# Experimental Phonology and Phonetics

**Katherine Morton**



This text/work book was first published in 1984. As a textbook the work is suitable for anyone interested in experimental phonetics, though of course many of the techniques are of historical interest only.

www.morton-tatham.co.uk – katherine_morton@morton-tatham.co.uk

This book was written in 1984, and the material in it had been used at Essex University as the basis for a number of elementary laboratory courses taught by me and Mark Tatham over a period of several years. Students should read the book now only with caution. Many of the principles and approaches are still valid, but much of the equipment and associated techniques are long since out of date. I think the material still has a value because it gives context to the published experimental work of the time. For example, several laboratories (including Haskins, UCLA and Essex) worked in the field of experimental electromyography in the 60s and 70s — this book describes the techniques those laboratories used, and sets out one or two of the classical experiments they performed. The same is true of experiments based on other techniques I included, but which are no longer used.

By the time I got around to thinking about formal publication of the book its contemporary value had been reduced by the change in experimental techniques — particularly the re-emergence in the 80s and 90s of a focus on acoustics work as a basis for automatic speech recognition and synthesis, and the rapid development of computer analysis of data. Web-based publication is intrinsically low cost, however, so the book is reproduced here for its historical value.

## INTRODUCTION

## PART I — INSTRUMENTATION

### Recording

- Recording Devices
- Direct Analog Magnetic Tape Recording
- Limitations of Magnetic Tape Recording
- Frequency Response
- Dynamic Range and Noise

### Transducers

- Microphones
- Electro-Manophone
- Aerometer
- Electroglottograph
- Photo-electroglottograph
- Accelerometer
- Electromyography
    - Detection of muscle fibre potential
    - Detection of the potential at the skin surface

### Visual Display

- Oscilloscopes
- Chart or Pen Recorders
- Inkjet Recorder (Mingograf)
- Ultraviolet Recorder
- Display Devices

## PART II — THE EXPERIMENTS

1. Calibration — The Mingograf
2. Vocal Cord Vibration
3. Voice Onset Time
4. Intraoral Air Pressure
5. Airflow
6. Amplitude and Stress
7. Duration
8. Electromyography — The Tense/Lax Controversy
9. Invariance
10. Stuttering

## APPENDIX A — Statistics

## APPENDIX B — Computing

## INTRODUCTION

This booklet accompanies your course in Experimental Phonology and Phonetics. For some of you the course parallels others in phonological and phonetic theory, others will have had an introduction to these earlier. The booklet is divided into two parts.

*PART I* deals with most of the equipment you will encounter. It is designed to provide a reference guide to particular pieces of equipment in the laboratory that you will either be using during the course or have demonstrated to you.

*PART II* takes you in some detail through a number of experiments. Not all groups will do all the experiments. The idea here is to set out the purpose of each experiment and how it ties in with other aspects of linguistics. The equipment to be used is discussed and the way the equipment is connected together set out. Often there is a sample of the kind of data you can expect the experiment to produce, together with some hints on measuring or evaluating that data. Brief sections deal with statistics and elementary computer processing of experimental phonetics data.

In some ways experimental phonetics, unlike experimental work in other areas of linguistics, is like experimental work in the physical sciences. For example, we can examine in detail the acoustic waveform of the sounds of speech or can examine the behavior of the musculature during articulation. The data we collect in such experiments reflects what actually happens when a speaker speaks, and we can call this *real world data*. Experimental work in, say, syntax or dialect studies is quite different. In these areas of linguistics we would use special techniques to elicit from an informant what he *feels* or *thinks about* language: we would be inquiring into his *intuition* as a speaker of the language. The contrast with phonetics is clear: here we are interested in discovering what the speaker *actually does* when he speaks, not what he feels or thinks about speech. Of course, people do have feelings about speech and it is important for linguistics to understand the mental, as well as the physical aspects of speaking. The study of the underlying mental processes involved in speaking is treated in linguistics for the most part under the heading phonology rather than phonetics. You should note though that the relatively new area of *Cognitive Phonetics*, developed here at Essex, tries to characterise some of the mental aspects of speaking which are not truly phonological.

Our interest in the physical aspects of speech leads us to formulate hypotheses about what a person does when he or she produces speech sounds. Hypotheses are rarely put together in a vacuum, however; they arise from our thinking as linguists in the context of the theory of phonetics and phonology. For example, phonologists know that a speaker of English feels that more articulatory ‘tension’ is involved in the production of [t] than in the production of [d], and label the former as [+tense] and the latter as [–tense]. The theory would predict that this feeling of greater or lesser tension has a *physical correlate* in speech production, and we might hypothesize that the musculature of certain articulators is actually held more tense for the one segment than the other. So, as experimentalists, we set out to discover in the laboratory whether we can find physical evidence to support the theoretically derived hypothesis. If we can, then we have provided support for the theory, but if we discover that the data reveals the opposite of our expectation, then we refute the hypothesis and our theory is thrown into question. This process of continuously generating hypotheses, designing a suitable experiment, gathering and evaluating data, supporting or refuting the hypotheses, and reinforcing or overthrowing part of the general theory is the usual way of proceeding with the development of a science, and is part of what is known as *scientific method*.

As we shall see, experimental phonology and phonetics are an important part of linguistics precisely because of their constant contact with the real world data we collect in the laboratory. This is a difficult part of linguistics, though, because it needs you to develop a feeling for experimental procedures which are somewhat different from the way you treat other areas of linguistics. But, as you will see, it is interesting because we constantly have to deal with difficult questions concerning the interaction between the abstract (or mental) and real (or physical) worlds.

## PART I — INSTRUMENTATION

### Recording

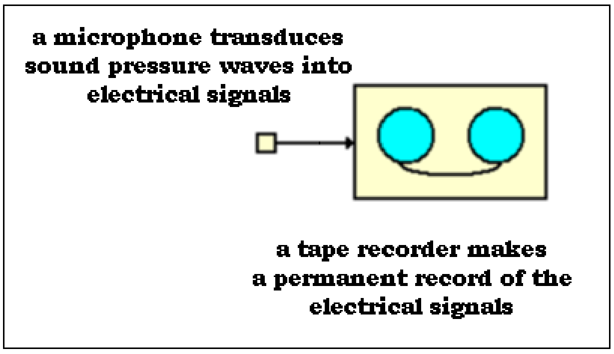
Speech is a *transitory phenomenon* whether we are considering the acoustic waveform, articulator dynamics or motor control of articulation: it operates in time. Indeed many of the linguistically significant elements or features of speech are changes from one acoustic or articulatory configuration to another. *Time* is an essential feature of speech. The rapid or transitory nature of speech events makes detailed scientific study difficult. One way around the problem is to record the speech event. Recording takes several forms, but basically we are concerned with capturing speech in a permanent and reproducible form, and with converting speech into some visual form. Because of the way in which human perception is organized it is easier for us to appreciate or analyze the details of speech if they are presented as a *visual transformation*. In other words: whether we are dealing with sound waves, articulatory movement, flow of air, neural signals, etc., we want to provide visual records of these parameters, and it is from these visual records that we make most of our scientific observations.

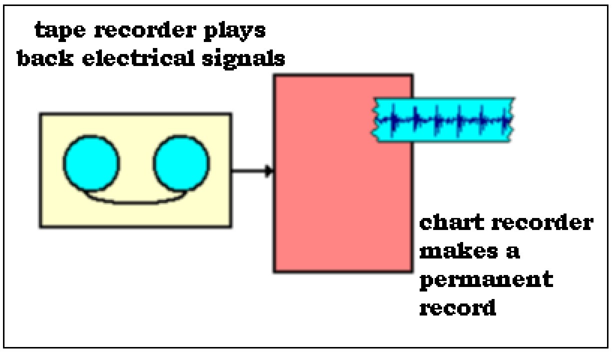
The first thing we do in the laboratory with any speech event is to *detect* it and *transduce* it into electrical energy. This is done in real time, or as the speech is actually happening. This electrical energy is then recorded in a permanent form such that it can be recovered at will on some future occasion. Later, on recovery, the electrical energy which is a transformed version of the original speech is itself transformed into some visual presentation — usually permanently on paper.

A visual record entails the use of one axis of a graph to represent time. The eye of the researcher is thus enabled to scan or go backwards and forwards over the speech in a way which would be impossible just by listening to someone speaking. Transducing, recording and subsequent visual presentation may be diagrammed like this:

*speech → permanent recording → visual recording or presentation*

Or, a specific example:



Figs. 1 and 2 — Making a tape recording of a microphone signal, and then using a chart recorder to make a visual record of the signal.

We shall deal with the components in this chain under three headings: Recording Devices, Visual Display, and Transducers.

## Recording Devices

The most common method of recording used in phonology and phonetics experiments is *direct magnetic tape recording*. We shall deal with this in some detail and summarize the main points of other methods of recording afterwards. In all cases, though, remember that we are not making a direct recording of sounds, or airflow, or articulator movement, etc., but of electrical signals which are transformations of these phenomena derived from a transducer specifically designed for the purpose.

### Direct Analog Magnetic Tape Recording

In most cases tiny particles of an oxide of iron adhered to tape (usually one eighth or one quarter of an inch wide) made of relatively non-elastic plastic film are magnetized differentially into magnetic patterns. Electrical signals from an appropriate *transducer* are fed to a non-permanent electromagnet, called a *recording head*, causing a magnetic field to be set up around the magnet. This field varies in sympathy with the electrical signals being applied to the head. The tape covered with magnetic oxide is passed steadily over the head and a magnetic pattern is recorded in the oxide as the head varies the magnetic field around it. Thus:

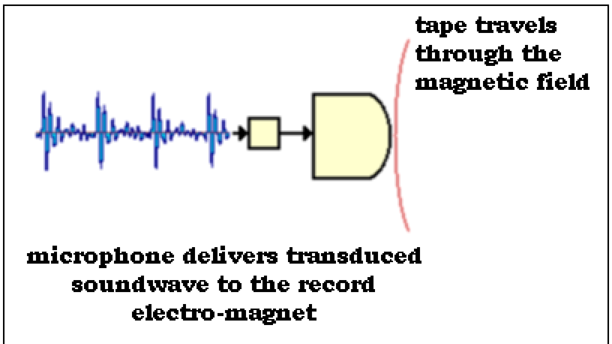



Fig. 3 — The process of recording a sound wave onto magnetic tape.

The difference between a blank tape and a recorded tape is that the oxide particles on the blank tape are randomly magnetized, whereas on the recorded tape they are magnetized into patterns which correspond to the electrical signals from the transducer. The magnetic patterns in the tape's oxide are semi-permanent. They can be erased by passing them through a very strong random magnetic field which will remove the patterns and recreate a random magnetization of the oxide particles ready for reuse.

To play back a recording, the tape is passed in front of a *reproduce head*, which works roughly in reverse to a record head. This time, magnetic fields are introduced into the head from the magnetic particles on the tape and in the corresponding patterns. These are then amplified and processed:

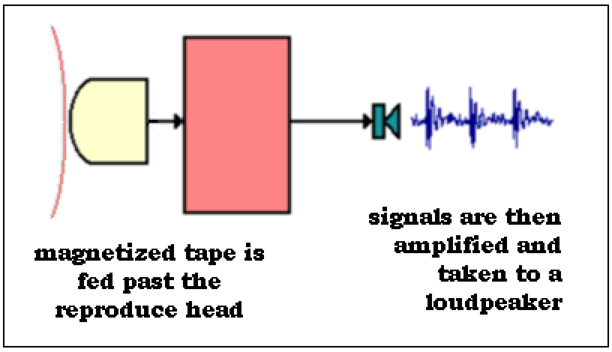



Fig. 4 — Playback of a recording *via* an amplifier and loudspeaker to produce a sound wave.

### Limitations of Magnetic Tape Recording

All recording systems have their limitations. We can only make good use of these systems if we understand what these limitations are and how they effect our interpretation of the data. Once the limitations are understood they can either be avoided or taken into account in our interpretations. With direct analog tape recording, of the kind just described, there are two major limitations which concern us: *frequency response* and *dynamic range*.

### Frequency Response

The acoustic signals, which have been electrically transduced, are constantly varying in time. The slowest and fastest varying elements of the signal denote the frequency range which is being transduced. Recording systems also have their own *intrinsic* frequency ranges — that is, a range between the slowest and fastest electrical signals to which they can respond accurately. Beyond the frequency range of a tape recorder no signal, even if present in the incoming signal, can be properly recorded. Clearly we must make certain that the frequency range of the recorder matches, or is wider than the signal we want to record.

Recording the full width of the frequency range of the incoming signal is not enough however. Frequencies at any point in the range must be equally efficiently recorded to preserve the correct relationships between them. Suppose we had a low frequency tone of moderate amplitude and a mid-frequency tone of somewhat lower amplitude to transduce and record. If the response of the tape recorder was more efficient in the mid-range it might well record the mid-tone with greater amplitude than the low tone, thus distorting the data on playback. This is why when you see the frequency response of a recorder quoted there are always two parameters: frequency range (for example, 50Hz–l5kHz), and amplitude response over that range (for example, ±2dB). This means that any tone in the range 50 to 15,000Hz can be recorded with an amplitude accuracy varying no more than 2dB above or below the actual amplitude present in the incoming signal. Thus all tone amplitude relationships will be preserved accurately within 4dB of each other which is usually accurate enough for our purposes.

Frequency response is often given in graph form:

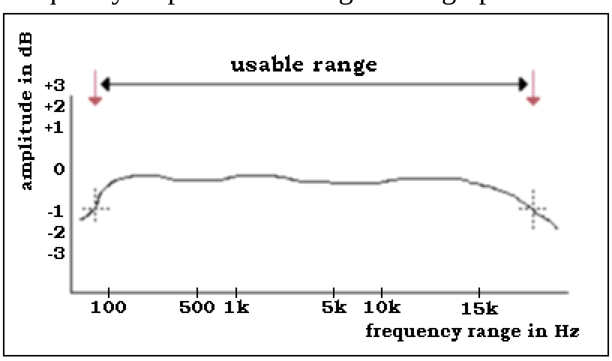



Fig. 5 — Typical frequency response graph for a tape recorder.

Frequency response in analog magnetic tape recording is dependent for the most part on how many magnetic particles travel past the recording head in a given time. There are two ways of influencing this:

- increase or decrease the number of particles per centimeter of the tape;
- increase or decrease the speed at which the tape travels past the heads.

In practice, since tapes are usually made with the maximum possible particle density, tape speed is the most important variable. Running the tape fast increases the number of particles passing the heads in a given period of time (say, 1s). The greater the speed the wider will be the *frequency range* and the more detailed and accurate the magnetic pattern recorded.

Here are some typical frequency responses for different types of tape recorder:

- cassette (4.76 cm/s) — 45Hz–12kHz ±3dB
- open reel (19 cm/s) — 35Hz–20kHz ±2dB
- open reel (38 cm/s) — 30Hz–25kHz ±2dB

The smaller the deviation value (expressed in decibels — dB) the greater the *recording accuracy*. The greater amplitude accuracy (2dB or better, as opposed to 3dB) of the open reel tape recorder reflects the increased width of the tape and therefore the 'track' of recorded magnetic patterns, among other things.

The sound waves of human speech rarely cover a range greater than 60Hz to 12kHz. This is within the capabilities of most well designed tape recorders. [Music can have a range 30Hz–20kHz, and human hearing is on average 20Hz–18kHz.]

### Dynamic Range and Noise

Dynamic range is the range between the lowest and highest amplitude signals. Once again there is, on the one hand, the dynamic range of the signal to be recorded, and, on the other, the *intrinsic dynamic range* of the recorder being used: these must be matched to achieve the most accurate recording.

The low end of the dynamic range scale of the tape recording system is not zero amplitude. In practice there is always some amplitude present even when a randomly magnetized blank tape is played. We obviously cannot record a signal of lower amplitude than this noise, or the noise itself will obliterate or mask it. At the other end of the range there comes a point where the ability of the oxide particles to accept further magnetization becomes saturated and further increase in amplitude results in severe distortion which would be unacceptable in our experiments. The dynamic range of a tape recorder is the range between the noise level (or *floor*) and the saturation level (or *ceiling*), and is expressed as a range of decibels. Thus:

- cheap cassette recorder — 35dB
- good cassette recorder — 50dB
- good open reel recorder — 65dB
- human speech — 45dB
- symphony orchestra — 110dB

Note that a good tape recorder can capture the dynamic range of human speech usually without much difficulty, but the same tape recorder makes a relatively poor job of music recording. That is one reason why recorded music does not sound quite as dramatic as it does in the concert hall.

When making a recording of speech, note that the record *gain control* must be adjusted to keep the signal's dynamic range within that of the tape recorder. Zero on the meters corresponds to an acceptable level of saturation. Aim never to exceed this, and the bottom end will take care of itself since the dynamic range of speech is usually narrower than that of the tape recorder.

