What's in a Voice?

An Ecological Approach to Non-Verbal Information from Speech

by Andy Wills

KEYWORDS: Speech, Voice, Ecological, Age, Sex, Height, Geographical area, Social Class

Presented for B.Sc. (Social Sciences) March 1993

ACKNOWLEDGEMENTS

Many people should be thanked for the help they have provided. In particular I wish to thank Alan Costall and Chris Colbourn for their advice, support, insight and patience. Thanks should also go Brian Newman, Martin Hall, Tina and Erica for their humour, and their help with the technical side of things; and to Violet Bailey and Amina Memon in helping me find thirty participants in such a short space of time! Finally, I would like to thank Rebecca Leyland, Alison Evans and the occupants of 98 Mayfield Road for putting up with my moods, scatterbrain-ness, and for their considerable moral support.

CONTENTS

ABSTRACT

In a very obvious sense, speech provides information about the speaker ie they can simply tell us their age, gender etc! The central hypothesis of this study was that information is also available on a different level - that of non-verbal vocal information. Sixteen voice recordings were collected from eight speakers, with each speaker providing two speech episodes - the reading of a set passage and the description of an object. A selection of these recordings were played to eleven participants (listeners) such that each episode was heard a total of four times. By a method of semi-structured interviews after each voice recording, the number and range of judgements listeners were willing to make about speakers was investigated. In addition to answering set questions about, among other things, age, sex, height, weight, intelligence and attractiveness, the listeners (as a group) provided between 10 and 30 other judgements about each speaker. These ranged from their typical dress and the colour of their hair to, in more than one case, their religious denomination! In the second part of the study, the most common of these judgements were investigated further. The object description episodes were dropped due to practical difficulties and time constraints. A (generally multiple choice) questionnaire was devised and presented to thirty-one participants in conjunction with the voice recordings. In this stage, all participants heard all set passage episodes, and all answered the same questions on them. It was found that ***age, sex*** and ***height*** appeared to be accurately specified. Some limited accuracy was also found in judgements of ***area of origin, marital status, sports participation, hair colour and length and gross earnings.*** Age may be implicated in judgements of hair colour and marital status, and gender in judgements of hair length. Gender may also be a factor in judgements of gross earnings and height, although there appears to be some gender-independent information. Nevertheless, the presence of accurate judgements suggests there is some information available for these judgements, whatever the source of that information may appear to be. The reasoning behind this stance, and its implications, are discussed. Further, it is suggested that speech provides non-verbal information for at least five accurate, independently specified variables. These variables were: ***combined physical and social age, combined physical and social gender, height, geographical tenancy relative to the listener*** and ***social position.*** Theoretical implications, limitations of the current study, and possibilities for further research are also discussed.

INTRODUCTION

Some definitions

In all of this report, the words vocal and verbal will be used extensively. In everyday usage these are often treated as interchangeable but here they have very specific meanings. Verbal means specifically of words, whilst vocal simply means of the voice.

Representationalism, The Ecological Approach and The Present Study

As the title indicates, the author takes an ecological perspective on perception. In contrast to representational theories, the ecological perspective considers that the environment provides sufficient information to specify itself. Perception need not be (and isn't) dependent on the construction of meaning from impoverished cues via 'previous knowledge'. The organism is considered to be an active agent in the environment that can 'pick-up' information from the perceptual array.

However, as we shall see in the discussion, it would be possible to interpret the findings of the present study from either a representational, or an ecological, viewpoint. In what sense, then, does this study embody an ecological approach? It is described as ecological because of what it studies rather than what it proves. The

representationalist perspective tends to go hand in hand with certain experimental methods and certain questions. The ecological approach asks different questions and often uses different types of experiment. It is not necessary to give a full review of all the differences. Instead, I will briefly discuss two areas relevant to the current investigation - the traditional view of sound and an ecological perspective on vision. Following this, ecological acoustics will be introduced.Ecological vision was chosen to provide the initial examples because it is a more mature field, and it contains some examples that parallel the aims of the present investigation.

Traditional Questions and Experiments in Sound

Traditional investigations of sound have concentrated either on speech or on tonal sounds. Brief, isolated and simplified sounds are generally utilised. In studies of speech, phonemes or syllables are the main units of investigation[1]; in non-speech, short, spectrally stable sounds. Yet such sounds, as they are presented, are uncommon in the everyday environment - one seldom hears a disconnected phoneme, for instance. Hence, there appears to be a large disparity between what is studied, and what is actually heard.

Nevertheless, the decision to maintain this style of investigation appears to be a deliberate one, and one driven by a particular (reductionist) philosophy. The argument is that, if we understand these sounds, which

---

[1] The are not the only ones, however. In particular, work on prosody (see later) is interesting.

are basic and fundamental, we will know more about the perception of other sounds. Two pieces of evidence in particular are taken to support this view - the work of Fourier, showing that any sound can be reduced to a number of sines waves, and some work that suggests the ear can perform such a reduction. In a similar way, that words can be synthesised from phonemes is taken as evidence that phonemes are basic units.

This is a reasonable, and internally valid, way to investigate the problems. However, it is not the only way. Just because such fundamental units exist (or even that the ear may be able to isolate them) does not mean that they are *necessarily* psychologically significant. There may be other, possibly more reliable and informative, sources of information. Some examples from ecological vision may help to clarify this point.

Some Questions and Experiments in Ecological Vision

The vast majority of work in vision has also concentrated on simplified, isolated and unnatural stimuli. However, some research, which may be generally labelled as ecological, has attempted to re-dress the balance. With varying degrees of success, researchers have attempted to use more natural situations and consider more meaningful, or non-'basic', sources of information. Compromises are often made to facilitate interpretation, but the aim is to make the experimental setting and the everyday world of the participant more similar .

