Methoden zu beschreiben, die ausgedehnt werden sollen auf Untersuchungen audiologischer Natur. Darüber hinaus wäre es interessant, einmal die Beständigkeit der Stimme einer Versuchsperson zu untersuchen, die sich selbst nachahmt.

Wie bereits angedeutet wurde, soll in Zukunft über weitere Ergebnisse berichtet werden, die bei der Analyse der einfachen Abstimmung eines Lautes musikalischen Charakters beim Sprechen oder Singen gewonnen wurden, wenn dieser Laut beständig oder wenn er kontinuierlichen Veränderungen unterworfen ist.

Examining the oscillograms one has only an aggregate picture of the phenomenon of the imitation, but one can observe very clearly the deviation of some elements, as the rhythm and the timbre, from the original voice. It may be observed for an instance that the ending of "le tue" tends to be synchronous after some time with the same termination of the voice to be imitated.

These examples are reported here to describe the method employed to study this problem, that shall be extended to investigations of audiological character. It would also seem interesting to study on the steadiness of the voice of a subject, imitating itself.

As has already been pointed out, other results will be exhibited in a future time, that have been obtained by the analysis of simple tuning, by speech or sing, with a sound of musical character, when this is steady as well as when it changes continuously.

Literaturverzeichnis / Bibliography

- [1] G. U. Righini, A Pitch Extractor of the Voice, "Acustica", vol. 13, 1963, Nr. 4 and more detailed in "Alta Frequenza", 1964, p. 18.

Arista-HÖRERSCHNÜRE