But what about signals whose frequency range falls outside the range of a tape recorder? Many signals other than sound waves fall outside the range — particularly at the low end, often going down as far as 0Hz. In these cases we need to use a special form of recording, known as *frequency modulated recording* (or FM recording), and this will be used to record such things as *airflow* and *air pressure*, both of which often vary very slowly at a very low frequency. In frequency modulated recording the incoming signals are transformed or encoded into frequencies which the tape recorder can handle. On playback the recorded frequencies are transformed back or decoded by mirror image circuitry to restore the original signal. Typically a FM tape recorder can handle frequencies in the range 0Hz–2.5kHz accurately. It is important to know, therefore, the frequency range of your data in order to choose an appropriate recorder for making the permanent record.

## Transducers

Transducers are devices which convert one form of *energy* into another. In the phonetics laboratory we usually convert all data into *electrical* energy for easy recording and visual display. The data can take many different forms, though. For example: sound pressure waves, airflow, air pressure, electro-physiological signals, etc.

### Microphones

Microphones are transducers which transform *sound pressure waves* into *electrical signals*. There are several different types, but usually they convert pressure variations into deflections of a diaphragm, resulting in a voltage whose rate of change reflects the frequency of the sound, and whose amplitude corresponds to the amplitude of the sound.

Microphones have inherent frequency responses and dynamic ranges — but although care has to be taken when recording difficult signals (e.g. music) almost all microphones will be suitable for speech recording on these two parameters. Microphones are, however, directional, ranging from omnidirectional to unidirectional. Care must be taken, for example, when recording under conditions of high ambient noise or in a reverberant room to use a unidirectional microphone. This is because its narrow ‘view’ of the signal tends to exclude reverberation or extraneous noise originating outside the angle of view.

### Electro-Manophone

The electro-manophone is a kind of special purpose microphone used for sensing *pressure changes* in the air right down to a steady pressure which is not changing (i.e. a change of 0Hz). The device is very sensitive since in speaking very low air pressures are involved. For sensing, say, intraoral air pressure, a flexible plastic tube is connected to the transducer and inserted into the front of the mouth. As the pressure varies in the mouth so the pressure of air in the tube changes, and this pressure variation is transmitted along the tube to the sensing diaphragm of the manophone outside the mouth. Once again, electrical signals are derived for treatment in the usual way. The frequency range of the electro-manophone is 0–1kHz, so *FM recording* must be used for a permanent magnetic record of the output signal.

In your experiments you will be using the electro-manophone to sense the amount of air pressure built up behind the stop in plosives. By inserting the tube between the lips and teeth into the mouth above the centre of the tongue the pressure build-up for bilabial and dental or alveolar stops ([p, b, t, d]) can readily be sensed. The pressure build-up behind velar stops cannot easily be sensed using this method because the site of the stop is too far back in the oral cavity. In addition it is obviously quite easy for the tongue to touch the end of the tube, blocking it and making nonsense of the pressure reading; a similar incorrect reading can occur if the tube gets saliva in it, and care must be taken to avoid this happening.

### Aerometer

The aerometer detects and transduces *airflow* into electrical signals. There are four transducers in the laboratory’s aerometer, one each for: mouth in, mouth out, nose in, nose out. The transducers are mounted in a face mask which has a horizontal dividing curtain of rubber separating nasal flow from oral flow. A surround of foam rubber prevents air leakage from the mask. Each transducer consists of a rubber valve which opens as air flows through, the amount of opening being proportional to the airflow. On one side of the valve is a tiny lamp, and on the other side a *photoelectric cell*. The more the valve opens, because of increased airflow, the more light can pass through the valve and fall on the photoelectric cell. Photoelectric cells have the property of being able to produce a voltage proportional to the amount of light falling on them.

So: the more airflow, the wider the valve opens, the more light passes through to fall on the photoelectric cell and the more voltage produced. The relatively small electrical signals coming from the transducers are amplified in the circuitry contained in the box connected to the mask. The frequency range of the system is 0–500Hz ±3dB — so for permanent magnetic recording a FM system is necessary.

### Electroglottograph

The electroglottograph is a device for observing the vibrating vocal cords by electrically transducing the area function of the glottal opening — the varying space between the vocal cords as they open and close. Two electrodes are positioned one on either side of the neck in the area of the larynx. As the vocal cords open and close, in effect an electrical signal flowing between the electrodes varies in strength proportionally to the degree of vocal cord opening. This is referred to as the area function of the glottis. This varying signal is amplified for recording and displaying purposes. The control panel of the equipment has a control for sensitivity, since the signal varies considerably in strength for different subjects. The electroglottograph is able also to detect the very low frequency movements (up and down) of the entire larynx. So, although vocal cord vibration itself rarely falls below 50Hz in frequency, the device has an overall frequency response of 0–300Hz necessitating FM recording of the output signal.

Strictly, the electrical signal produced by the electroglottograph (and the *photo-electroglottograph* described below) corresponds to the varying area function of the *glottis*. Conveniently, but only as a first and rough approximation, it turns out that mathematically this varying area function matches the acoustic waveform of the sound produced by the vocal cords as they open and close. It is normal, then, in phonetics experiments to treat the output signal of the electroglottograph as equally representing either the area function or the acoustic waveform itself. You can check this by amplifying the signal and feeding it to a loudspeaker: you will be able to hear the ‘sound’ produced by the larynx without the subsequent filtering produced by the resonance characteristics of the cavities above the larynx.

### Photo-electroglottograph

This is another device for detecting and electrically transducing the area of opening between the vocal cords. A very bright light derived from a powerful quartz-iodine source (a film projector lamp) is focussed using a light pipe onto the neck wall below the larynx. The light is bright enough to diffuse into and illuminate the *trachea*. As the vocal cords open and close so varying amounts of light, proportional to the degree of opening of the vocal cords, pass through and into the *pharynx* above the larynx. A tiny *photoelectric cell*, attached to wires and suspended in the pharynx via the nasal opening and cavity, detects the varying amount of light, and sends back to the amplifier a proportional voltage varying with changes in the amount of light as the vocal cords open and close. Low frequency elements in the signal (as for example when the vocal cords are held open during breathing) require the use of the FM recorder for permanent magnetic recording of the signals. Frequency response is 0–500Hz ±3dB.

### Accelerometer

The accelerometer is a device developed to detect *acceleration* or *deceleration* (changes in rate of movement) of an object. It is used in the phonetics laboratory to detect, for example, changes in *neck wall vibration* as the vocal cords vibrate within the larynx, or vibration of the outside wall of the nose if the air column in the nasal cavity is in resonance because a nasal sound is being produced. The electrical output of the transducer is extremely small and is locally pre-amplified to bring the signal up to standard levels of amplitude for acceptance by a normal tape recorder amplifier or display device.

### Electromyography

Electromyography is a technique for detecting and recording or displaying for measurement the minute electrical signals generated within a muscle when that muscle receives a *motor command* for contraction.

#### *a. Detection of muscle fibre potential*

There are three methods commonly used for detecting the muscle fibre potential (the electrical signal). These subdivide into two main approaches:

- detecting the signal within the muscle, and
- detecting it at the skin surface.

There are two electrode methods used for detecting signals within the muscle:

- needle electrodes,
- hooked wire electrodes.

With needle electrodes a *concentric bipolar electrode* is formed from a small hypodermic needle. The canula of the needle forms one electrode, and the second consists of an insulated wire held in position within the canula by a bonding resin. The wire is cut level at the end of the needle. With hooked wire electrodes two insulated copper wires with bared tips are inserted side-by-side into the muscle. This is done by firstly preparing two 6-inch lengths of fine copper insulated wire with bared tips, passing these through a medium sized hollow hypodermic needle and bending over the ends of the wires to form hooks. The needle, together with the wires, is then inserted into the muscle, and then gently withdrawn. The wires remain in the muscle anchored by the hooks. The wires can be withdrawn after the experiment by a gentle tug.

With both needle and hooked wire electrodes sterilization is essential. The usual method is to boil the electrodes for twenty minutes or so in water before insertion after cooling. Insertion is not painful if the needle is sharp, but often a topical anaesthetic of weak xylocaine, which is effective for around one minute, is applied to the skin surface. Any pain is felt only on the surface and after insertion this disappears.

#### *b. Detection of the potential at the skin surface*

The *electrical potential* developed from contracting muscle fibres is sufficient to diffuse outside the muscle body, and if the muscle is just below the surface it can be detected using *surface electrodes*. These are constructed of small discs approximately 2mm in diameter, punched from thin sheet silver or platinum. A slight indentation is formed in the disc, and a wire is attached by soldering to the domed side. Two electrodes are applied to the skin surface, separated approximately 2mm, and held in position by a special very elastic, but very sticky, surgical tape called Blenderm. Before placement the indentations on the electrodes are filled with a gel containing *sodium chloride* (common salt) which improves electrical conductivity. The skin surface is shaved, abraded slightly to remove dead skin cells, and thoroughly cleaned with alcohol on a pad of cotton wool; all these measures are to ensure the closest possible contact between the electrodes and the skin surface.

*Intramuscular electrodes* are used to detect potentials in tiny muscles or muscles which are not immediately below the skin surface. They have a tendency, though, to pick up signals from only a small number of *muscle fibres*, and the behavior of these fibres may not be indicative of the behavior of the muscle as a whole. Surface electrodes, on the other hand, can only be used with larger muscles immediately below the skin surface, and 'see' a relatively large number of muscle fibres, thus providing a signal which is more indicative of whole muscle behavior. We use surface electrodes whenever possible in our laboratory, partly because of this property of reflecting better the behavior of an entire muscle, but also because their use is not invasive (i.e. does not penetrate the subject's skin surface).

The signals derived by the electrodes are extremely weak. They range from 5–50 micro-volts or millionths of a volt from intramuscular electrodes to 15–200 micro-volts from surface electrodes. Careful and correct amplification is essential. The biggest problem with amplification is unwanted *noise*. Human beings act like radio antennae picking up interference from broadcast radio transmissions or even from the lighting and electrical wiring in the laboratory. This interfering and unwanted signal can often well exceed by several orders of magnitude the emg signal itself. Special amplifiers, called *differential amplifiers*, are used to help eliminate the unwanted interference signals and to provide the high gain necessary (up to 1v) before the signal can be adequately recorded. To drive a differential amplifier a further signal besides that from the electrodes is required. This is usually obtained from a surface electrode on some part of the body away from the muscle being investigated, and away from any contracting musculature (especially the heart). This electrode will be picking up the unwanted interference signal, diffused throughout the body, which we'll call X. So, the emg electrodes have two signals (emg and X) and the reference electrode has one signal (X). The differential amplifier operates by subtracting the reference signal from the emg electrode signals. Thus:

(emg + X) – X = emg

[i.e. subtract X from the emg which includes X and you are left with the emg]

This leaves only the tiny emg signal, which goes on to be amplified. Unwanted noise is thus eliminated.

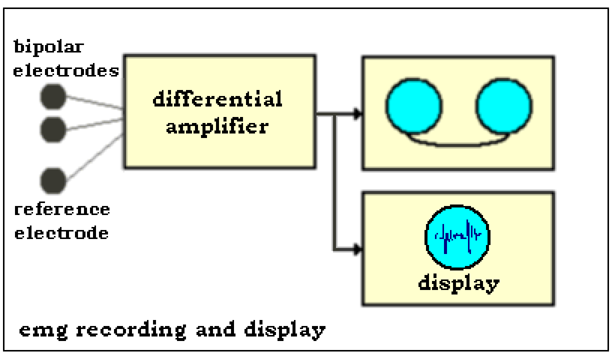



Fig. 6 — Equipment arrangement for detecting, recording and displaying electromyography signals from surface electrodes.

Emg signals can be recorded just like other signals produced in the laboratory. They range from 50–300Hz, and can easily be accommodated therefore on an ordinary direct record tape recorder. However, sometimes, if badly applied or after long periods of use, the electrodes may move, producing *artifacts* which influence the emg signal. These electrode movement signals, as they are called are of low frequency (a few Hz only). We normally deliberately record these low frequency signals by using FM recording so that we always know when an artifact is present. Direct recording would eliminate the artifact, leaving its effects only, which we would not be readily able to spot. 50–300Hz is within the audio range. Try listening to the signals via the loudspeaker of the tape recorder or through an external amplifier and loudspeaker system.

## Visual Display

Visual display of our transduced electrical signals is essential if we are to observe and measure what is going on in those signals. This section looks at a few different methods of visual display, each with their uses in the laboratory.

### Oscilloscopes

The oscilloscope is a device for presenting a visual display of an electrical signal in real time. That is, the display changes with the signal, and disappears with the signal. Unless the display is photographed the oscilloscope does *not* supply a permanent record.

The display is achieved on a *phosphorescent screen* at the end of a *cathode ray tube* (CRT). At the narrow end of the tube a coil driven by a high voltage generates a stream of electrons which is electrically focussed into a beam using magnets:

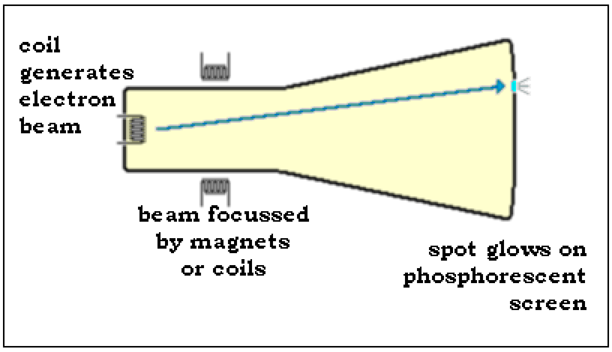



Fig. 7 — Schematic of a cathode ray tube oscilloscope.

When the beam of electrons hits the screen the phosphorescent material on the inner side of the glass glows, transforming the electrical energy into light. In the Essex Speech Laboratory oscilloscope the phosphorescent material glows blue and has a yellow *afterglow* — that is, it continues to glow for a couple of seconds in a different color when the electron beam has passed on elsewhere on the screen. This gives the image a *persistence* which is helpful in examining fast changing signals.

Magnetic coils on either side of the tube's neck attract the beam from side to side as it travels towards the screen. [Sometimes plates within the tube's neck are used instead of coils to perform the same beam deflection function.] The rate at which this happens can be varied from very slow to very fast by changing the rate at which the coils alternately attract the beam. This is called the timebase, because it is the left-to-right spatial axis (the *x–axis*) of the display which is used to represent time. The coils or plates are called *deflection coils* or *deflection plates.*

Magnetic coils above and below the tube's neck deflect the beam upwards or downwards from the centre of the screen. The signal is applied to these coils from one of our transducers. So, varying voltages from a transducer result in an upward-downward deflection of the beam, and an internally generated timebase gives left-to-right representation of time. It is important to remember that the horizontal axis (the *x–axis*) is time, and is internally generated, whereas the vertical axis (the *y–axis*) is from the transducer. A third axis (the *z–axis*) of brightness is generally not used in the Speech Laboratory (but is used, for example, in the presentation of a television image on the phosphorescent screen because these images contain a varying brightness element).

Although normally oscilloscopes do not provide a permanent record except when the screen is photographed, it is possible to hold a display on the screen after the event has gone. This is done in the laboratory by putting the signal into a computer-like memory and repeatedly presenting it to the y–axis deflection magnets every time the screen is scanned from left to right by the timebase. This gives the illusion of a steady unchanging display. Such an arrangement is called a *digital storage oscilloscope* — digital because the memory operates digitally, and storage because the normally transient signal is held.

Because the oscilloscope is wholly electronic involving nothing mechanical the system is relatively inertia free and permits a frequency response of zero to many mHz. In other words its display can be considered absolutely accurate for handling the signals we get in the phonetics lab.

### Chart or Pen Recorders

Pen recorders make a permanent record with ink on a strip of moving *chart paper*. A roll of paper is made to unwind at a certain speed (typically we use 100mm/s in the Speech Laboratory) and pass beneath a felt-tipped pen which is held against the paper. Magnetic coils on either side of a magnetically sensitive armature holding the pen deflect it. Thus, an x–axis is drawn by the motion of the paper, and corresponds to time. A y–axis is created by deflecting the pen armature by applying signals from transducers to the deflecting electromagnets.

Pen recorders are extremely limited in *frequency response* because of the *inertia* of the pivot holding the pen armature and the enormous friction between pen and paper. Typically the usable frequency response is 0–100Hz, and not really suitable for work in the phonetics laboratory. They are, however, cheap to buy and run, and can be of some restricted use if employed properly and provided the frequency response limitation is kept in mind.

### Inkjet Recorder (Mingograf)

The inkjet recorder also records with *ink on paper,* but does not use a pen. Instead, the ink is sprayed onto the moving paper using a miniature spray gun mounted on a shaft which rotates from side to side under the influence of electromagnets carrying the y–axis signal, arranged about the shaft. The frequency response of a typical inkjet recorder, such as the Mingograf in the Essex Speech Laboratory, is 0–700Hz ±3dB, with a x–axis accuracy (paper speed accuracy) of around 5%.

The Mingograf is the most heavily used means of providing *permanent paper records* of signals in the laboratory despite its apparent limitations in frequency response. It is very reliable and extremely cheap to run, producing a permanently clear and easily measurable record. In practice, the frequency response limitation is not too important as you will see in the various experiments described later, which use the Mingograf.