Although it is not appropriate to review the field here, it is worth mentioning some studies that parallel the purpose of the present investigation. Perhaps the most famous is the work of Shaw & Pittenger (1975), who demonstrated that participants could accurately judge age from skull profile . Further, this ability was found to be linked to the perception of a transformational invariant[2] , and the manipulation of the value of this invariant produced predictable changes in the judgement of age. Line drawings were originally used, but in later studies these results were replicated using realistic three dimensional models (Mark & Todd 1983).

The group of studies that followed Johansson's original work are also interesting. Kozlowski & Cutting (1978), for instance, showed that participants performed at better than chance levels in judging the sex of a walker from a point-light display of their body profile (all that can be seen here are small lights placed at the walker's joints). In an similar vein, Bassili (1979) showed that participants were reasonably good at judging expressed emotion from a point-light display of a face (where the dots were placed semi-randomly).

The informative sources used in these studies are not entirely natural, but they do demonstrate that the human body can specify more than shade and outline form. Reliable, accurate information (such as age and gender) that is useful for the direction of behaviour can be picked-up

[2] The term transformational invariant comes from Gibson (1966,1979). It refers to some property of the optic array that is constant through change.

even when the body image is seriously degraded. By analogy, one could suggest that auditory information also provides useful, reliable and accurate information. The investigation of this assumption is central to ecological acoustics, and this is the field of study to which we now turn.

An Introduction to Ecological Acoustics

It is easy to forget that ambient light is not our only source of information. When two large objects collide, we would normally not only see, but also hear, the crash. In fact, if we were not looking in the right direction, we would be alerted to the crash by the sound alone and would then (probably) turn to look. Evidence suggests that people can locate the source of a sound quite accurately (Makous & Middlebrooks 1990).

Sound can also specify information about the *nature* of events, rather than just 'look, something has happened!'. This can be readily observed in everyday life - if someone drops a glass in another room, you can easily tell whether it has smashed or bounced. Experimental evidence confirms this common-sense hypothesis (see Warren & Verbrugge 1984). Further, the 'background hum' of everyday life seems to indicate to us 'all is well'. One can observe this by walking into a totally sound proof room - a rather unnerving experience.

However, even given the evidence of everyday life, many people would still assert that they can tell nothing, or very little, from sound alone.

However, experimental evidence would suggest otherwise. VanDerveer (1979) showed that people could accurately identify common sounds and events (eg keys jangling) in a free description task. All 30 sounds tested showed good agreement between listeners and a high accuracy of judgements (most objects or events were correctly identified in 80-95% of cases - the lowest accuracy was 66%). McNaughten (1987) , developing VanDerveer's original work, found that people could distinguish between recorded rolling and sliding events. Further, participants performed at better than chance levels in identifying the shape and material (eg metal, plastic etc) of a dropped object in a forced-choice test.

Bill Gaver's current work (Gaver, in press) is also interesting. He suggests that we should classify sounds by the events they specify. For instance, dripping, pouring and splashing are all different types of 'liquid' events, whilst wind and explosions are types of 'gas' events. His work is intimately related to practical applications in human-computer interaction. Gaver has developed auditory versions of icons ('earcons'!) to specify, for instance, how big a particular computer file is. A large file makes a 'thump' when selected[3], an empty file makes a 'tock' (see Gaver 1989). Preliminary reports indicate that users find this both useful for certain operations.

[3] An icon is a small picture that represents a program, piece of text etc.

Ecological Acoustics and Speech

As we can see above, ecological acoustics has made some progress in investigating the sounds of our everyday environment and yet, in one sense, it is a step back from traditional investigations. Speech, one of the most common everyday sounds, has received *no* attention whatsoever. Admittedly, this can largely be seen as an intentional move - the perception of speech is something of a burgeoning area and, until the work of people like VanDerveer and McNaughten, the information in any sound *apart* from speech had received scant attention. However, a justified concentration on non-verbal information has also meant a concentration on non-vocal information. These two areas are quite separate. It remains, at the very least, a logical possibility that speech contains non-verbal information about the speaker. Given this, it is perhaps surprising that (as far as the author could ascertain[4]) there has been no work in ecological psychology on this subject.

The Contribution of Acoustics and Phonolgy

A search for published work led to the disciplines of acoustics and phonology. However, non-verbal speech *information* receives scant attention in both these areas. Considerable work has been done on the acoustic characteristics of the voice,but most of it addresses the effects of disease (eg laryngitis), or the production and identification of phonemes and vowels.

[4] An extensive search of the relevant journal and the use of the 'BIDS' on-line search system revealed no articles.

Two notable exceptions are Ryan (1972) and Kirchener (1970). Ryan identifies a number of changes in the voice with age. For instance, old people (defined here as over 70) have a lower vocal intensity, and a lower mean w.p.m. (words per minute). Kirchener found that male speech tends to have a lower fundamental frequency than female speech - an interesting, if common-sense, result. Other information can also specify sex-differences, for instance some languages (eg Thai) have different lexical sets for male and female speakers (Haas 1964). Other studies have looked in considerable detail at changes in the vocal tract with age and sex. Hoit & Hixon (1992) showed that laryngeal airway resistance changes with age, and Swartz's study (1992) suggests there are sex differences in voice onset time.