### Ultraviolet Recorder

To further reduce friction and inertia in the galvanometer ink is replaced in the ultraviolet (UV) recorder by a fine beam of *ultraviolet light*. The beam is focussed by means of a series of lenses such that it reflects off a tiny *mirror* attached to the shaft of a galvanometer. Side-to-side movement of the galvanometer is under the control of y–axis driven electromagnets.

The reflected beam of UV light falls onto unrolling photo-sensitive paper. The image appears on the paper with exposure to bright light for a few seconds. This image is not stable and disappears altogether on continued exposure to bright light, but may be fixed by chemical developing and fixing, or simply by avoiding prolonged exposure to bright light. UV-sensitive paper is expensive compared with the ordinary paper used in pen or inkjet recorders, but the gain is a wide frequency response, typically 0–5kHz ±3dB, but easily extendable up to, say, 20kHz with carefully designed galvanometers.

## Display Devices

All the display devices mentioned (oscilloscope, pen-recorder, inkjet recorder, UV recorder) are able to handle more than one channel of data simultaneously. This facility becomes useful for side-by-side comparison of different types of data obtained simultaneously from a subject while talking. Thus, for example, channel A may be displaying the acoustic waveform of, while channel B may be displaying intraoral air pressure. Synchronization of the two data channels on, firstly the tape recording and later on the display, permits comparison of the relative timing of significant events on the two channels of data, The laboratory's direct analog tape recorders each have two channels, the FM recorder seven channels, the oscilloscope two channels and the chart recorders (inkjet and UV) each four channels.

## PART II — THE EXPERIMENTS

### 1. Calibration — The Mingograf

Calibration is very important because the characteristics of any piece of electrical or mechanical apparatus can vary. To make accurate observations in experiments the researcher must be fully aware of his equipment's characteristics.

#### Equipment required

- Mingograf
- sine-wave generator
- voltmeter (digital display)

#### Setup

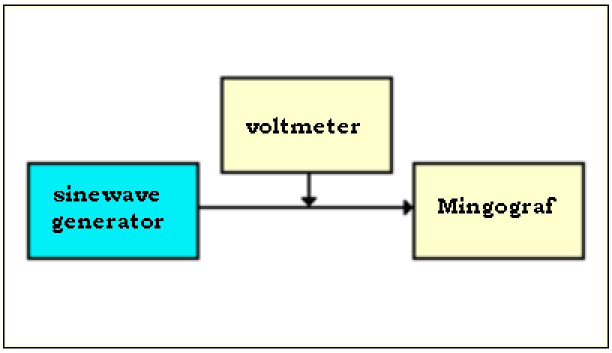



Fig. 8 — Block diagram of the equipment arrangement for calibrating the Mingograf.

#### Procedure

Set the output control (voltage) of the generator to a simple level (say, 1V), monitored off the volt meter. During the experiment monitor this constantly and adjust the output control of the generator such that 1V is always being delivered.

Set the frequency control of the generator to the lowest possible frequency (say, 20Hz), and adjust the Mingograf gain controls for a trace width of, say, 3cm. There are two controls for gain:

- a gross, switched control, and
- a continuously variable fine control.

Adjust the trace position control for a suitable placement of the trace on the chart paper. Do not touch the Mingograf controls again during the entire calibration procedure. *[Q: Why not?]*

Run the Mingograf at, say, 100mm/s, noting the *width of the trace* with the 20Hz sine wave. Now, keeping the *voltage output* from the generator constant, change the *frequency* to, say, 100Hz, and run the Mingograf again for, say, 250mm of chart. Repeat the procedure, using different frequencies up to around 1500Hz. Draw up a *table of results*. Here is a typical sample table — yours may be slightly different:

| Frequency of input signal (Hz) | Trace width for standard input voltage (cm) |
|---|---|

| | |
|---|---|
| 100 | 3.1 |
| 200 | 3 |
| 300 | 2.9 |
| 400 | 3 |
| 500 | 3 |
| 600 | 2.9 |
| 700 | 2.5 |
| 800 | 2 |
| 900 | 1.5 |

Now, plot a graph of the results table. Here is the graph of the above results:

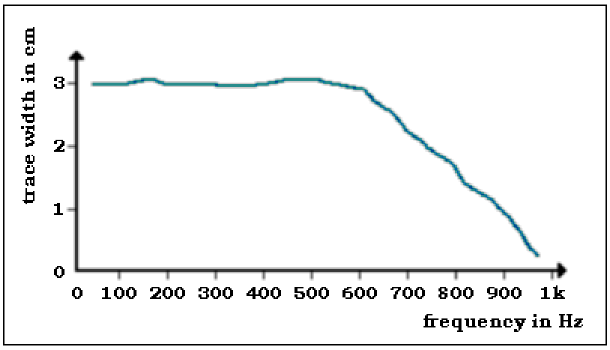



Fig. 9 — Typical frequency response graph for the Mingograf.

This graph is a *frequency response curve*. We are interested in the *frequency range* over which the galvanometer can maintain an accurate amplitude of the displayed signal. Our input signal was always 1V. In spatial terms we set the equipment to make 3cm equivalent to 1V. But note that as the frequency exceeded 600Hz, even though a 1V input was being maintained, 3cm width of trace was not being drawn. That is, *amplitude distortion* arose: the galvanometer no longer gave a reliable representation of the amplitude of the incoming signal. Reasonable amplitude accuracy was maintained only over the region 0–600Hz; any signals recorded above that range are unreliable with respect to the amplitude parameter. The *inertia and friction* in the galvanometer system make it unable to respond accurately outside the 0–600Hz range.

## 2. Vocal Cord Vibration

Vocal cord vibration studies form a major part of laboratory work in phonetics, since this is the primary source of sound in speech. The nature of the waveform, the on/off timing of vibration, the general relationship with the phonological feature [voice], etc., all form major areas of experimental investigation.

*Equipment*

- Electroglottograph

- Oscilloscope
- Mingograf
- tape recorder

*Setup*

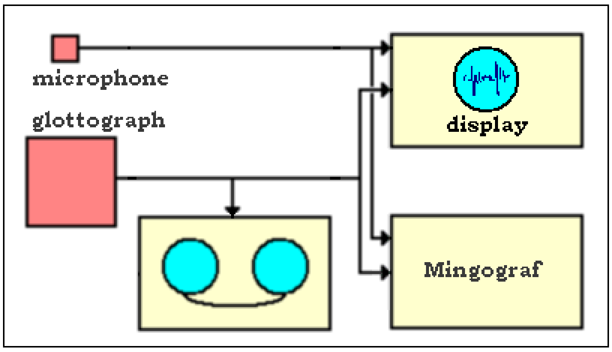



Fig. 10 — Equipment arrangement for examining vocal cord vibration.

*Demonstrations*

1. **Monotone versions of different vowels**. Note that the waveform should be the same, no matter which vowel is being attempted, since the quality of a vowel is not dependent on the source sound produced at the vocal cords but on the independent resonance properties of the supra-glottal cavities. Compare the waveforms from the vocal cords during different vowels with those from a microphone: the microphone waveforms will clearly differ — indicating that the differences between vowels are added after the larynx tone is produced.
2. **Pitch changes**. Notice that with a rise in pitch the periods of the waveform (that is, the repetition rate of the near-identical portions of the waveform) get closer together or occur more close spaced in time. When the pitch falls then the spacing increases.
3. **Connected speech**. Notice the apparent continuity of the waveform, and that it pauses only during portions of voiceless consonants.
4. **Whisper.** There should be no apparent vocal cord vibration on the oscilloscope screen during whisper. Whisper is done by producing a fricative type of sound using tense vocal cords rather than by causing them to vibrate as they do in the voiced portions of normal speech.
5. **Listening**. Record the above and play them back to hear what the vocal cord vibration sounds like without oral cavity resonance. Note the similarity of sound between the different vowels and during the voiced consonants.

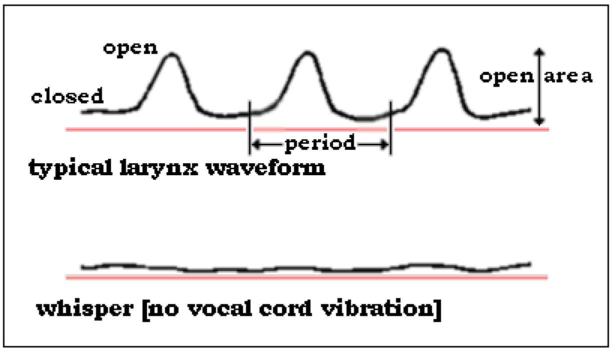



Fig. 11 — a typical larynx waveform during voicing and the corresponding signal during whisper.

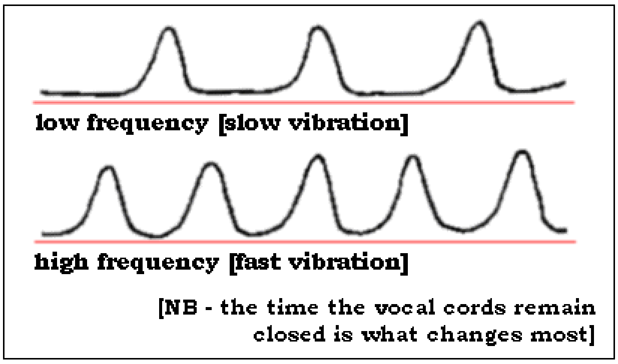



Fig. 12 — larynx waveforms for low frequency and high frequency vibrations. Notice that the amount of time the vocal cords remain closed is what changes most.

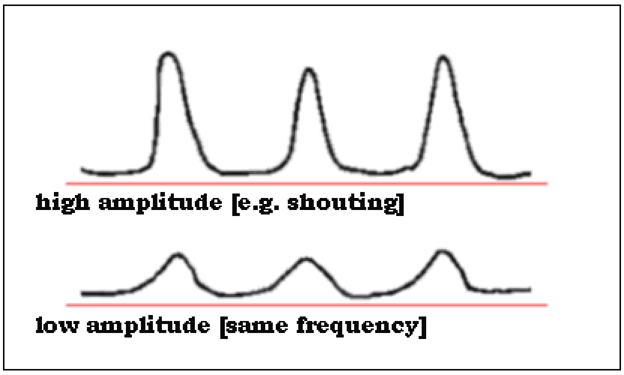



Fig. 13 — high amplitude and low amplitude larynx waveforms compared.

### 3. Voice Onset Time (VOT)

The term Voice Onset Time refers to the timing of the beginning of vocal cord vibration in CV sequences relative to the timing of the consonant release. The theory proposes that this timing is critical for accurate perception of the voiced/voiceless phonological contrast between consonants. This holds for *C* = *stop* and to some extent for *C* = *fricative* — that is, when the consonant is a plosive and often when the consonant is a fricative.

In *traditional phonetics* the focus of attention was on the sound or the airflow which filled the voice onset time. This was usually referred to as *aspiration*.

In *phonology*, consonants are contrasted and differentiated on the [voice] feature. Thus /b, d, g/ are marked [+voice] and /p, t, k/ are marked [–voice] in the phonology. The abstract nature of the phonology implies firm boundaries to the segment, and a straightforward conversion from abstract to concrete (mental to physical) as the phonetics realizes the phonology. Thus:

| | C | V | |
|---|---|---|---|
| consonantal | + | – | Implies that for the duration of the stop consonants there is or is not voicing — where voicing is realized as vocal cord vibration |
| vocalic | – | + | |
| stop | + | – | |
| voice | +/– | + | |

In fact, it can be shown that in English and many other languages, vocal cord vibration does not occur throughout a [+voice] stop, and that it does not begin simultaneously with the beginning of the vowel following a [–voice] stop. Typically the phonetic realisation is as follows:

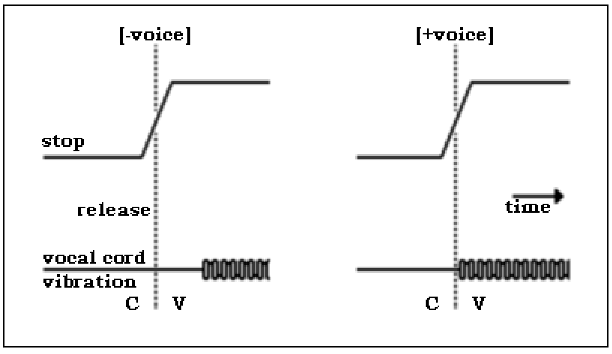



Fig. 14 — Comparison of the different stop releases for a voiceless and a voiced plosive. Notice the delay in vocal cord vibration following the release of the voiceless stop.

VOT may be *negative, zero* or *positive*. Zero indicates that vocal cord vibration has begun simultaneously with the release of the plosive consonant; negative indicates vibration beginning earlier than the release; positive indicates vibration beginning after the release. Different languages have different methods of phonetic realisation of this phonological feature. Thus:

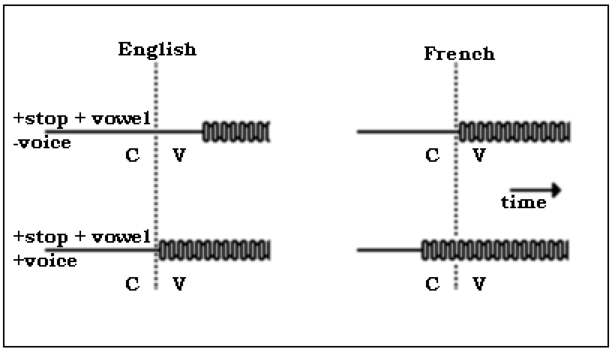



Fig. 15 — Comparing the voiced and voiceless stops in English and French. Notice although there are two pairs of plosive (two for English and two for French) these are in fact realised as only three different signals: the waveform in the French voiceless instance is the same as that for the English voiced instance.

Notice that zero VOT in French cues the perception of a [–voice] stop, whereas in English the same auditory cue indicates a [+voice] stop.

### Experiment — VOT

This VOT experiment has two aims:

- to verify the above observations;
- to determine typical VOT times for CV sequences in English and French.

*Equipment*

- microphone
- electroglottograph
- Mingograf
- tape recorder for a permanent record

*Setup*

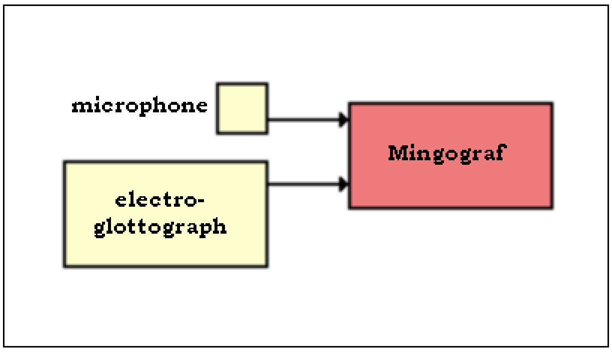



Fig. 16 — Arrangement of equipment for voice onset time experiments.

*Procedure*

Record vocal cord vibration and microphone signal for an English speaker saying the following each 15 times:

| | |
|---|---|
| /pa/ | /ba/ |
| /ta/ | /da/ |
| /ka/ | /ga/ |

and repeat, if possible, for a *French* speaker or speaker of some other *Romance language*.

Play the tape back making a permanent paper visual record using a chart recorder running at a paper speed of 100 mm/s. For each token measure the voice onset time. That is, measure on the chart recording the distance in mm from the plosive release to the start of vocal cord vibration. Sometimes you will notice that the vocal cord vibration begins simultaneously with the release, in which case the VOT is 0. On most occasions the vocal cord vibration begins after the release, in which case you have a positive measurement. When the vibration begins before the release you have a negative measurement (see the section above on *Voice Onset Time*). Since the chart recorder's paper speed was 100 mm/s each millimeter measured represents 10ms (milliseconds). Set up tables, thus:

/pa/ English:

| tokens | VOT |
|---|---|
| 1 | +x |
| 2 | +y |
| 3 | +z |
| etc. | |

/ba/ English:

| tokens | VOT |
|---|---|
| 1 | –a |
| 2 | 0 |
| 3 | –b |
| etc. | |

Calculate the arithmetic mean of the results (see the Section on *Statistics*).

You will notice that for the phonological sequence of [–voice] consonant followed by a vowel in English the VOT is between around 20–50 ms; for [+voice] followed by vowel sequences the VOT is between –2 and +2 ms approximately. By contrast, in French the [–voice] + vowel sequence has a VOT of around 0 to +2 ms, and the [+voice] + vowel sequences a VOT of some –20 ms. The actual numbers vary from speaker to speaker and from consonant to consonant. *[Q. Does it look as though particular consonants have a consistently longer or shorter VOT?]*

*Conclusions*

Abstract phonological specifications like [+voice] or [–voice] do not always have a direct one-to-one phonetic realisation. *Phonetic features* (in this case that of vocal cord vibration) do not necessarily correspond directly with phonological features (in this case [voice]), nor do phonetic features necessarily reflect the abstract segmental divisions of phonology, but tend to 'blur' very often across where a segmental boundary might be. Notice also that different languages often realize the same abstract phonological specifications differently in the phonetics. Here the phonological feature [+voice] is realized in English by a zero VOT, but in French by a comparatively long negative VOT, whereas the [–voice] feature was realized in English by a long positive VOT, but in French by a zero or very short positive VOT.