What all the studies described so far fail to address is the ecological significance of such information. It may be interesting to illuminate the difference between people in the production of speech and noise but, on its own, such work is of little use to the study of human perception. What needs to be discovered is whether people actually *use* such information. To be fair, these investigations are not really central to acoustics or phonolgy. However, one might expect them to be central to psycholinguistics.

Psycholinguistics and the 'Stereotype Principle'

Studies in psycholinguistics (Palmer (1973) among others) strongly suggest that listeners form definite impressions of the people they hear on audio recordings. Further, they seem prepared to make a number of

judgements about them on the basis of such recordings. Ryan, Carranza & Moffie (1977), for instance, found that the personality traits attributed to non-native speakers of English (Spainards in this case) depended on their level of accentedness. Further, Frazier (1973) concluded that black and white American English speakers were rated more positively on a number of scales if their accent was seen as typical of their perceived race. Ryan & Capadonno (1978) found that old female speakers were rated as more reserved, passive, 'out of it' and inflexible than younger female speakers.

This appears to be a promising lead but, on closer examination, turns out to be a little disappointing. Most of the writers in this area seem to be content to show just how many judgements a listener will make about a speaker, without progressing to assess whether or not these judgements are accurate. The purpose appears to be the provision of yet more evidence for an already well-established principle. Namely, people will, at least in experimental settings, make many judgements on the basis of what appears to be a minimal amount of information.

Non-Verbal Speech Information

In their attempts to uphold this 'stereotype principle', the authors provide some evidence that the voice *can* actually provide accurate information about the speaker. Ryan & Capadonno (1978), in their study of stereotypical personality attributions, found a good correlation

between the estimates of the speaker's age and their actual age. Sachs, Lieberman and Erickson (1973) showed that listeners were over 80% accurate in judging the sex of pre-pubescent speakers from just one sentence.

There are also an number of interesting studies that fall somewhere between acoustics, phonolgy and psycholinguistics. VanDommelen (1990) found that people could recognise speakers previously known to them simply by recordings of their repeated utterance of the 'ma' syllable. Schwartz (1968) found that (in English) male and female speakers were discriminable even when the fundamental frequency had been removed from their speech[5].

Perhaps the most intriguing publication is that of Lass et al (1979). The question they attempted to investigate was whether one could judge a speaker's height and weight from a recording of their voice, and whether certain manipulations of this recording would affect this ability. Unfortunately, their interpretation of the results was fundamentally flawed. Their index of accuracy was the difference between the mean height of all the speakers and the mean judgement of height of all the listeners. Hence what they were really assessing was the listeners ability to the average height and weight of the general population! Without access to their raw data it is difficult to get any

[5] This is a result that can also been seen, to a certain extent, in everyday life.
The nature of the telephone system means that the fundamental frequency of most male speakers is not transmitted - yet we seldom have problems distinguishing the sex of a speaker in this situation.

meaningful results from this study, but their summary graphs suggest relatively *inaccurate* mean estimates of height and weight for the individual speakers (some errors were as large as 10 inches and 40 lbs).

Even given the occasional analysis error, what one can see appearing from this group of disparate studies is the idea that speech can provide useful and reasonably accurate information about the speaker. This is obviously true in a linguistic sense - the person that says 'I am fifty four years old' is (probably!) providing accurate information about his or her age. However, it also appears to be true in a more fundamental sense. Speech appears to specify information about the speaker in a non-verbal form.

A Note On Prosody

The study of prosody (a term from linguistics) would suggest that speech can convey subtle aspects of meaning and emotional content via changes in tone, amplitude and rhytmn. This would appear to be one area that has touched upon non-verbal speech *information* , and one I believe would benefit from more extended psychological investigation. Whilst this area is differentiable from non-verbal information about *stable characteristics*, it is possible that some of this form of information could be prosodic - for instance there may be characteristic changes in voice rhythm with age. However, this possibility is not investigated by the empirical work described in this study, but the issue is returned to, indirectly, in the discussion.

Introduction to the Initial Study

Very little work has been done, in any area, on non-verbal speech information. In ecological psychology, as far as the author could ascertain, there is no published work. The basic question I wanted to ask was 'what can you tell about the speaker from his or her voice?' As seen earlier, parallel questions have been asked about visual information and the results, in some cases, have been quite surprising. I therefore decided that the best initial move would be to investigate just what sort of information people *thought* they could pick-up from the human voice. In a sense, I wanted to see how far they would go. The basic idea was to record some people speaking and then to play these recordings to a different group of people.

The voice recordings were used in the initial study were pressed back into use for the follow-up. It is for this reason, and for reasons of clarity, that their collection is described separately.

VOICE RECORDINGS

METHOD

Design

Thirteen participants (hereafter the 'speakers') were asked to read a set passage, and then describe an object handed to them. These two vocalisations were recorded separately on audio tape, giving twenty-six speech 'episodes' in all. Each speaker was also asked to provide some biographical information (their age, sex and so forth) which was recorded on printed sheets. Permission was sought to obtain more biographical information from them if this was found to be necessary. All participants concurred and, in eight cases, further information was requested some weeks later.

Participants

All speakers were employees or students of the University Of Southampton. All volunteered, and none were paid for their participation. The sample was a convenience one, although it was ensured that there was a range of ages (20-63 years) in both sexes. No range in any other variable was explicitly sought.

Materials & Apparatus

Voice recordings were made in a sound-attenuating but echoic room, on a standard, high-qualitycassette recorder with a good microphone and audio tape (TDK AR-90). Recordings were monitored on high-quality,

low-leakage headphones (Sennheisner HD450). All instructions for the participants, and the set passage and the biographical information questionnaires, were presented on clear, large-print, laser-printed A4 sheets. An example set can be found in Appendix I.

The set passage was taken from Feynman's (1988) "What do you care what other people think?" and selected on the criteria of being easy to read and having little manifest content[6]. The object for description was a small, simply shaped and decorated, aromatherapy oil burner (minus candle and oil). This object was used because, it was assumed, few people would know what it was, and would therefore have to speak for some while to describe it.

Procedure

Ethical approval for obtaining voice recordings (and subsequently playing them to participants) was requested and obtained from the department of psychology's ethical board.

The speaker was invited in to the sound-attenuating room and asked to sit down on a set chair. This gave a reasonably constant distance to the microphone, which was mounted on a fixed table-top stand. No adjustments of recording level were made, but no participant spoke so loudly or softly as to cause problems of distortion or loss. The experimenter was in the room at all times.

---

[6]Apologies are due here to any particular fans of Feymans's meandering prose style!

The speaker was handed the printed instruction sheets and asked to read the first page. It was checked that he or she understood the instructions and then (s)he was asked to read the set passage. This vocalisation was recorded and the tape-recorder stopped.

Next, the speaker was handed the oil-burner and asked to describe it. He or she was told that, if they knew what it was, they could name it but they should also describe its physical properties. This second vocalisation was recorded and the tape recorder stopped.

Finally, the speaker was asked to fill in the biographical questionnaire. The following information was requested from participants: -

- age
- sex
- whether they smoked
- any medical condition that might affect their voice
- the area of the country they have spent most of their life in
- a self-description of their accent
- a self-report of their general emotion state when making the recording
- any other information they thought I might find interesting

The speaker was thanked for their participation and, if requested, the purpose of the investigation was described in more detail. Further, all participants were told how to contact the experimenter if they were interested to find out the study's results.

The last question of the biographical questionnaire sought permission to request further information, and a space was left for their contact address. For reasons detailed later in this report, further details were requested from eight consenting participants via the second biographical questionnaire (see end of appendix I). The information requested included:-

- Height, weight, hair and eye colour, hair length, facial hair, skin colouring and shoe size.
- Job, salary and education
- Accommodation type and location, car ownership and marital status
- Frequency and type of social and sport activities.
- Typical dress.

This may seem a little odd and excessive but was driven by the results of the initial study (see following section).

RESULTS

All twenty-six speech episodes recorded well, although some did include a few incidental noises (such as bumps, clicks and so forth). No attempt was made to edit these out. The audio recordings are not presented here for obvious reasons but they can be obtained from the author[7].

All participants volunteered the biographical information, and all agreed to provide more information if it was needed. All did so when this was requested. Biographical information for all participants is presented in summary in appendix II, although names and contact addresses are obviously omitted. No speaker reported any voice problem more serious than a slight cold, and only two reported smoking (on an occasional basis).

[7] I can be contacted at 98 Mayfield Rd, Swaythling, Southampton or the Department of Psychology, University Rd, Southampton.

INITIAL STUDY

METHOD

Design

Eight speakers (and hence sixteen speech episodes - one of a set passage and one of an open-ended object description) were selected semi-randomly from previously made voice recordings. The selection criterion was that the speakers, as a group, gave a good range of ages in both sexes. Both 'set passage' and 'description' episodes were used in order that these two sources of information could be compared. From the sixteen episodes, a selection of six were played to eleven participants (hereafter called 'listeners').The selection differed from listener to listener in order that each episode could be presented a total of four times with no speaker appearing twice in the same selection. After hearing each episode, the listener was asked about their impressions of the speaker. Some set questions were always asked, but opportunities were given for the listener to volunteer information. The topics (but not the content) of volunteered information were often used to prompt subsequent listeners. Questions to which a particular listener could provide no information were not pursued with that listener. None of the listeners had participated in the production or collection of the voice recordings.

Participants (Listeners)

The participants were friends and associates of the experimenter and, at the time of the study, all lived in the Southampton area. There were roughly equal numbers of males and females, and all were aged between 19 and 26.

Materials & Apparatus

Voice recordings were played back on the machine they had been recorded with via Sennheisner HD450 high-quality headphones. A selection of sixteen speech episodes taken from the original recordings were used. The selection contained both 'set passage' and 'description' episodes. The participants' responses were recorded on specially prepared, pre-printed sheets (see appendix III). No standard room was used for playback.

Allocation Procedure

The six recordings a particular participant would hear were pre-determined by an allocation sheet (see appendix V). The organisation of this sheet allowed each episode to be presented a total of four times without any one listener hearing the same speaker twice. Listeners heard both 'set passage' and 'description' episodes.

Procedure

This procedure was approved by the department's ethical board, under the application detailed in the 'voice recordings' section.

The experimenter sought permission from volunteers, warning that participation may involve half to three-quarters of an hour of their time. Participation took place in a number of settings (the experimenter's room, participants' houses and so forth) but all reasonable measures were taken to ensure a quiet, undisturbed environment.

The participant was handed a printed instruction sheet (see appendix IV) which encouraged them to relax, listen to the recording, and report any impressions they had of the person speaking. The sheet also included some questions such as 'How old do you think they are?' and 'How tall do you think they are?'. These were intended to give the participant some initial ideas on how to approach the task. However, it was stressed to the participant that this was the only purpose of the questions, and the experimenter took the printed sheet back as soon as the participant had read and understood it.