## 4. Intraoral Air Pressure

During the *stop phase of a plosive consonant*, air continues to be expelled from the lungs through the glottis where there may or may not be vocal cord vibration. This causes an increase in intraoral air pressure which explodes immediately following release of the stop. It has been hypothesized that a distinguishing cue in the differential perception of stops, particularly those phonologically differentiated by the tense/lax features, depends on hearing a different audio from the explosion due to the different amounts of air accumulated under pressure.

Specifically it is hypothesized that in the case of a [+tense, –lax] stop the explosion will have greater amplitude than in the case of a [–tense, +lax] stop. The object of this experiment is to determine whether there are differences in intraoral air pressure between phonological [+tense] and [–tense] (correlating with [–voice] and [+voice]) stops.

*Equipment*

- Electromanophone
- Mingograf
- microphone

*Setup*

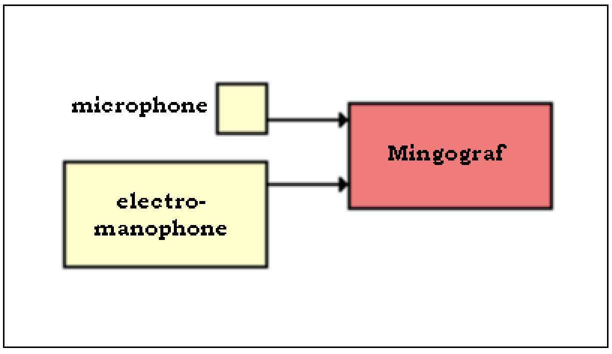



Fig. 17 — Equipment arrangement for examining intraoral air pressure.

*Procedure*

Because of the difficulty of detecting air pressure behind back consonants since the tube has to pass into the pharynx *via* the nasal passage, we confine ourselves to bilabial consonants [p] and [b].

However you might like also to try the alveolar stops [t] and [d]. Have one or two subjects repeat 15 times:

/apa/ and /aba/.

Typical Mingograf traces will look like this:

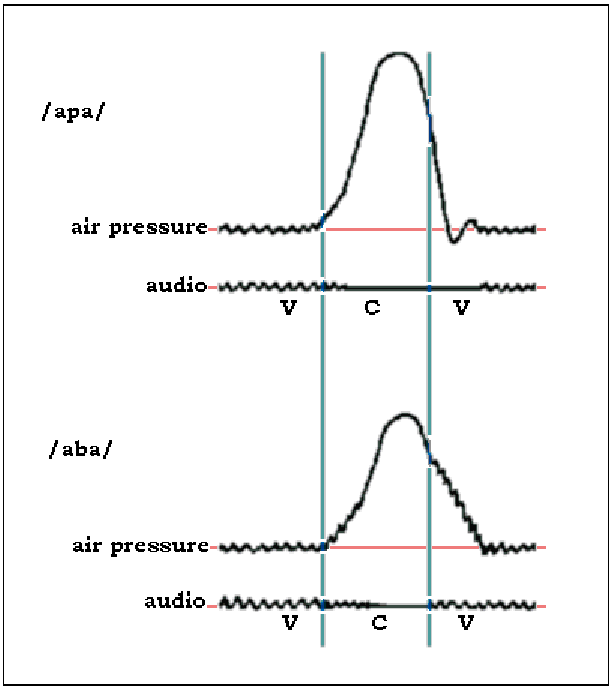



Fig. 18 — Typical Mingograf traces for intraoral air pressure and audio for /apa/ and /aba/.

Notice the following:

1. vocal cord vibration detectable on the air pressure (a-p) trace;
2. VOT is positive for the realisation of /p/; zero or negative for /b/;
3. air pressure rise time is faster for /p/ (the slope is steeper);
4. air pressure decay time is faster for /p/;
5. air pressure peak pressure is some 20% lower for /b/;
6. V1 (the first vowel) vocal cord vibration continues longer when /b/ follows;
7. air pressure falls slightly immediately before the release;
8. air pressure falls below zero often after release for /p/.

Try to explain these observations before reading on.

*Explanations*

1. If there is vibration there must be pressure changes — vibration is pressure change. The vocal cords are cyclically interrupting the smooth airflow through the glottis, resulting in pressure variations in the oral cavity. The manophone is sensitive enough to register these.
2. cf. the VOT experiment. Notice that usually vocal cord vibration does not continue throughout [b] in English even though /b/ is phonologically [+voice], even here in the

immediate context of two [+voice] segments. This is because in order to vibrate the vocal cords need the supra-glottal air pressure (the intraoral air pressure) to be significantly lower than the sub-glottal pressure. If the air passage is stopped the air cannot escape, so the pressure builds up to a critical level and vibration ceases because the required supra-/sub-glottal pressure ratio no longer holds.

3. The rise time is faster for [p] because there is no glottal vibration to impede flow of air into the oral cavity. With [b] glottal stricture (for vibration) slows the airflow. Notice that when, in [b], vocal cord vibration stops the rise time sharpens.
4. Similarly, vocal cord vibration early in [b] slows release time of the air pressure. Also less air was compressed, so the explosion is slower.
5. Because of impeding glottal stricture, less air has flowed into the oral cavity, and so the pressure peak is lower. Much less glottal impedance is presented during realisation of /p/; there is therefore more air flowing, so the peak pressure is higher.
6. Vocal cord vibration continues, as explained in 2. Theoretically the subject tries to keep the vocal cords vibrating as long as possible.
7. In anticipation of release, the musculature of the lips relaxes momentarily. This causes the air pressure to push the lips forward a little, thus enlarging the cavity in which the air is held — resulting in a drop of pressure. Volume and pressure are inversely correlated.
8. The pressure is falling so rapidly that it actually overshoots the zero baseline and momentarily goes negative. This oscillatory effect makes the onset of vocal cord activity difficult and partly accounts for positive VOTs following the realisation of [–voice] plosives. These have the highest air pressure peaks during closure: the higher the peak air pressure achieved the more likely is this oscillatory effect to occur.

*Treatment of Results*

The table of results should be subjected to Mann-Whitney *U*-tests (see the Section on *Statistics*) to establish the reliability of the differences of air pressure peak between the realizations of /p/ and /b/. If it is practical to measure the duration of closure (i.e. the time from pressure rise to pressure fall) then these should also be compared using the *U*-test.

## 5. Airflow

This procedure is designed

- to show some of the *aerodynamic effects* in speech, and
- to verify an auditory observation concerning *nasalization of vowels* in English.

Obviously air flows in and out of the mouth and nose. The aerometer has four transducers which detect the four airflow possibilities: mouth in/out, nose in/out. There is also a microphone in the mouth section of the mask to enable an audio recording to be made simultaneously with the airflow.

> **Caution:** Because the microphone is in a confined and resonating space its signals are often anomalous, so it should not be used to provide a signal for accurate measurement, but merely for monitoring purposes.

*Equipment*

- electro-aerometer
- Mingograf

*Setup*

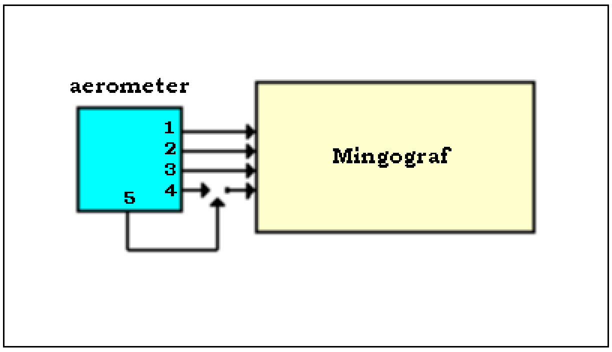



Fig. 19 — Equipment arrangement for examining airflow. The connections on the aerometer are: 1. Mouth out; 2. Nose out; 3. Nose in; 4. Mouth in; 5. Microphone channel.

Notice alternative connections for the Mingograf 4th channel — ideally we should use a 5–channel chart recorder.

*Procedure A*

Get the subject to breathe normally in and out of

- the nose,
- the mouth,
- in through the nose, out through the mouth.

Take note of the traces on the Mingograf showing airflow.

Ask the subject to utter various isolated vowels, breathing in between the vowel utterances. Notice the oral flow out on vowels and the absence of nasal flow out — the velum is closed in English for vowel segments to prevent any nasal resonance. Get the subject to alternate English vowels with French nasal vowels, noticing the airflow through both mouth and nose during nasal vowels. If you know any other language with nasal vowels try these (e.g. Portuguese).

Ask the subject to utter VCV sequences with stops: as /ata/, /ada/, /apa/, /aba/, etc. Notice the stoppage of airflow during the consonant and the high level of airflow immediately following the explosive release, as the pressure drops rapidly in the oral cavity. Compare these traces of airflow with air pressure traces:

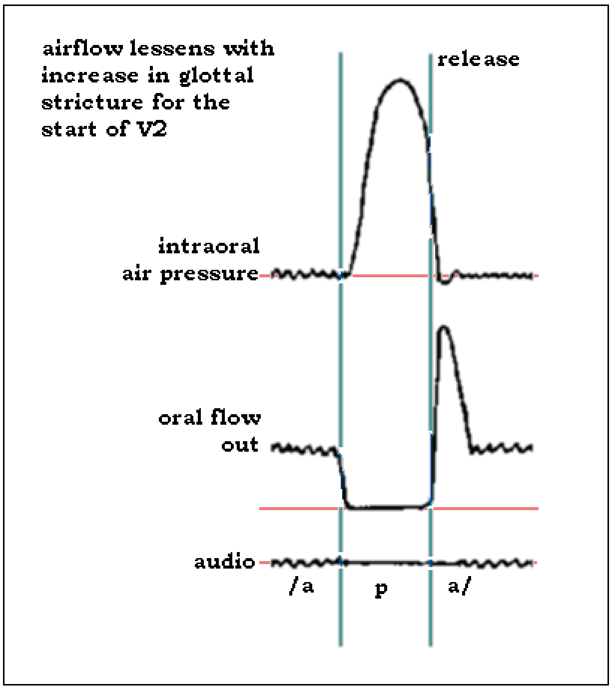



Fig. 20 — Typical intraoral air pressure, oral outward flow and audio signals associated with an instance of /apa/.

Have the subject repeat alternately /ada/ and /ana/; also /aba/ and /ama/, where, phonologically /d/ and /b/ are [–nasal] counterparts of /n/ and /m/ which are [+nasal]. Notice that there is no nasal airflow out during /d/ and /b/, but nasal airflow only during /n/ and /m/. Notice the timing of onset and offset of nasal flow, and notice that this does not correspond exactly with oral closure and release.

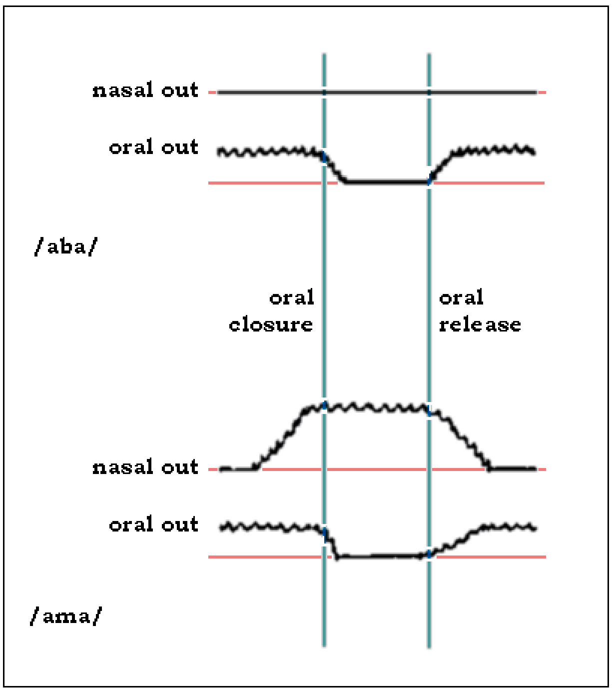


Fig. 21 — Outward nasal and oral airflow comparing instances of /aba/ and /ama/.

Now ask the subject to utter a few voiceless fricatives, like /f/ or /s/, and note the airflow. Have him utter the sequence /sa/, for example. Note a change in the airflow rates between the fricative and the vowel, and also that the vowel does not immediately take over from the fricative. This is because the fricative's intraoral constriction has raised the supraglottal air pressure above the critical level for voice. The pressure must fall, which takes time and involves flow out between the fricative and the vowel, before the vocal cords can begin to vibrate properly (see the Section on *VOT*).

*Procedure B*

No vowels in English are phonologically nasal at the underlying contrastive level. Thus, phonologically the [–nasal] vowel in bad is the same as that in man. Note that it has been observed, though, that auditorily the V in man has 'nasal quality' — that is, has a nasal formant present. This observation would suggest that there is air flowing into the nasal cavity causing resonance.

The segment sequence /bad/ is phonologically as follows:

| | | | |
|---|---|---|---|
| consonantal | + | – | + |
| vocalic | – | + | – |
| stop | + | + | + |
| nasal | – | – | – |
| voice | + | + | + |
| etc. | | | |

and /man/ is:

| | | | |
|---|---|---|---|
| consonantal | + | – | + |
| vocalic | – | + | – |
| stop | + | – | + |
| nasal | + | – | + |
| voice | + | + | + |
| etc. | | | |

Note that /b/ differs from /m/ and /d/ from /n/ on the [nasal] feature, and that in both words the vowel is [–nasal]. From the phonological specification we would predict an abrupt cessation of nasal airflow after the /m/, to be resumed immediately after the vowel for realisation of the final /n/.

Ask the subject to alternate /bad/ and /man/ several times. And then have him say /man/ with an American accent, or better get an American to say the word. Note that in fact there is continued nasal airflow through the vowel, despite the phonological prediction or requirement; though this is perhaps at a quite low level. Observe that there is no nasal airflow during the vowel in /bad/. The American version will probably have greater nasal airflow during the vowel than the British English version.

The phenomenon of the observed nasalization of the vowel is known as *coarticulation.* There is a certain amount of inertia in raising and lowering the velum (which is responsible for stopping off the nasal airflow). This inertia cannot be overcome to cause a completely abrupt stop/start of nasality at the segmental boundary — so the result is a blurring or overlap of the phonetic parameters: this accounts for nasality at the beginning of the vowel. Nasality at the end of the vowel occurs in anticipation of the following nasal consonant. Usually this overhang at the start and the anticipation at the end of the vowel produces a continuous nasality throughout the theoretically non-nasal vowel1 In American English this is even more pronounced an effect. This indicates perhaps that the different dialects vary in their attempts to overcome the inertial and anticipatory phenomena.

Thus the nasalization effect is an artifact of' the mechanical lowering and raising of the velum and its control system. Because there is no phonological contrast in English between oral and nasal vowels a certain degree of phonetic nasality does not matter and cannot give rise to confusion during perception. The listener, knowing that a genuine nasal vowel is never intended by a speaker of English, tends to disregard the artifact. This is an important phenomenon in Speech Perception.

## 6. Amplitude and Stress

A possible correlate of the abstract phonological features [stress] is phonetic acoustic amplitude. It can be hypothesized that stressed vowel syllable nuclei are realized with greater amplitude than unstressed vowels. We shall try to see whether this is the case.

### *Registration of Amplitude*

Amplitude is extracted from the audio signal by a two stage processing of the electrical signal transduced from the sound pressure wave. The equipment used consists of a rectifier and an integrator or smoothing circuit, arranged as follows:

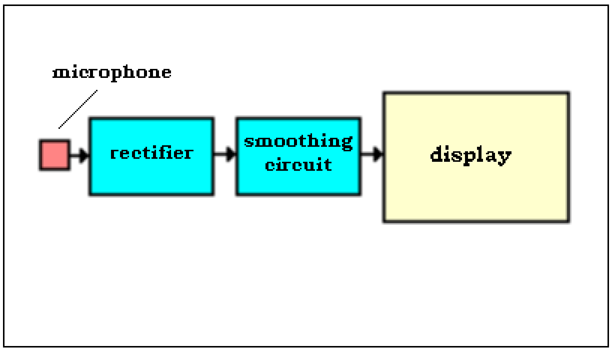



Fig. 22 — Processing a microphone signal to display audio amplitude.

*Rectification*

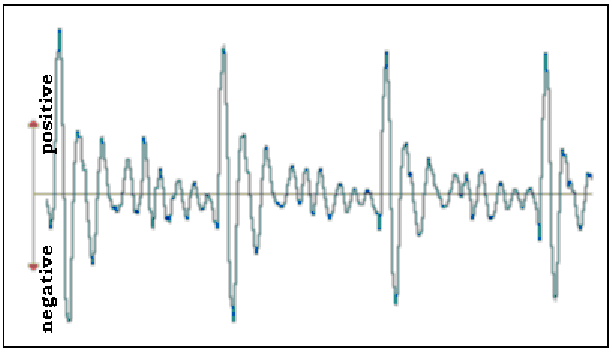



Fig. 23 — A typical waveform for a vowel. Notice the signal consists of both positive and negative going elements.