The participant put on the headphones and listened to the first speech episode. Following this, he or she was invited by the experimenter to remove the headphones and discuss the person they had just heard with them. Although the form of the discussion varied slightly from participant to participant, it tended to follow the same basic pattern:

The experimenter would first ask some set questions (in no particular order). These were chosen on something of an arbitrary basis. Some (such as age, sex and height) were borrowed from the acoustics,

phonology and psycholinguistics literature. The rest were simply questions that the experimenter thought it might be interesting to ask. The set of questions asked were as follows:-

- "How old do you think the speaker was?"
- "Were they male or female?"
- "How tall were they?" - An estimate in feet and inches was encouraged.
- "How heavy are they?" - No particular measurement system was encouraged.
- "What area of the country do you think they come from?"
- "What job might they have?"
- "How would you describe their general emotional state?"
- "How intelligent are they?" - The use of a 5 point scale was encouraged
- "How attractive are they?" - A 5-point scale was again encouraged.

If a particular participant could consistently provide no information for a question, it was dropped for that participant.

Next, the experimenter encouraged the listener to describe any other impressions they had of the speaker. As more participants were interviewed, the experimenter began to use descriptors other listeners had used to prompt the current listener. Hence a 'snowball' effect was created, later participants generally providing more information than earlier ones.

When the discussion between experimenter and participant had come to an end, the participant was invited to listen to another recording. A similar process was followed until six recordings had been heard and

discussed. The participant was thanked for their assistance and invited to ask questions about the purpose of the investigation. The experimenter also gave a contact address so the participant could, at a later date, enquire about the study's findings.

RESULTS

A summary of all responses can be found in Appendix VI. Discussion transcripts are not available as the sessions were not recorded. What *is* presented is a summary of the experimenter's notes, which were taken as the participants were speaking. The notes are ordered by speaker, with each speaker section being sub-divided into information categories (eg age judgements) and each information category sub-divided into individual listeners. All comments that were noted down at all are recorded here. The reader may find it informative and interesting to examine this appendix, as it is the basis of all that will be said in the next section.

Some aspects of the results have been summarised in Tables I, II and III. These tables have been complied from their more comprehensive counterparts in appendices VII, VII and VIII respectively. For the convenience of the reader, Tables I thro' III have between included in the body of the discussion as the text refers to them extensively.

DISCUSSION OF RESULTS

Listener Complicity

The responses made by participants were interesting from a number of angles. Firstly, the experimenter was quite surprised by the complicity of all listeners. Some of the set questions were personally felt to be rather unorthodox, and possibly unreasonable (in the sense of un-judgeable) to ask. However, as Table I shows, pre-set questions were almost always answered. This finding is particularly surprising as (it is felt) the participant was under little or no pressure to produce an answer. At the very least, they were under considerably less pressure than in the common questionnaire situation that stresses 'you should answer *all* the questions'. If a participant said they had 'no idea', this was taken on board by the experimenter and the matter was left at that. Furthermore, the experimenter's personal experience of the interviews suggests that the participants were not unduly bothered by the questions and had little difficulty in producing an answer.

Passage answered	%age of Pre-Set Questions
Set Passage	99%
Free Description	98%

Table I: Mean percentage occurrence of an answer to pre-set questions in the set passage and free description episodes.

Magnitude of Information

Related to this apparent readiness to provide information was the sheer amount that was produced. In addition to the pre-set questions, the participants (as a group) made between *ten and forty* other judgements about a particular speaker (see Table II). The large number is obviously partially due to the 'snowballing' procedure used by the experimenter, as a large number of questions were asked in total. It is also likely that some listeners contributed disproportionately to this total. Nevertheless, the participants seem willing to make a large number of judgements about the speakers from both the set-passage, and the free description, episodes. Because of the relatively uncontrolled nature of this initial study (the design is neither entirely within nor between participants), it would be rather difficult to determine whether one type of episode was producing more judgements than the other. However, this wasn't really the intention behind this summary - table II is only really intended to give a rough indication of total volume.

Passage	**Speaker**							
	1	2	3	4	5	6	7	8
Set Passage	14	17	11	31	25	25	22	24
Free Descr	22	40	**	18	20	15	28	18

Table II: Total number of judgements (excluding those arising from pre-set questions) made about a particular speaker in response to 'set passage' and 'free description' episodes. ** indicates that this data value was missing from the results.

Analysis of Judgements

Nature Of Judgement	Mean No. Of Listeners Judging (max. 8)
• Marital or Parental Status	5.4
• Details of Hair (Colour, Length etc)	5.0
• Details of Typical Dress	4.0
• Personality Judgements (including Sociable, Troubled, Happy, Nice, Caring and Diligent)	3.5
• Educational Status	3.0
• Sporting activities	1.1
• Eyesight (whether they wore glasses)	0.75
• Accommodation (Type of house, urban/rural)	0.75

Table III: Some of the most common judgements made about speakers (excluding those made in response to pre-set questions) with the mean number of listeners making such a judgement across speakers. The maximum number of listeners for one speaker is eight (four 'set passage' and four 'free description')

An analysis of the type of judgements produced (such as that shown in Table III) is interesting from two angles. Firstly, it shows the large *range* of judgements mad - what is being judged does not seem to be limited, in any obvious way, to a particular sub-section of biographical information. In addition to the list given in Table III, judgements about personal wealth and social class were also fairly common. They do not appear in the table because they were mainly implied by, or combined within, responses to the 'job' pre-set question (and hence rather difficult to analyse).

Secondly, the frequency with which some judgements are made is interesting. For instance, judgements about a speaker's marital or parental status appear a mean of 5.4 times (across the eight speakers) out of a possible eight (four presentations of a speaker in the set-passage condition and four presentations in the free description condition). The relative frequency of different judgements may have been affected by the frequency with which the experimenter asked relevant questions (as the questions were not standard across all discussions). However the occurrence of some information about a person's hair in almost 70% of all possible occasions of occurrence *is* rather striking and suggests a consensus view that such information is discernible. As Table III shows, judgements of what is discernible appear to be rather more far reaching than one might expect.