A waveform consists of positive- and negative-going elements *oscillating* about a zero baseline. The process of *rectification* is designed to eliminate the negative-going elements leaving only the positive part of the waveform. Two types of rectification are possible:

- half-wave, and
- full-wave.

In half-wave rectification the negative elements are detected and discarded completely, leaving only the positive elements of the signal, thus:

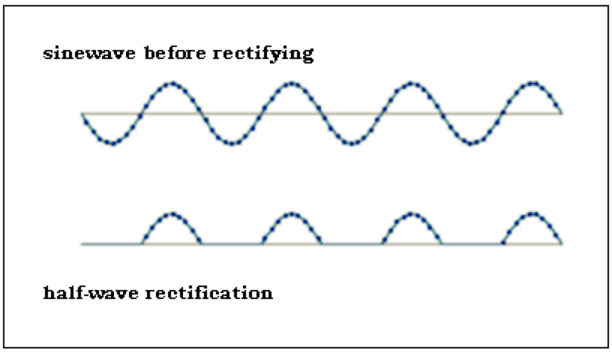



Fig. 24 — A simple sinewave before and after half-wave rectification.

In full-wave rectification (the method used in the Essex Speech Laboratory) all negative-going elements are detected and have their sign changed to positive. This method is more accurate in many ways for the purposes of deriving amplitude:

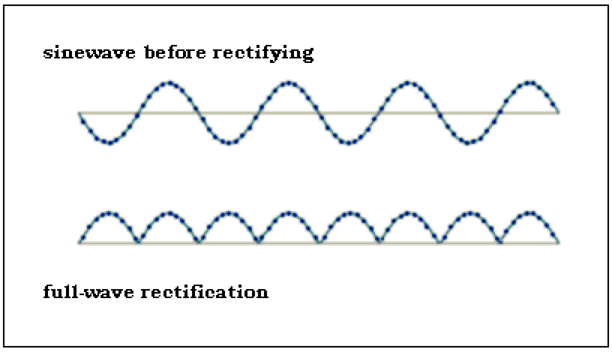



Fig. 25 — The same sinewave, but this time with full-wave rectification.

*Smoothing*

Smoothing is an elementary form of integration, and is accomplished by adding inertia to the path the signal is made to follow. Inertia limits the possibilities of abruptness of change of direction of the signal (i.e. change from positive-going to negative-going and *vice versa*). Thus in the above rectified waveform, note the sharpness of the lower edge of the trace:

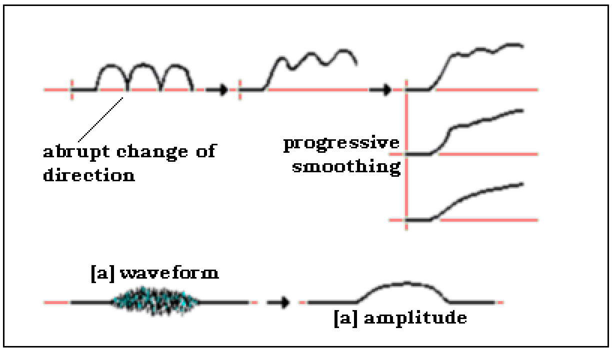



Fig. 25a — The progressive application of inertia to a full-wave rectified signal. Notice the correlating increase in the smoothness of the tracing until little if anything of the original 'ripples' is left.

The distance from the curve to the baseline is said to be the amplitude of the signal at any given moment in time. Applying rectification and smoothing to the waveform of, say, a vowel, we derive an amplitude curve for that waveform.

*Equipment*

- microphone
- rectifier
- integrator
- display device

*Procedure*

a. *General.* With the subject uttering single vowel sounds display amplitude curves and waveforms on the Mingograf simultaneously. Note the effects of varying the electrical inertia of the system, i.e. varying the integration time, which is expressed in *ms*. Choose an integration time which just leaves a trace of vowel ripple at the top of the curve: a greater integration time will obscure important detail later, such as short-lived plosive bursts. Try this by saying [ɑpɑ] with different integration times, noting how the release peak disappears as the time is lengthened.

b. *Stress and Amplitude Correlation.* Ask the subject to say the words *cart* and *kit*. Notice that both these words uttered in isolation bear primary stress, but notice also that the [ɑ] of *cart* has a greater amplitude than the [ɪ] of *kit*. Vowels have different *intrinsic amplitudes* just as they have other different intrinsic properties, like *duration*. With [ɑ] the oral cavity is wide open. This is a low, or open vowel, and airflow (and therefore sound dispersion) is relatively unimpeded; but with [ɪ] the oral cavity is much more constricted. This is a high, or close vowel, and airflow is relatively curtailed giving rise to an acoustic waveform of intrinsically lower amplitude. Similarly, with [ɑ] the second formant is lower in frequency than the second formant of [ɪ]. Remember that the lower the frequency of a formant the greater its amplitude, by and large. So, equally stressed vowels phonologically may have different intrinsic amplitudes.

What we need to see is whether a stressed [ɑ] has greater audio amplitude than an unstressed [ɑ] — i.e. the same vowel. Comparing different vowels without some kind of normalisation to take care of their intrinsic differences in amplitude would give anomalous results, for stressed [ɪ] may well have less amplitude than unstressed [ɑ]. Try this.

Take a word like *Alabama*, comparing the initial vowel (unstressed) and the third vowel (stressed) with respect to sound wave amplitude. Repeat the word some 15 times and calculate averages of the peak amplitude achieved for each token for each of the two vowels. Draw up a table, and use a Mann-Whitney *U*-test to see if there is any statistical difference between the two vowels. Try other words and different vowels.

Notice that very often a stressed vowel changes its quality altogether when unstressed. This is a phonological phenomenon, known as *vowel reduction*. So, for example:

- the underlying [æ] in the second syllable of salad becomes the reduced central vowel [ə],
- the underlying [ɔ] in the second syllable of method becomes the reduced central vowel [ə].

You should find that although often amplitude does correlate well with phonological stress, this is not always the case, and is perhaps an unreliable correlate overall. *[Q. Why then are we always able to tell stress placement when we hear words and sentences pronounced, even though amplitude may not be a reliable cue?]*

## 7. Duration

Speech segments vary considerably in length, but human beings are relatively consistent in keeping the audio length of individual segments nearly the same. In other words, the vowel in *cart* for example, is much longer that the vowel in kit, yet repetitions of the long [ɑ] are much the same length, as are repetitions of the short [ɪ]. In this experiment the object is to discover some of the intrinsic length attributes of particular phonetic segments.

### *Vowels*

Using a microphone and Mingograf to display the waveform, utter the following words in a suitable frame, making sure the overall tempo remains as constant as you can make it. Obviously speaking slowly or fast is going to have some effect on segment duration, so keep a fairly constant speed or rate of utterance.

*Frame*

*Say the word — again.*

| | |
|---|---|
| carp | [kɑp] |
| cap | [kæp] |
| kip | [kɪp] |
| cop | [kɔp] |
| cope | [koup] |
| coop | [kup] |
| cup | [kʌp] |

Each word should be repeated at least five times, and the *arithmetic mean* calculated for the vowel duration of each different word. This should be measured as follows:

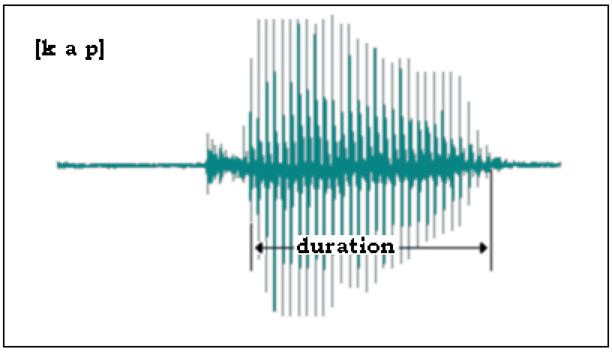



Fig. 26 — The measurement of duration for an inter-plosive vowel.

Measure from the *first vocal cord vibration* that you can detect to the *last vocal cord vibration*. Use a fairly high gain setting on the Mingograf to emphasize vowel ripple to make the start and finish points easier to spot.

This method follows that phonetic school which believes that the VOT or *aspiration* between the release of the consonant and the vocal cord vibration belongs to the consonant segment realisation. A more modern school would assume that the realisation of the vowel begins coincident with the consonant's release. In this case, the appropriate measurement would be:

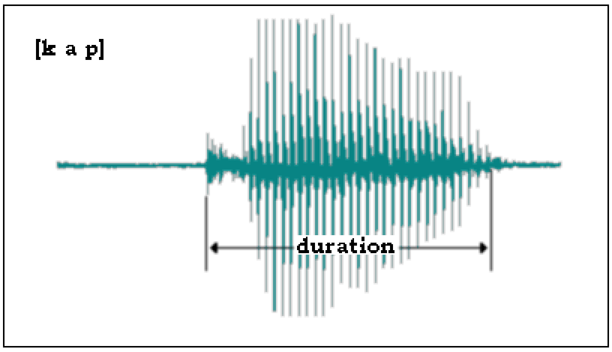



Fig. 27 — An alternative measurement of duration for an inter-plosive vowel.

In both cases the end point of the vowel is taken as the moment of closure for the stop consonant. Notice, though, that we are only inferring this from the audio tracing produced. A more reliable method uses air pressure or airflow to indicate closure. With pressure it is the moment the pressure begins to rise; with airflow the moment the flow ceases. *[Q. Why should these two non-audio parameters be more reliable?]*

Take both types of measurement for vowel duration and compare them. You will notice that you can rank order the vowels with respect to their durations

We are out initially to discover the intrinsic durations of vowels. Intrinsic durations may in fact be concealed by contextual effects like coarticulation. Try the following:

| | |
|---|---|
| garb | [gɑb] |
| gab | [gæb] |
| gib | [gɪb] |
| gob | [gɔb] |
| ‘gobe’ | [goub] |
| ‘goob’ | [gub] |
| ‘gub’ | [gʌb] |

It does not matter that some of these words are not English — they are potentially English because they do not break the rules for combining segments in English. These rules are discussed in phonology under the heading morpheme structure conditions — these are the conditions placed on what sequences of segments are possible in any given language.

Measurement is going to be more difficult:

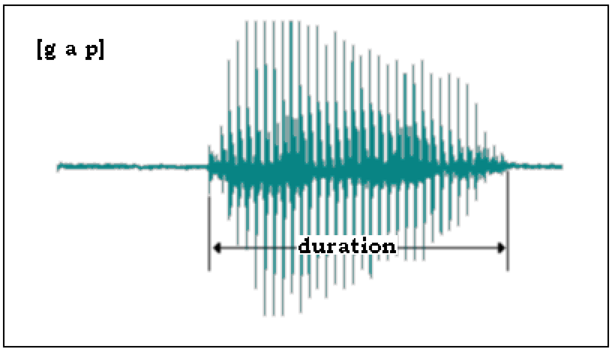



Fig. 28 — Measurement of vowel duration between voiced plosives.

Measure from the release of C1 to get the start of vocal cord activity associated with the vowel realisation. Towards the end of the vocal cord activity you should see a quite sudden lessening in amplitude of the tracing corresponding to the stop closure, although vocal cord activity may continue because C2 is phonologically [+voice]. If detecting this proves too difficult, measure to the end of the visible vibration. Remember, though, that for comparison purposes you must consistently measure all tokens in the same way, whichever one you choose.

You should now be able to rank order the vowels with respect to length as before. Is the rank order the same as when the vowel was in the context of phonologically [–voice] consonants? It should be. Do you find that all vowels are longer in this context than before? — even though the rank ordering is preserved? There is a phonological rule in English which lengthens vowels when they occur before [+voice] consonants. This rule assumes that the *canonical* or intrinsically determined form occurs when the vowel is followed by a [–voice] consonant.

$$V \rightarrow \begin{bmatrix} \overline{\phantom{+\text{long}}} \\ +\text{long} \end{bmatrix} \Big/ \; \text{—} \begin{bmatrix} C \\ \hline +\text{voice} \end{bmatrix}$$

Stress can have an effect on length. In fact, besides amplitude, length is sometimes considered to be one of the cues for indicating phonological stress. Remember amplitude of the acoustic signal seemed unreliable in a previous experiment.

Try a few sentences like the following:

*He didn't say cap, but carp.*
*He didn't say carp, but cap.*
*He didn't say cop, but cup.*
*He didn't say cup, but cop.*

comparing the same words placed differently in the sentence with and without contrastive stress. Do five repetitions each time and take the arithmetic mean of your measurements.

*Consonants*

*Stops*

Using the *electro-manometer* to detect *intraoral air pressure*, measure the closure durations of [p], [b], [t], [d] in various contexts. Be careful with [t] and [d] to get the tube correctly behind the place of closure, and be careful not to block the tube with saliva or touch its open end with your tongue.

Using the frame: *Say the word — again*, record the following:

| | |
|---|---|
| [ápa] | [ába] |
| [ípi] | [íbi] |
| [áta] | [áda] |
| [íti] | [ídi] |

and then with the stress position reversed:

| | |
|---|---|
| [apá] | [abá] |
| [ipí] | [ibí] |
| [atá] | [adá] |
| [ití] | [idí] |

each five times, taking the arithmetic mean of the measurements. Measure as follows:

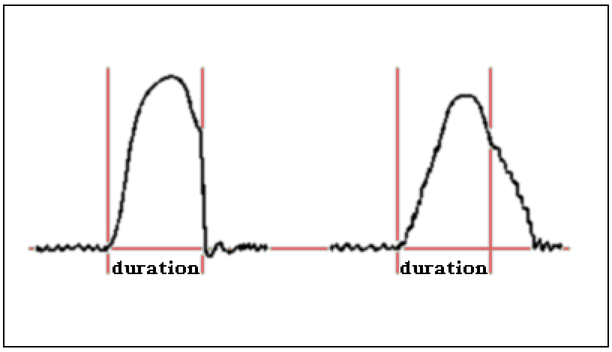



Fig. 29 — Measurement of closure duration for stop consonants using intraoral air pressure.

*Fricatives*

Fricative duration is very difficult to measure from the audio waveform on the Mingograf. The reason is that the energy in fricatives often has a frequency of several thousand Hertz, and of course the Mingograf's *galvanometers* respond reliably up to only about 700Hz as you discovered earlier when you calibrated the instrument. Fortunately a method has been devised to help get around this problem: it is based on the idea that because of the *inertia* introduced by integration the amplitude curve of the fricative will register on the Mingograf. The waveform is transformed into what is known as a *duplex oscillogram.*

The duplex oscillogram is derived by splitting the audio waveform into two parts on the frequency axis. This is accomplished by taking the outputs of two filters, each set to 4kHz — one a low-pass filter and the other a high-pass filter. The low-pass filtered signal contains most of the vowel-associated energy and can be displayed in the usual manner directly onto the Mingograf. The high-pass filtered signal has the fricative energy and is first rectified and then smoothed before displaying. In practice both signals are displayed on the same Mingograf trace by electrically combining the signals before display. Thus:

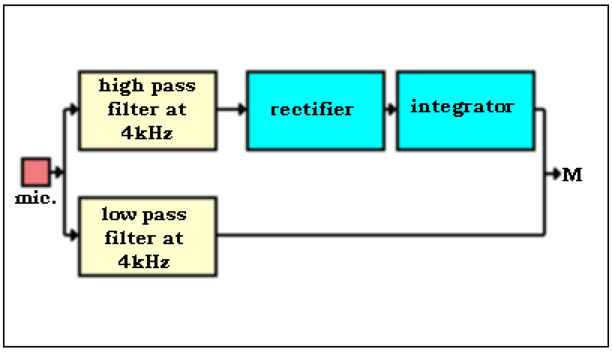



Fig. 30 — Equipment arrangement for displaying a duplex oscillogram.