Other Judgements

If the judgements so far described seem somewhat bizarre, it is probably worth noting that these were simply the most *common* ones. Some listeners, in reaction to some speakers, produced extremely specific or unusual judgements. Here are just a few examples, paraphrased from the original interview notes. The dialogue has been reconstructed but every effort has been made to keep these 'para-quotes' in the spirit of what was said. All comments can be traced back to the notes in appendix VI:

"She has a slightly reddish complexion, is reasonably well educated, possibly artistic and slightly flirtatious" (of speaker 1, by listener 10)

"She is religious, but out of guilt rather than faith. Possibly a Catholic. Probably an emotional arguer. Has had a few problems in her life" (of speaker 1, by listener 3)

"He probably wears pin-stripe suits at work but corduroys and slippers at home. He was born in, or has ancestry in, the west country. He is scholarly. Probably a family man with two or three children - the youngest about 14, the oldest about 25. He sounds very much like Michael Heseltine. Used to smoke, has a boardroom manner and a thin-face. Is probably analytical and lateral thinking" (of speaker2, by listener 11)

"She is single, has a good education and works hard. Probably a public school education. Likes to be in charge" (of speaker 3, by listener 8)

"She is aware of environmental issues. She doesn't wear glasses. Has middle of the back length mousy blond hair. Is possibly religious and is a social smoker and drinker. She has had a fair bit of training" (of speaker 3, by listener 3)

"She probably isn't the life and soul of the party, and doesn't go out much. Is probably an excellent employee" (of speaker 5, by listener 10)

" He is probably visiting on business for an Italian company. He is single, good-mannered. He possibly has an English mother" (of speaker 6, by listener 11)

"She has permed grey hair but it is not blue rinsed. She was once blond. Wears old but nice floral dresses. Her husband is dead. She wears her glasses on a string" (of speaker 8, by listener 4).

Summary Of Findings

It would therefore seem that, in response to an audio recording of a speaker, listeners are willing to answer questions about the speaker's age, sex, height, attractiveness, occupation and so forth. In addition, they seem able to produce, or respond to questions about, a large number (10-40) of other judgements. The most common of these concern the speaker's accommodation, hair, marital or parental status, typical dress, personality, education, wealth, social class, participation in sport and use of spectacles. It is not entirely clear whether the set passage or the free description episodes lead to the most judgements, but Table II does show that both types provide at least some information. The extent to which different listener's agree, and the extent to which they are accurate, is also unclear. However, a skim through the raw results in appendix VI suggests some level of agreement, although there are some gross disparities in some cases (eg a speaker being rated as 20 by one listener, and over 50 by another). An analysis of these two questions would be difficult because the experiment was rather uncontrolled and has neither entirely a within participants, nor entirely a between participants, design. Most importantly, the question of accuracy was considered essential to investigate further. The results given so far are interesting but, to a certain extent, participants could be 'playing along' when they produce these large number of judgements. What needs to be addressed is whether these judgements are accurate, or, at the very least, consensual.

FOLLOW-UP STUDY

Introduction

The initial study showed that people are willing to make a large number and range of judgements about a speaker from recordings of their voice. The aim of the main study was to investigate this in more detail. Specifically, the researcher was interested in whether there would be any level of agreement between participants' judgements and whether these judgements were actually *accurate*. The relative informativeness of 'set passage' and 'free description' passages would also have been an interesting question but, due to time constraints, the researcher decided to concentrate on the first two questions. It was decided to limit the follow-up study to the 'set passage' episodes because the 'free description' ones varied widely in length (from less than ten seconds to over four minutes!) and in general content. This was considered to be uncontrolled and unacceptable variance. There would be various ways of getting round this problem but, at the same time, there is a limit to what can be investigated in one study.

The experimenter decided that a structured, forced-choice questionnaire style study was most appropriate for this follow-up study. Although such a method obviously restricts the participants responses (and therefore may miss some interesting information), this sacrifice was considered worthwhile in order to obtain directly comparable

judgements. As all participants would always be asked the same questions in exactly the same format, this left two problems. Firstly, the total number of judgements must be restricted to avoid being over-demanding of the participants. Secondly, the questions must be phrased in a clear, concise way that allowed for some reliable index of accuracy.

Choice of Questions

The full list of judgements that the experimenter considered to occur with reasonable frequency is presented in appendix X. This is an extended list in comparison with the list given in the initial study's results, as it includes some slightly less frequent judgements. For all but two of these judgements (attractiveness and intelligence), relevant information about the speakers was available from the extended follow-up biographical questionnaire (see initial study results). This gave a possible twenty-nine judgement areas - an excessive number for experimental purposes in the experimenter's opinion! Fortuitously, this problem was largely self-solving. For one reason or another, many of the variables appeared unsuitable for a structured judgement task. The variables excluded, and the reasons for exclusion, are given below:

- ***Facial hair, emotional state and typical dress*** were excluded because there was very little range in the voice sample (only 1 had an facial hair, nearly all reported being relaxed and all reported their typical dress as broadly casual or smart casual).

- ***Weight and parental status*** were excluded because there was non variation with respect to another variable (height and marital status respectively). That is to say, no speakers were over or under-weight, and no unmarried speakers had children.