The trace of the audio waveform of [asa] displayed on an oscilloscope or UV recorder, both of which can easily handle high frequencies, looks like this:

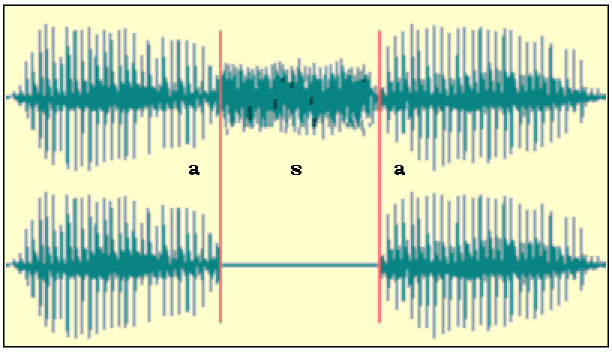



Fig. 31 — [asa] displayed on an oscilloscope and on the Mingograf. Notice the loss of high frequencies associated with the [s] in the band limited inkjet trace.

i.e. the [s] is lost in the display because of its intrinsic narrow frequency response. The duplex oscillogram of the same waveform, displayed on either an oscilloscope or the Mingograf, looks like this:

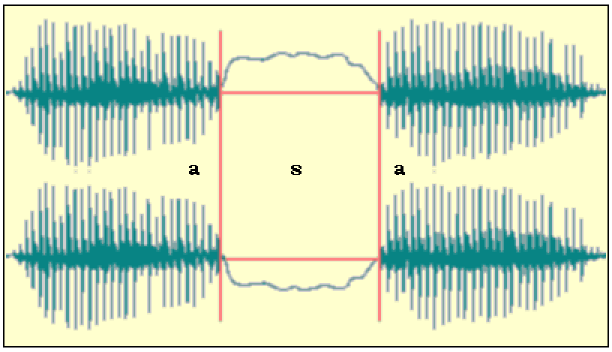


Fig. 32 — Alternative displays of the duplex oscillogram.

In the second version the high-frequency component curve is inverted — this sometimes makes it easier to interpret the graph. All negative-going components in the amplitude parts of the trace should be regarded as positive-going.

Make duplex oscillograms of the following:

| | |
|---|---|
| [asa] | [aza] |
| [isi] | [izi] |
| [aʃa] | [aʒa] |
| [iʃi] | [iʒi] |
| [afa] | [ava] |

Measure duration of the fricative thus:

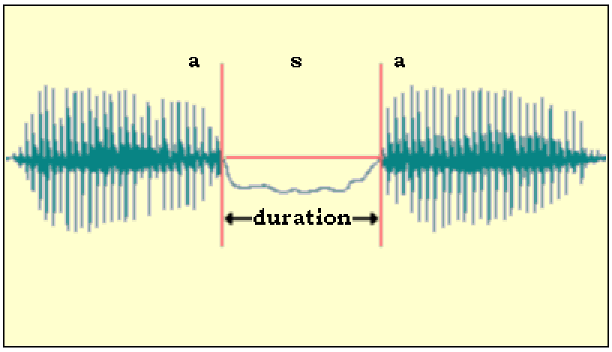


Fig. 33 — Measuring intervocalic fricative duration on a duplex oscillogram.

1. Construct an arbitrary baseline slightly above the baseline, or below if the amplitude display is inverted, of the regular waveform. This is because amplitude displays always contain a certain level of noise.
2. Measure duration between the points where the fricative amplitude curve crosses this line. Positioning the constructed baseline must be identical with respect to the zero of the waveform for all tokens to ensure comparability. Once again, take the arithmetic mean of five tokens of each type.

*[Q. Do fricatives have their own differentiating intrinsic durations?]*

*[Q. Can you rank order them?]*

## 8. Electromyography — The Tense/Lax Controversy

Electromyography (*emg*) is an instrumental technique for registering the *electrical signals* occurring when the *fibres* within a muscle contract on receipt of a *motor command* from the brain's *motor cortex*. These are minute electrical signals, and very special techniques are involved in the detection and amplification of the signals. Transducing is not involved because the signals are already in electrical form, and direct pickup is possible. (See the Section on *Instrumentation*)

In phonology, consonantal segments are divided into two major classes by the feature [tense]. Segments which are [+voice] are usually [–tense] and vice versa. A major pursuit in phonetics since the middle sixties has been defining the phonetic correlates of the phonological distinctive features, and much of this work is taking on a new significance as some phonological theories take on a less abstract nature. So, we ask the question: what is the phonetic property which is used to realize the [tense] phonological feature? That is, what are the *phonetic correlates* of the feature [tense]?

The pre-transformational grammar literature used the term *tense* also, and phoneticians were quite sure that [+tense] consonants involved more muscular tension and a more deliberate articulation than the [–tense] consonants. Thus, it might be hypothesized that [+tense] realizes phonetically as more muscular force. If this is so, then this increased muscular force should be apparent in a higher amplitude emg signal, since the emg signal can be taken to reflect the intended degree of contraction of a muscle — and more contraction is required for more force.

*Equipment*

- emg apparatus
- Mingograf
- rectifier and integrator
- microphone
- oscilloscope for monitoring
- tape recorder for permanent record

*Setup*

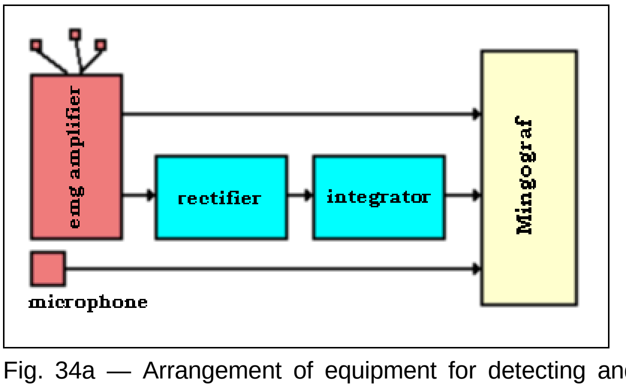



Fig. 34a — Arrangement of equipment for detecting and measuring electromyographic signals using bipolar surface electrodes.

Emg signals are rather like the audio waveform, and quite difficult to interpret. For our purposes here, where we are interested in the *degree of muscular contraction*, *amplitude* is the important parameter, since it is this which corresponds to degree of muscle contraction. As with the audio waveform, amplitude is derived by a process of *rectification and integration.* The raw, unprocessed signal should be displayed alongside for reference purposes together with the audio waveform.

*Procedure*

We want to determine whether the prediction of emg amplitude difference to be inferred from the abstract phonological [tense] marking for consonantal segments has any *physical reality*. Selecting the null hypothesis, we hypothesize no difference in emg amplitude in the realisation of [+tense] and [–tense] consonants.

*M. orbicularis oris* which runs around the periphery of the mouth is an easy muscle to examine since it is just below the surface, and is a primary articulator for the *bilabial set of stop consonants*. The muscle is sphincter and responsible for lip closure, lip rounding and lip protrusion. The superior portion of the muscle (above the upper lip) is a good site for the electrodes since here there is minimal pickup of potentials from other muscles. Electrode placement should be like this:

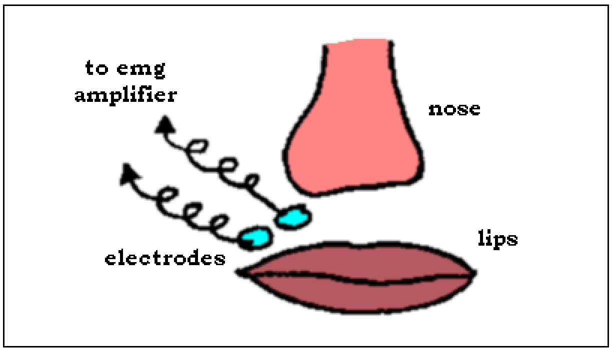



Fig. 34b — Electrode placement for recording signals from m. orbicularis oris.

Ideally the skin surface should be abraded slightly after shaving, cleaned with alcohol, and the electrodes attached with Blenderm tape after having their indentations filled with sodium chloride gel. An appropriate ground electrode should also be attached.

Have the subject repeat fifteen times:

- Say the word ‘apa’ again. [apa]
- Say the word ‘aba’ again. [aba]

The only difference between these two sentences along the phonological feature of tensing is that i. has [+tense] on /p/ in /apa/ and ii. has [–tense] on /b/ in /aba/.

The raw emg, processed emg and audio records should look something like this:

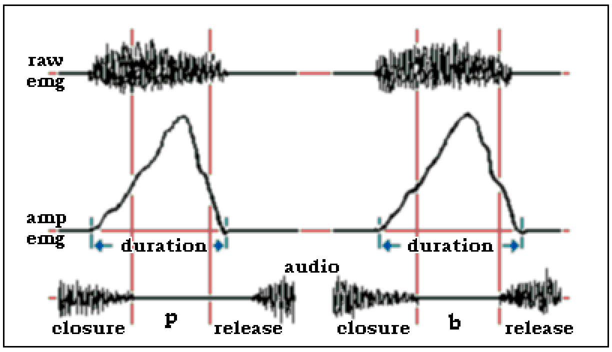


Fig. 35 — Raw emg, processed emg and audio traces.

- Measure the *peak amplitude* of the processed emg signal.
- Measure the *duration* of the signal associated with contraction to obtain lip closure, having constructed an arbitrary baseline (see the Section on the Duplex Oscillogram and the amplitude of fricatives).

Notice that the emg signal begins significantly before the moment of closure: obviously because the lips have to be moved together before they will close. Notice then that with the entire complex musculature of speech, synchronization is going to develop into a major problem: large, heavy articulators must have their movement initiated before light, small muscles — especially if both large and small muscles are to arrive ‘on target’ simultaneously.

Notice that contraction all but ceases before release, otherwise release could not occur if contraction were held too long.

*Note:* Your records may look like this for the processed emg:

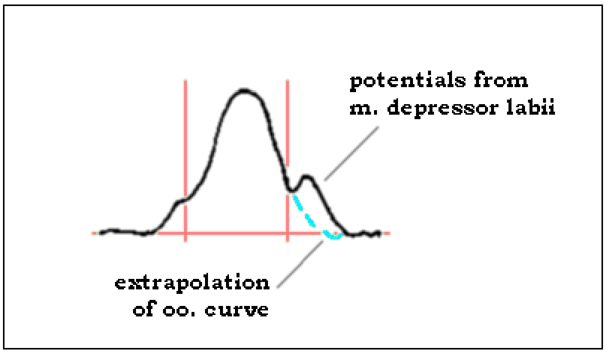


Fig. 36 — Processed emg signals from m. orbicularis oris, showing crosstalk from m. depressor labii

A second peak roughly approximating the release time may be visible. This peak is usually spurious pickup of signals from *m. depressor labii* which run (on each side of the chin) down from the corners of the mouth, and is used for opening the lips on time, following closure by *m. orbicularis oris*. Be careful not to include this signal in your duration measurements, but extrapolate the curve from *oo*. as shown.

Make up a results table as follows:

| | /apa/ | /aba/ |
|---|---|---|

| token | peak emg amplitude | emg duration | peak emg amplitude | emg duration |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| . | | | | |
| . | | | | |
| . | | | | |

## 9. Invariance

The question has been asked whether the fact that identical specifications are used in the phonology for segments occurring in *different contexts* in words and sentences *implies identical motor programming* of these segments when it comes to phonetic realisation. Some regard this still as a controversial issue. Let us conduct a simple experiment to see if we can throw light on the controversy.

Emg captures muscle fibre potentials which, for the most part, arise from deliberate motor control. The emg signal therefore reflects intention to contract, and to contract to a particular force for a particular time. A simple examination of sameness or difference of the *amplitude and duration parameters* for phonologically same segments in different contexts should reveal whether, at the phonetic level, there is intention of sameness or difference.

Using the previous experimental setup, have the subject repeat 15 times each:

| | |
|---|---|
| /pit/ | /tip/ |
| /bid/ | /dib/ |

in the frame *Say the word — again*. Notice that two consonants have been used in extreme word contexts, *initial* and *final*. The previous experiment gave you *word-medial*.

Perform Mann-Whitney *U*-tests, comparing initial /p/ with final /p/, and then initial /b/ with final /b/. *[Q. Does context affect motor programming intentions?]*

## 10. Stuttering

### Introduction

Stuttering is the most prevalent of all the speech disorders and affects some 4% of the population: that is, 2,000,000 people in Great Britain are stutterers. 75% of stutterers are male. Stuttering usually has a fairly sudden onset, and the peak year of occurrence onset is 4. The vast majority of stutterers belong to the middle classes.

Stuttering is in almost all cases ‘curable’ — in the sense that it can be rendered inaudible and in normal conversation no stuttering is apparent. It is sometimes argued, though, that some of the underlying effects are still present although they might no longer present any problems to the speaker or listener.

Stuttering is a phonetic phenomenon and has nothing to do with phonology. That is, it is a problem of the *motor-control of speaking*, and is not connected with the speaker’s knowledge of or ability to handle phonological processes in a language, a mental task.

It has been fairly recently observed that spasms of the vocal cord tensing musculature occur in stutterers. In fact, such spasms occur in normal speakers as well, but at a much lower frequency — say, 5 per minute for the normal speaker, compared with 5 per second for the stutterer on average. These spasms take the form of an involuntary short-lived contraction of the larynx musculature resulting in momentary tensing of the vocal cords or in a shock-wave which travels along the edges of the vocal cords.

Interestingly, control of the vocal cords in speech is one of the most critical adjustments that has to be made. Correct vibration will not occur unless the vocal cord tension is just right to balance between subglottal and supraglottal air pressures. Imagine, then, a situation in which a speaker is carefully adjusting the vocal cord tension to begin a voiced sound — he hasn’t much time because such adjustments have to be made very rapidly to keep the flow of speech going — when suddenly one of these spasms occurs, upsetting all his careful adjustments and causing the system to break down. He can do one of two things:

- try to adjust again rapidly, usually with the result that he overdoes it and locks the musculature too tense, resulting in stopping the necessary airflow between the vocal cords, or
- he can stop altogether and try again, hoping that this time a spasm will not occur.

The first strategy results in what is commonly known as a ‘block’, in which the person is trying desperately and far too vigorously to produce the sound, with no effect at all except to totally disrupt the fluency of speech. The second strategy results in a repetition of a sound or syllable which may occur up to three or four times before normal fluency is restored, only to break down again when another spasm occurs at a critical moment. Why these spasms occur or what their underlying mechanism is not known, but we can observe such spasms in other muscles.

As mentioned above, control of the vocal cords is always a difficult operation. In the normal flow of speech the *degree of precision* varies with the general aerodynamic requirements. The most critical moments occur when a plosive (which disrupts and hold up the airflow) is released into a vowel (which requires the vocal cords to vibrate) and at the beginning of a vowel which starts off a word or phrase. If a spasm occurs during this critical period a block or repetition is likely to occur; more so than at any other time in speaking more so than halfway through a vowel or voiceless fricative in word-final position, for example.

Since the spasms occur 50–100 times more frequently in the stutterer than in the normal speaker the sheer statistical probability of one occurring at a critical moment is much greater. But in the stutterer such occurrences are not only much more frequent but in addition the anxiety level is so high that the effect of the spasm is that much greater than in a normal speaker. Often a moment of panic is induced, which tends to increase the rate of occurrence of the spasms, hence blocks or repetitions also increase in rate. The spasms are so infrequent in the normal speaker that either they rarely occur at a critical moment, or they do not cause the faltering, and the effect passes often unnoticed by either speaker or listener.

## Stuttering — an Experiment

*Object*

- to examine emg of a primary articulator in a stutterer;
- to examine vocal cord activity of a stutterer;
- to compare primary articulator emg of the stutterer with typical emg from a normal speaker.

*Equipment*

- emg electrodes and amplifier,
- rectifier and smoother (for processing the emg signal into an amplitude display),
- accelerometer (with pre-amplifier),
- FM tape recorder (for a permanent record of the signals),
- Mingograf.

*Setup*

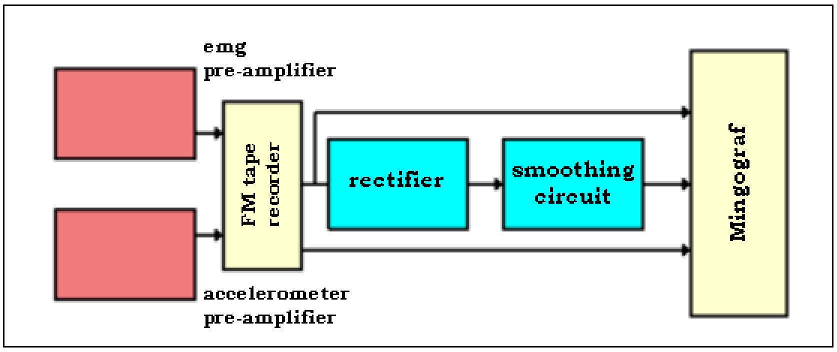



Fig. 37 — Arrangement of equipment for the stuttering experiment.

*Procedure*

- Record emg signals from *m. orbicularis oris* by placing electrodes on the upper lip of the subject. Record *vocal cord activity* from the subject simultaneously by fixing the *accelerometer* transducer to the neck wall in the region of the larynx.

  *Note:* The accelerometer registers vibration present at the neck wall over the larynx resulting from vibration within the larynx. It is sensitive to acceleration and deceleration of the tissue movement: its signals should not be confused with those from an *electroglottograph* which registers the time varying area function of the glottis. We are using the accelerometer here because of its unusually high sensitivity to any vocal cord movement.

- Have the subject say the phrase *Maybe he should pay very poor people more*. There should be at least 12–15 repetitions.
- Record the signals on the *FM tape recorder* and later play back displaying on the Mingograf the accelerometer recording, the raw emg and the emg signal processed through the *rectifier and smoother*.

  Note that severe stutterers often produce very erratic emg signals. Sometimes these present recording and measuring difficulties in both the amplitude and duration parameters which need some considerable experience to carry out. Concentrate on the after-therapy examples from a subject. If at all possible use a subject who has not undergone recent therapy.

- Measure peak amplitude and duration of the emg for each of the bilabial closures in the sentence: Maybe he should pay very poor people more. Some durations will be more difficult to measure than others: for example, the /p/ in poor will probably run into the following rounded vowel, making duration measurement impossible. Amplitude measurements of the /m/s may be difficult because there may be no clear peak in the signal: choose what seems to you to be the highest peak.

  It is highly likely that even a severe stutterer will not audibly stutter during the experiment — stutterers often don't stutter in certain stress conditions or when performing a repetitive task like this one. That does not matter. We are interested in the stutterer's general control of lip-closure and vocal cords on the assumption that stutterers behave differently from normal speakers all of the time.

*Results*

Make up *tables of amplitude and duration* of the emg signal for the measurements before and after therapy. Calculate arithmetic means, standard deviations and coefficients of variation for each example of bilabial closure.

- Compare, using the Mann-Whitney *U*-test, one example of each of /p/, /b/ and /m/ in the amplitude parameter. Compare these by using /p/ as a standard and applying the statistic between /p/ and /b/, and then between /p/ and /m/.
- Compare variability, or precision, of each successive bilabial. Remember, narrow variability — small values of $v$ — indicates greater precision in the articulation. Compare these variations with those obtained in data you have collected from normal subjects, pooled by averaging coefficients of variation across several subjects, to obtain some standard measure of precision for the average normal speaker.

Notice the plosive release is easily detectable in the accelerometer trace. The sudden release in the intraoral air pressure develops a shock wave which momentarily vibrates the vocal cords — you can usually see this. Measure VOTs in *pay* and *people*. Determine coefficients of variation and compare with your data from normal subjects. The hypothesis is that stutterers — even when not audibly stuttering — exhibit greater variation in VOT than normal speakers.

*Questions*

1. Have you discovered that sometimes variation is less, or precision is greater, in a stutterer than in a normal speaker? If so, why might this be?
2. Does variability change through an utterance? If so, is there any patterning in the change?
3. Is the amplitude of the emg signal between /p/, /b/ and /m/ different?
4. Is the variation in VOT greater or narrower than that of a normal subject in initial /p/? Why?

## APPENDIX A

### Statistics

One of the major problems associated with the measurements and evaluation of any aspect of human behavior is its inherent variability. This is as great a problem in experimental phonetics and phonology as it is, say, in psychology. Variability simply means that values obtained for a measured parameter will not be the same on different occasions, making it impossible to say which value was 'right'. In fact the question: *Which is right?* can't be asked — all the measurements are right in some sense.

We recognize though that a human being may report his intention to perform identical acts of behavior even though he may produce some variability. His reported intention can be considered an abstraction which contains no variability: what he *wants* to do might be called a mental idealized concept free of any physical constraints which might impose variations in the execution of that concept. Bearing in mind that the abstract concept contains no variability, researchers in psychology, linguistics, etc., have sought to determine an automatic method of moving from the measured variable data to an invariant abstraction which can be equated with the speaker's intention.

The simplest abstract measure we can have of the behavior resulting from intention would be a simple arithmetic mean of the actual measured values of some parameter of the behavior over a large number (say, more than fifty) repetitions of the same intention. We hypothesize that the subject intends the mean value as his target each time; that is, he intends invariance, but that inherent limitations of physical performance produce variability about that mean.

The reliability of the arithmetic mean as a quantified value representing the realisation of the subject's intention is dependent on two variables:

- the number of samples measured, and
- the range of the variability itself.

Take, for example, the following data sets representing measurements of any given parameter in arbitrary units:

| A | B |
|---|---|
| 35 | 36 |
| 38 | 20 |
| 34 | 45 |
| 36 | 40 |
| 38 | 52 |
| 35 | 46 |
| 36 | 26 |
| 37 | 26 |
| 36 | 32 |

The arithmetic means for both set A and set B are the same: 36. But what does this number actually mean? There are nine measured tokens in each of A and B. Calculating the arithmetic mean enables us to make some kind of prediction bout what a tenth score might be. In each case we would predict that the next score stood a fair chance of being 36. Notice however that we can be more confident of our prediction in respect of A than B. In actual fact the next score for A could be anything from 34 to 38 (although less likely, a number outside this range), but equally for B could be anything from 20 to 52 since in the B set we already have scores over this whole range. Taking the mean is going to be more useful to us if we can express how confident we are that the next score will be the calculated number.

To do this we must state not just the mean. You can see that saying that A averaged 36 and that B also averaged 36 conceals important aspects of the differences between A and B. So we must also give some indication of the spread of scores obtained. The wider the spread (set B) the less reliable the prediction concerning the next score.

However, in practice in experiments with human beings, as opposed to most experiments in the physical sciences, we get the occasional 'wild' value which might falsify or bias our general inferences. For this reason the measure of spread or range taken is the standard deviation which gives us an indicator of the range of most of the scores, leaving out any that are wildly outside that range.

The mean is calculated as follows:

$$\bar{x} = \frac{\sum x}{n}$$

The standard deviation is calculated as follows:

$$\sigma = \sqrt{\frac{\sum x^2 - n\bar{x}}{n-1}}$$

So, taking our above examples A and B:

$$\bar{x}_A = \frac{\sum x}{n}$$

$= 36.1$

and

$$\bar{x}_B = \frac{\sum x}{n}$$

$= 35.89$

$$\sigma_B = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{12517 - (9 \times 1288.09)}{8}}$$

$$= \sqrt{\frac{12517 - 11592.83}{8}}$$

$$= \sqrt{\frac{924.17}{8}}$$

$$= \sqrt{115.52}$$

$$= 10.75$$

and

$$\sigma_A = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{11751 - (9 \times 1303.21)}{8}}$$

$$= \sqrt{\frac{11751 - 11728.89}{8}}$$

$$= \sqrt{2.76}$$

$$= 1.66$$

These are interpreted as: the mean of the values in column A is 36.1, and the standard deviation indicates a predictable spread of ±1.66 from that value — i.e. from 34.44 to 37.76. The mean of the values in column B is 35.89 and the standard deviation indicates a predictable spread of ±10.75 from that value — i.e. 25.14 to 46.64. In general the greater the standard deviation the less usable the mean as a numerical value for abstract intention of the human being.

When making a comparison of the extent to which two different sets of data contain variability it is necessary to make some adjustments so that the sets are in fact directly comparable. In the example so far the arithmetic means of A and B were virtually the same. Sometimes we may want to consider range of variability in two sets of data where the means are quite different. For example: suppose one set of data has a mean of 25 and a standard deviation of 5, and the other has a mean of 50 and also a standard deviation of 5. Clearly the variability is greater in the former than in the latter because 5 is a greater percentage of the mean. In order to make a direct comparison we normalize the data. In this case this is done by expressing the standard deviation as a ratio with the mean: 5 as a percentage of 25 is greater than 5 as a percentage of 50. The calculation derives what we call the *coefficient of*

$$v = \frac{\sigma}{\bar{x}}.100$$

*variation:*

So, for A:

$$v_A = \frac{1.66}{36.1}.100$$
$$= 4.6$$

and, for B:

$$v_B = \frac{10.75}{35.89}.100$$
$$= 29.95$$

### Difference

Look at the following data sets:

| C | D |
|---|---|
| 78 | 79 |
| 86 | 82 |
| 87 | 70 |
| 70 | 84 |
| 91 | 75 |
| 72 | 86 |
| 84 | 79 |
| 76 | 71 |
| 89 | 74 |
| 75 | 80 |

*Are these data sets actually different?* might be the question we wish to ask. Another way of saying this is: although there is variability in the two sets C and D, did the speaker intend the same event to occur or a different one?

The mean of C is 80.8 and the mean of D is 78. The question we are asking is: is there sufficient evidence in C and D to enable us to predict that the next score for C will not be 78 or of D will not be 80.8? Or, put another way: are the scores for C equally probable for D, and vice versa?

We test for difference between two sets of data like these by application of the *Mann-Whitney U-test.*

*Proceed as follows:*

Rank order the data using rank 1 for the lowest score — in this example the lowest score is 70 (once in C and once in D). When the same score occurs twice or more (that is, when it ties) assign the mean of the rank. So, 70 would be rank 1 if it occurred only once, but since it occurs twice both occurrences are ranked as 1.5 (the mean of 1 and 2). The next score up is 71 (occurs in D), so that is assigned rank 3 (1 and 2 having already been allocated to 70). And so on, to 20 in this case ( (n=10) x 2 ).

Next, sum the ranks for each column to obtain RC = 117 and RD = 93, thus:

| C | | D | |
|---|---|---|---|
| scores | rank | scores | rank |
| 78 | 9 | 79 | 10.5 |
| 86 | 16.5 | 82 | 13 |
| 87 | 18 | 70 | 1.5 |
| 70 | 1.5 | 84 | 14.5 |
| 91 | 20 | 75 | 6.5 |
| 72 | 4 | 86 | 16.5 |
| 84 | 14.5 | 79 | 10.5 |
| 76 | 8 | 71 | 3 |
| 89 | 19 | 74 | 5 |
| 75 | 6.5 | 80 | 12 |
| R = 117 | | R = 93 | |

$$U = n^2 + \frac{n(n+1)}{2} - R$$

Now, using the formula (for columns of equal numbers of tokens):

substitute the larger value of R, to obtain:

$$U = 10^2 + \frac{10(10+1)}{2} - 117$$
$$= 100 + 55 - 117$$
$$= 38$$

Now, turn to Table K III, which will give the critical values of U. n is the same for both columns C and D, so look at the matrix where the row 10 and column 10 intersect. The critical value of U is 23. If our calculated value of U is greater than 23, we can say that any hypothesis that C and D are not different (the null hypothesis) is not refuted (it would have been refuted if our U had been smaller than 23) — i.e. we proceed, by convention, as though C and D are the same. We are confident of this at the 5% level (that is, there are only 5 chances in 100 of error in the assertion).

### Correlation

Sometimes, when obtaining data for two parameters simultaneously or measuring two parameters simultaneously (i.e. on the same data) it is useful to establish whether they correlate — that is, whether when one parameter increases in value so does the other (positive correlation), or when one parameter decreases the other increases (negative correlation).

We test this using the Spearman Rank Correlation Coefficient, $r_S$ . Look at the following data:

| parameter X | | parameter Y | | | |
|---|---|---|---|---|---|
| score | rank | score | rank | $d_i$ | $d_i^2$ |
| 31 | 3 | 79 | 7 | –4 | 8 |
| 40 | 9.5 | 92 | 10 | –.5 | .25 |
| 26 | 1 | 74 | 3 | –2 | 4 |
| 33 | 4.5 | 78 | 5.5 | –1 | 1 |
| 39 | 8 | 82 | 8 | 0 | 0 |
| 40 | 9.5 | 86 | 9 | .5 | .25 |
| 37 | 7 | 77 | 4 | 3 | 9 |
| 33 | 4.5 | 78 | 5.5 | –1 | 1 |
| 35 | 6 | 72 | 1 | 5 | 25 |
| 30 | 2 | 73 | 2 | 0 | 0 |
| | | | | $\Sigma d_i^2 = 48.5$ | |

Rank order the data. within each column, using rank 1 for the lowest scores (be careful of ties). Next, subtract, for each row, the ranks of Y from X to give a number ($d_i$) for each row. Now, square di to obtain for each row $d_i^2$.

$$r_S = 1 - \frac{6\sum d_i^2}{n^3 - n}$$

Now, calculate *r*, using the formula:

So, substituting:

$$r_S = 1 - \frac{6 \times 48.5}{1000 - 10}$$

$$= 1 - \frac{291}{990}$$

$$= +.71$$

The plus sign means positive correlation (a minus sign would have meant negative correlation). 1 indicates maximum (or absolute) correlation, and 0 indicates no correlation. So we have a scale:

-1 --------------- 0 --------------- +1

showing a range for *r* from –1 to +1 through 0. Our value of +.71 indicates a good (though not perfect — that would be +1) correlation between parameters X and Y.

**Summary of Formulae**

Arithmetic mean:

$$\bar{x} = \frac{\sum x}{n}$$

$$\sigma_X = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n-1}}$$

Standard deviation:

Coefficient of variation:

$$v = \frac{\sigma}{\bar{x}}.100$$

Mann-Whitney *U*-test:

$$U = n^1 n^2 + \frac{n(n+1)}{2} - R$$

Spearman Rank Correlation Coefficient:

$$r_S = 1 - \frac{6\sum d_i^2}{n^3 - n}$$

Table I — Critical values of U for a one-tailed test at alpha = .001 or for a two-tailed test at alpha = .002

| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | 0 | 0 | 0 | 0 |
| 4 | | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 3 |
| 5 | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 5 | 5 | 6 | 7 | 7 |
| 6 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 7 | 3 | 5 | 8 | 7 | 8 | 9 | 10 | 11 | 13 | 14 | 15 | 16 |
| 8 | 6 | 6 | 8 | 9 | 11 | 12 | 14 | 15 | 17 | 18 | 20 | 21 |
| 9 | 7 | 8 | 10 | 12 | 14 | 15 | 17 | 19 | 21 | 23 | 25 | 20 |
| 10 | 8 | 10 | 12 | 14 | 17 | 19 | 21 | 23 | 25 | 27 | 29 | 32 |
| 11 | 10 | 12 | 15 | 17 | 20 | 22 | 24 | 27 | 29 | 32 | 34 | 37 |
| 12 | 12 | 14 | 17 | 20 | 23 | 25 | 28 | 31 | 34 | 37 | 40 | 42 |
| 13 | 14 | 17 | 20 | 23 | 28 | 29 | 32 | 35 | 38 | 42 | 45 | 48 |
| 14 | 15 | 19 | 22 | 25 | 29 | 32 | 30 | 30 | 43 | 46 | 50 | 54 |
| 15 | 17 | 21 | 24 | 28 | 32 | 36 | 40 | 43 | 47 | 51 | 55 | 59 |
| 18 | 19 | 23 | 27 | 31 | 35 | 39 | 43 | 48 | 52 | 56 | 60 | 65 |
| 17 | 21 | 25 | 29 | 34 | 38 | 43 | 47 | 52 | 57 | 81 | 88 | 70 |
| 18 | 23 | 27 | 32 | 37 | 42 | 48 | 51 | 50 | 81 | 68 | 71 | 78 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 25 | 29 | 34 | 40 | 45 | 50 | 55 | 80 | 66 | 71 | 77 | 82 |
| 20 | 26 | 32 | 37 | 42 | 48 | 54 | 59 | 65 | 70 | 76 | 82 | 88 |

Table II —Critical values of U for a one-tailed test at alpha = .01 or for a two-tailed test at alpha = .02

| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | |
| 2 | | | | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4. | 4 | 4 | 5 |
| 4 | 3 | 3 | 4 | 5 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 |
| 5 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 6 | 7 | 8 | 9 | 11 | 12 | 13 | 15 | 16 | 18 | 19 | 20 | 22 |
| 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 | 21 | 23 | 24 | 26 | 28 |
| 8 | 11 | 13 | 15 | 17 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 |
| 9 | 14 | 16 | 18 | 21 | 23 | 26 | 28 | 31 | 33 | 36 | 38 | 40 |
| 10 | 16 | 19 | 22 | 24 | 27 | 30 | 33 | 36 | 38 | 41 | 44 | 47 |
| 11 | 18 | 22 | 25 | 28 | 31 | 34 | 37 | 41 | 44 | 47 | 50 | 53 |
| 12 | 21 | 24 | 28 | 31 | 35 | 38 | 42 | 46 | 49 | 53 | 56 | 60 |
| 13 | 23 | 27 | 31 | 35 | 39 | 43 | 47 | 51 | 56 | 59 | 63 | 67 |
| 14 | 26 | 30 | 34 | 38 | 43 | 47 | 51 | 56 | 60 | 65 | 69 | 73 |
| 15 | 28 | 33 | 37 | 42 | 47 | 51 | 56 | 61 | 68 | 70 | 75 | 80 |
| 16 | 31 | 36 | 41 | 46 | 51 | 56 | 61 | 66 | 71 | 78 | 82 | 87 |
| 17 | 33 | 38 | 44 | 49 | 55 | 60 | 66 | 71 | 77 | 82 | 88 | 93 |
| 18 | 36 | 41 | 47 | 53 | 59 | 65 | 70 | 76 | 82 | 88 | 94 | 100 |
| 19 | 38 | 44 | 50 | 56 | 63 | 69 | 75 | 82 | 88 | 94 | 101 | 107 |
| 20 | 40 | 47 | 53 | 60 | 67 | 73 | 80 | 87 | 93 | 100 | 107 | 114 |