- ***Vocational and Education questions*** were excluded because the sample was highly unrepresentative of the general population - all but one had at least 'A' levels, more than one had a pHd, all had jobs connected in some way or another to the University of Southampton. However, a question about gross earnings was considered interesting, and this variable had a good range (0-£35,000 per annum)

- ***House type*** was excluded on grounds of the difficulty of phrasing a relevant question to the listeners. Limitations of the information obtained about the speakers' accommodation was also a factor.

- ***Intelligence and attractiveness*** were omitted because it was considered too troublesome, at least for the present study, to attempt to create a reliable rating of such for the speakers.

- ***The speakers' shoe size and whether they wore glasses*** were excluded out of a total lack of interest in the questions!

The remaining judgements were age, sex, height, hair colour and length, eye colour, gross earnings, car ownership, marital status, area of

origin, participation in sports, frequency of social outings and typical group for social outings. All of these have a good or reasonable range across the speakers, as in shown is appendix X. This just left the problem of how the remaining variables could be assessed in a structured situation.

Question Phrasing

In addition to the problems of formulating clear and concise questions, was the conversion of the judgement areas into measures that allowed the results to give an idea of the accuracy of the listeners' judgements. Age, sex, gross earnings and car ownership were considered reasonable to ask unmodified. Height, hair length, participation in sports, marital status and frequency of social outings - previously free response questions - were modified to a multiple choice format (or yes/no in the case of sports and single/married in the case of marital status).

Hair colour, eye colour, area of origin and form of social outing were more tricky. These were also converted in to a multiple choice format, but their ordinal properties were less than obvious. The experimenter decided to impose an ordinal sequence on these variables so that some idea of accuracy could be obtained. Hair colour and eye colour were arranged by shade (light to dark). Form of social outing was organised on a friends-family scale. Finally, and perhaps most troublesome, area of origin was organised on a scale ranging from south, through the Midlands and the North, to abroad. This scale was also arranged so that the bigger the difference between the rating and the actual value,

the bigger the geographical distance between the two areas. On reflection, the organisation of 'area of origin' was considered to be unsatisfactory and therefore it was decided that this variable would be analysed categorically (see results).

Finally, with the odd nature of some of the questions, the experimenter thought it might be interesting to see how confident the listener's were of their ratings. Confidence scales were added to age, sex, gross earnings and marital status. They were not added to scales such as eye colour because it was anticipated that most participants would answer in the maximally unconfident, through pre-existent expectations if nothing else.

The combination of variables and confidence ratings gave a total of 17 ratings which needed to be made for each of eight speakers. The experimenter considered this an adequately small number (just!). A trial run of the questionnaire, with one participant, indicated that the whole rating process would take about twenty minutes, which was considered reasonable by the experimenter. This test was covered by the ethical permission detailed in the 'procedure' section below. All that remained was to present the questionnaire to a group of more than one.

METHOD

Design

Recordings of eight different speakers reading a set passage were played to three groups (n=12, n=7, n=12, Total =31) of participants. The recordings were the same as those used in the initial study (see above), with the order of presentation being fixed but randomly determined. All members of a particular group heard the recordings simultaneously, and the same recordings were played to each group. After each recording, participants were given approximately two minutes to answer set questions about the speaker on a printed sheet. After all recordings had been played and responded to, participants were asked to comment on how they approached the task and to provide some information about themselves (age, sex etc).

Participants

The three participant groups were as follows:

Group A (n=12) Volunteers from 'A' level students and their parents visiting the Department of Psychology at the University of Southampton.

Group B (n=7) As above, later the same day.

Group C (n=12) Volunteers from one 'A' level Psychology class at Southampton Technical College.

Overall, there were 7 males and 24 females. Their mean age was 20.3 years (s.d. 5.816), although ages ranged between 17 and 42. Almost

84% came from the south or south-west of England, 10% came from abroad and the remaining 6% from Wales, the North and Northern Ireland.

Materials & Apparatus

The voice recordings were played from a high quality duplicate of the original tapes, with the order re-arranged for ease of presentation. The order in which the eight recordings appeared on the tape was fixed, but initially randomly determined by means of a hand-drawn lot.

Recordings for groups A and B were presented via the machine they had been recorded on. For group C, as the original tape recorder was difficult to set up and transport, recordings were presented on a good quality, stereo tape recorder (Sony CFS-W304L)[8].

Instructions to the participants and all required response sheets were presented in the form of a stapled booklet (see appendix IX for an example).

Procedure

As the procedure for this study was substantially different to that used in the initial study, a further request for ethical permission was submitted to, and approved by, the department's ethical board.

---

[8] As Brian, a member of the technical staff noted, 'replication is for people with Volvo estates'!

For groups A and B, volunteers were asked over a lunch break in the Open Day whether they would be prepared to participate in an experiment. They were informed that it would involve listening to some voice recordings and answering some questions on them, and might take up to twenty minutes. It was stressed that their participation was entirely voluntary and in no way formed part of the assessment procedure. They were also informed their responses would remain anonymous and that they could withdraw their consent retrospectively if they so wished.

For group C, the class teacher had previously informed the class that the first half of the lesson would be taken up by an outside experiment. When the experimenter arrived it was stressed that, even though this was a lesson activity, participation was voluntary and they were not required to take a questionnaire if they did not want to do so. Further, it was made clear that their responses would remain anonymous and that they could withdraw their consent retrospectively if they so wished.

The procedure for all three groups was otherwise identical. Participants were handed the questionnaire (see appendix IX) and asked to read the front page. They were then asked if they understood, and any misunderstandings were clarified. Participants were also asked if they had any hearing difficulties. Following this, they were asked to turn the page and the first recording was played. They were then asked to respond to all the questions on that page. When all participants

appeared to have finished, the experimenter asked "Everyone finished?" and, if so, the next recording was played. If people were still responding after two minutes they were urged to finish off. This was done to ensure the experiment did not exceed the stated duration (ie twenty minutes).