Table III — Critical values of U for a one-tailed test at alpha = .025 or for a two-tailed test at alpha = .05

| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | |
| 2 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 |
| 4 | 4 | 5 | 8 | 7 | 8 | 9 | 10 | 11 | 11 | 12 | 13 | 13 |
| 5 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 20 |
| 8 | 10 | 11 | 13 | 14 | 16 | 17 | 19 | 21 | 22 | 24 | 25 | 27 |
| 7 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 |
| 8 | 15 | 17 | 19 | 22 | 24 | 26 | 29 | 31 | 34 | 36 | 38 | 41 |
| 9 | 17 | 20 | 23 | 28 | 28 | 31 | 34 | 37 | 39 | 42 | 45 | 48 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 20 | 23 | 28 | 29 | 33 | 36 | 39 | 42 | 45 | 48 | 52 | 55 |
| 11 | 23 | 26 | 30 | 33 | 37 | 40 | 44 | 47 | 51 | 55 | 58 | 62 |
| 12 | 26 | 29 | 33 | 37 | 41 | 45 | 49 | 53 | 57 | 61 | 65 | 69 |
| 13 | 28 | 33 | 37 | 41 | 45 | 50 | 54 | 59 | 63 | 67 | 72 | 76 |
| 14 | 31 | 36 | 40 | 45 | 50 | 55 | 59 | 64 | 67 | 74 | 78 | 83 |
| 15 | 34 | 39 | 44 | 49 | 54 | 59 | 64 | 70 | 75 | 80 | 85 | 90 |
| 18 | 37 | 42 | 47 | 53 | 59 | 64 | 70 | 75 | 81 | 86 | 92 | 98 |
| 17 | 39 | 45 | 51 | 57 | 63 | 87 | 75 | 81 | 87 | 93 | 99 | 105 |
| 18 | 42 | 48 | 55 | 61 | 67 | 74 | 80 | 86 | 93 | 99 | 106 | 112 |
| 19 | 45 | 52 | 58 | 65 | 72 | 78 | 85 | 92 | 99 | 106 | 113 | 119 |
| 20 | 48 | 55 | 62 | 69 | 76 | 83 | 90 | 98 | 105 | 112 | 119 | 127 |

Table IV — Critical values of U for a one-tailed test at alpha = .05 or for a two-tailed test at alpha = .10

| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | 0 | 0 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
| 3 | 3 | 4 | 5 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 11 |
| 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 10 | 17 | 18 |
| 5 | 9 | 11 | 12 | 13 | 15 | 16 | 18 | 19 | 20 | 22 | 23 | 25 |
| 6 | 12 | 14 | 16 | 17 | 19 | 21 | 23 | 25 | 26 | 28 | 30 | 32 |
| 7 | 15 | 17 | 19 | 21 | 24 | 26 | 28 | 30 | 33 | 35 | 37 | 39 |
| 8 | 18 | 20 | 23 | 28 | 28 | 31 | 33 | 36 | 39 | 41 | 44 | 47 |
| 9 | 21 | 24 | 27 | 30 | 33 | 36 | 39 | 42 | 45 | 48 | 51 | 54 |
| 10 | 24 | 27 | 31 | 34 | 37 | 41 | 44 | 48 | 51 | 55 | 58 | 62 |
| 11 | 27 | 31 | 34 | 38 | 42 | 46 | 50 | 54 | 57 | 61 | 65 | 60 |
| 12 | 30 | 34 | 38 | 42 | 47 | 51 | 55 | 60 | 64 | 68 | 72 | 77 |
| 13 | 33 | 37 | 42 | 47 | 51 | 56 | 61 | 65 | 70 | 75 | 80 | 84 |
| 14 | 36 | 41 | 46 | 51 | 56 | 61 | 66 | 71 | 77 | 82 | 87 | 92 |
| 15 | 39 | 44 | 50 | 55 | 61 | 66 | 72 | 77 | 83 | 88 | 94 | 100 |
| 16 | 42 | 48 | 54 | 60 | 65 | 71 | 77 | 83 | 89 | 95 | 101 | 107 |
| 17 | 45 | 51 | 57 | 64 | 70 | 77 | 83 | 89 | 96 | 102 | 109 | 115 |
| 18 | 48 | 55 | 61 | 68 | 75 | 82 | 88 | 95 | 102 | 109 | 116 | 123 |
| 19 | 51 | 58 | 65 | 72 | 80 | 87 | 94 | 101 | 109 | 116 | 123 | 130 |
| 20 | 54 | 62 | 69 | 77 | 84 | 92 | 100 | 107 | 115 | 123 | 130 | 138 |

## APPENDIX B

[*REMINDER:* WRITTEN IN 1984!]

### Computing

Many of the tasks performed by hand and some performed by the equipment itself in this booklet can be done by computer. Even the smallest machine will enable us to do most of the statistical tasks faster and easier than by hand, and at the other end of the scale a large scientific computer enables detailed and accurate analysis of an *acoustic waveform* to provide data similar to that produced by the *sound spectrograph*. Computers are only able to store and process numbers, so in the case of acoustic analysis the signals coming from a microphone have to be converted from their original analog form into an equivalent series of numbers. This is done by means of an analog-to-digital (A/D) converter. Any of our data can be input to a computer in this way.

An A/D converter works by dividing the time parameter of the signal into small slices and measuring the amplitude of the signal during each consecutive slice. A good example to enable us to understand what is going on might be an amplitude processed emg signal:

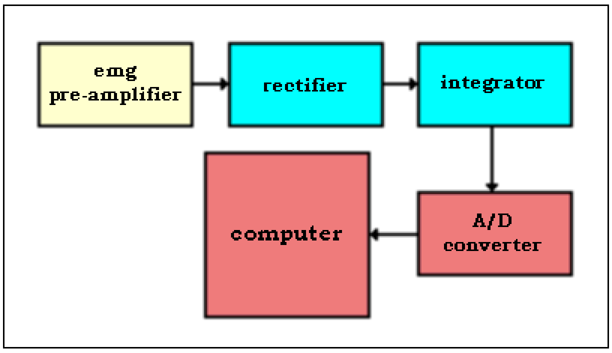



Fig. 38 — Amplitude processing of an emg signal.

The emg amplitude curve is examined in a manner analogous to placing a *grid* over the signal as it appears on an oscilloscope or chart recorder. The grid has two axes: *time and amplitude*.

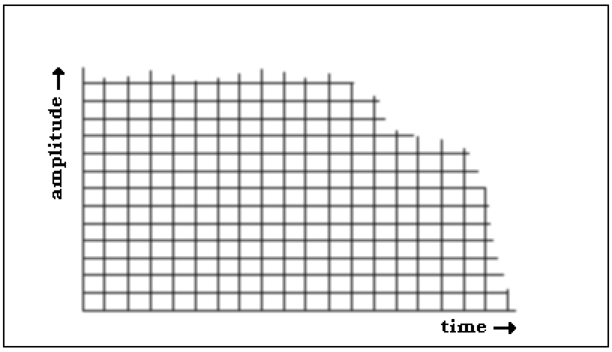



Fig. 39 — Time *vs*. amplitude grid.

The computer operator can select the actual scale on the two axes to be used. For our example we'll choose a vertical amplitude division to have the value 1, giving us a scale of 1–25 on the y–axis. For the x–axis we'll say that each division equals one one-hundredth of a second, so the first time slice will be the first 1/100s, the next the second 1/100s, the next the third 1/100s, and so on.

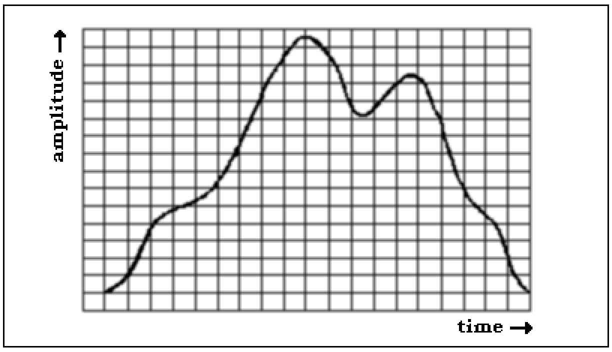



Fig. 40 — Processed emg signal superimposed on the time/amplitude grid.

By examining the curve with the superimposed grid the A/D converter makes a series of amplitude readings, one for each 1/100s. These numbers are then passed to the computer which sets up a table in its memory for storing the readings:

| Time slice | Amplitude reading |
|---|---|
| 1 | 0 |
| 2 | 1 |
| 3 | 4 |
| 4 | 5 |
| 5 | 5 |
| 6 | 6 |
| 7 | 8 |
| 8 | 11 |
| 9 | 13 |
| 10 | 15 |
| . | 13 |
| . | Etc. |

We now have a numerical version of the earlier analog emg signal, since the computer has been told that each time slice is equal to 1/100s.

As a simple example of the kind of task the computer can perform on this data, note measurements of peak amplitude of emg signals. Refer to *Experiment 8*. Here we look at the chart-recorder curve, find the highest point (the point of greatest amplitude) and measure down to the baseline to determine its value. The computer simply scans the table it now has in its memory: the highest amplitude reading is the value of the peak emg during this sample. The computer can also tell us that this peak occurred during a particular time slice, or so many 1/100s's into the signal1

By means of a D/A (digital-to-analog) converter operating like a reverse A/D converter the table of data can be pulled out of the computer's memory and be reconstituted as a graph:

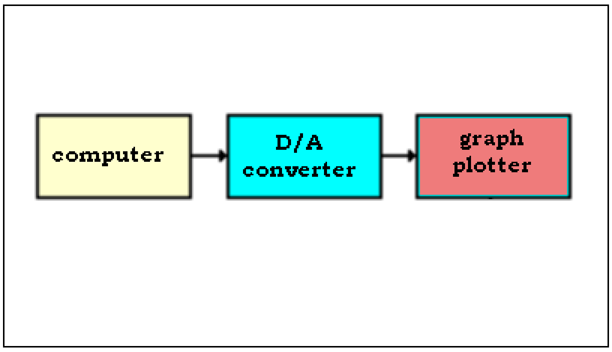



Fig. 41 — Plotting the stored data.

The pen of the graph plotter is moved in increments along the x–axis corresponding to the value of each time slice in the table, while at the same time positioned vertically along the y–axis a distance corresponding to the amplitude value for that time slice found in the table. A graph like this is produced:

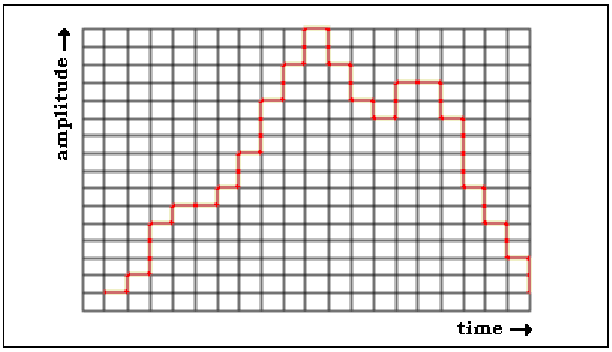



Fig. 42 — Final graph of the computer stored data.

Notice that the graph, while following the original curve, is lacking in the smooth detail of the original. More detail can be incorporated by making the cells of the original A/D conversion finer — that is, by taking an amplitude reading every 1/1000s instead of every 1/100s, and dividing the amplitude scale into 100 levels rather than 25.

### Programming

An obvious use of computers is to perform the tedious arithmetic operations in statistical tests. We have seen above how measurements can be made automatically by the computer by using an A/D converter. Those measurements can be made on a series of data samples in much the same way as we make hand measurements from our chart recordings, and tables can be set up in the computer's memory. The numerical data can also be fed into the computer by hand from the keyboard. That is, we perform the function of the A/D converter to provide the computer with some numbers to process.

The computer will need to know how to process the numbers, just as we need to know how to perform a statistical test. In Appendix A — *Statistics* you find that the arithmetic mean is calculated using the formula:

$$\bar{x} = \frac{\sum x}{n}$$

In a sense, that formula tells you how to calculate the mean. Similarly the computer has to be told what tasks to perform on the data in its memory. The instructions are set out in the form of a program or procedure consisting of a step-by-step description of what the computer must do with the data.

Programs are written in languages specially developed for the task. There are basically two types of programming language: *procedural languages* like FORTRAN, ALGOL, PASCAL, BASIC, etc., and descriptive, or *declarative languages* like LISP or PROLOG. A procedural language takes the computer step-by-step through a series of tasks it must perform. For example: 'calculate x by multiplying y and z together' (x = y . z), or 'replace the string S with the sequence of two strings NP and VP' (S → NP VP). A declarative language is used to tell the computer how pieces of data relate to one another in a logical fashion. For example: 'John is the brother of Sue who is Anne's daughter'. Each type of program enables us to get the computer to tell us the answers to questions. So, we can ask: 'What is the product of 13 and 6?' or 'What is the relationship of John to Anne?'. Answers: '78' and 'son' respectively.

Procedural languages are most useful for getting the computer to perform tasks involving *numerical computation*, whereas declarative languages are useful when it comes to examining consequences of *logical relationships* in semantics.

For our statistical tests we shall use the procedural language BASIC since this is the most widely used language. It is important to know, though, that, just as with human languages, programming languages have dialects and the details of a program in BASIC can vary from computer to computer.

Take, as our first example, the simple calculation of the arithmetic mean. The numbers we have obtained from our measurements must be got into the computer's memory. It must be programmed to acquire or input those numbers via its keyboard. The simple instruction `INPUT X` will enable the computer to acquire a single piece of data. Repeating this instruction in a loop for the number of times that we have pieces of data will enable it to input a set of data to form a table. So, the first thing the computer must do is ask us how many data tokens there are:

```
PRINT "How many tokens";
INPUT N
```

'How many tokens' is put up on the screen as the program interacts with the user. The ? comes up whenever there is an `INPUT` instruction, so we don't need to put it in the text to be printed to the screen.

The operator will input the number of tokens from the keyboard and the computer will assign this number to the variable `N`. Now the computer knows how many times to loop the input of the individual pieces of data: it's `N` times,

So:

```
FOR J = 1 TO N
INPUT X
NEXT J
```

What this says is: using a counter `J`, start with a value of 1 and input the first piece of data `X`. Now go to the next `J`, which must be 2 and input the next piece of data. In other words the program loops back from the third line to the first line as many times as it takes to increment the counter `J` from 1 to the value of `N` (the number of tokens) — each time inputting a token value `X` from the keyboard.

In the calculation of the arithmetic mean we need to have the computer add up all the token values. This could be done once they are all in the computer's memory, just as we add them up once we've made all the measurements. An easier way from the computer's point of view, however, is to add them up as it goes along. That is, to perform the addition as part of the loop:

```
FOR J = 1 TO N
INPUT X
```

```
S = S + X
NEXT J
```

Here a variable accumulator `S` is having added to it the value of `X` each time the program goes round the loop. To explain this let's have some actual numbers. Say our readings are: 7, 4, 9, 6. Four pieces of data, so `N=4, S` begins by being equal to zero.

So the first time round the loop `X` equals 7 and `S` is incremented by 7 — i.e., is made equal to 0+7. The second time around `X` equals 4 and this number is added to the current value of `S` to give 11(7+4). Next time around `X`=9 and S becomes 20 (11+9). The fourth time around X=6 and S becomes 26 (20+6). `J` now equals `N`, so the loop stops and we have `S=sumX` or 26.

The next step: let `M` be the mean. Then:

```
M = S / N   (/ means 'divided by')
```

The complete program would look like this:

```
10 CLS   [clear the screen]
20 S = 0  [set the accumulator to zero]
30 PRINT "How many tokens";
40 INPUT N
50 FOR J = 1 TO N  [data acquisition loop]
60 INPUT X
70 S = S + X  [accumulate the sum of X]
80 NEXT J
90 M = S / N  [calculate the mean]
100 PRINT M   [display the result]
110 END
```

Notice that the lines are numbered so that the computer knows the sequence the operations are to be performed in. The numbering here goes up in tens so that we can add lines between if we want to.

The following short program is not explained here in detail. It will be used in class when you come to evaluating some of your measurements:

Program to calculate the arithmetic mean, the standard deviation and the coefficient of variation of a data set

```
10 CLS
20 PRINT "How many tokens";
30 INPUT N  [size of data set]
40 CLS

50 S = 0  [set accumulators to zero]
60 Q = 0

70 FOR J = 1 TO N  [loop to acquire data]
80 PRINT "No."; J;
90 INPUT X  [acquire the data]
100 S = S + X  [sum the data]
110 Q = X^2  [sum the squares of the data]
120 NEXT J

130 CLS

140 M = S / N  [calculate mean (M)]
```

```
150 D = SQR ((Q – N x M^2) / (N – 1))  [calculate std. dev. (D)]
160 V = D / M x 100  [calculate coef. of var. (V)]

170 PRINT "Mean:"; M  [print results to screen]
180 PRINT "St. Dev.:"; D
190 PRINT "Coef. Of Var.:"; V

200 PRINT
210 PRINT "Another data set (Y/N)";  [loop to start or exit]
220 INPUT N$
230 IF N$ = "Y" THEN GOTO 10

240 END
```