The experiment proceeded in a similar way until all eight recordings had been played and responded to. Participants were then asked to fill in the last two sheets, which asked how they approached the task, and requested some biographical information. Finally, participants were thanked for their help.

For groups A and B, the experimenter then encouraged participants to ask questions about the nature and purpose of the research. This was also done for group C but, in addition, at the experimenter's invitation and the teacher's request, the experimenter held a short lecture and discussion session on the psychological, ethical and statistical issues raised by the current research. It was also promised that a summary of results would be sent at a later date. None of these things were done for groups A and B because of the time constraints the participants were under, and because, as they came from all over the country, distribution of results would be time-consuming and costly.

RESULTS

All responses to all questions, for all eight speakers, are presented in appendix XI[9]. Due to the vast amount of information these results represent, it is not possible here to provide any entirely complete summary. Instead, the author will simply present some of the most *interesting* findings. Nevertheless, there is still a very large amount of information presented over the next few pages, so the reader may find it useful to read the results summary at the start of the discussion first, to avoid getting lost in details

Participants' Reactions

All but one of the thirty one participants provided at least some information for each of the seventeen variables. Nearly all responded to the vast majority of questions. The thirty-first participant refused to answer any of the questions, excepting the age and sex ones, on the grounds that these were not possible to rate. This participant is treated as a 'missing value' in the assessment of all other judgements. This participants general (derogatory!) comments may be found interesting and are therefore presented at the end of appendix XIII.

Participants produced some interesting information in the evaluation section of the questionnaire (a full summary is provided in appendix

[9] These were complied via Excel (v. 3) on the Macintosh, from which they could be exported to a number of statistical packages. The data were also converted in a variety of different formats to facilitate analysis. The form they appear in appendix XI is the one in which they were originally typed. Computer-based copies of the data are available from the author.

XIII). Over two-thirds reported having clear images of the speaker some of the time or all of the time. Almost two-thirds reported approaching the task by using the 'image in their head' and/or utilising comparisons with people that they knew. Most of the remaining third reported that they used what may be roughly described as 'physical' characteristics of the voice eg tone and resonance.

Over a third of participants reported judgements of sex as being easy, almost a third judgements of area as difficult, and over a third judgements of earnings as difficult. Almost a third stated that one or more of the judgements were not possible to make from the voice recording.

Participants Responses

A lot of variables were assessed, so it is probably easiest to deal with them one by one. The author will deal with binomial variables first, followed by area of origin, and ending up with analysis of the ordinal variables. The large number of variables means it would make little sense to include all summary tables in a notional 'results' section and then discuss them several pages hence. Therefore the results are individually discussed in this section. The discussion section will draw together all these separate comments.

Considerable reference is made to appendices for the full details of statistical tests. This strategy is employed to reduce clutter but the

interested reader may wish to read through these appendices. The appropriate appendices can be found via the contents pages.

Binomial Variables

Sex

The contingency table presented in table IV says all there really needs to be said about sex judgements - they are, for this sample, basically error free :

	Speaker Is Male	*Speaker Is Female*
Rated As Male	123	**0**
Rated As Female	1	**124**

Table IV: Contingency table for judgments of speaker sex across eight speakers and thirty-one participants.

This accuracy is reflected by the participants' reported confidence in making the judgements. Mean reported confidence across speakers was 4.89 (range 4.87 - 4.94). 5 was the maximum confidence rating. The full set of means and standard deviations for age confidence are presented in appendix XIII.

Participation in Sports

	Plays Sport	*Doesn't*	*TOTAL*
Rated as Sports Players	**110** (73%)	**32** (36%)	142
Rated as Non-sports	**40** (27%)	**57** (64%)	97

Table V: Contingency table for judgements of speaker sports participation across eight speakers and thirty-one participants.

The full statistical tests performed in this section can be found in appendix XIV.

Table V suggests there is some judging ability. There was a significant difference between the number of times individual listeners rated sports players as playing sports, and the number of times they rated non-sports players as such ($Z=4.607$ $p<.0001$). This held true when only married persons ($Z=4.354$ $p<.0001$) or females ($Z=3.258$ $p=.0011$) were considered. Note that the apparent response bias to participation is justified as more speakers actually do play sport than not. Hence it appears that listeners are performing at better than chance levels in the judgement of sports participation.

One hypothesis would be that age is an important factor in this judgement, and some support was found for this. When the initial test was repeated using only the four oldest speakers, there was still a significant difference ($Z=4.354$ $p=.0001$), but when the four youngest sports players were considered, the difference was non-significant ($Z=1.40$ $p=.16$). One speculative conclusion might be that general

physical condition or fitness is what is being tapped here, and perhaps younger non-sports players are not being identified because they approximate the physical condition that allows sports participation. The author appreciates there would be may other explanations for these results, and suggests this hypothesis would be worthy of further investigation. This will be returned to in the general discussion.

Marital Status

	Married	*Unmarried*	*TOTALS*
Rated as married	78 (65%)	**21** (17.5%)	99
Rated as unmarried	42 (35%)	**99** (82.5%)	141

Table VI: Contingency table for judgements of speaker marital status across eight speakers and thirty-one participants

Table VI suggests there is some ability to judge this variable. There is a significant difference between the number of married persons rated as married, and the number of unmarried persons so rated ($Z=4.692$, $p<.0001$,). Note, however, that listeners were more likely to rate any speaker as unmarried than married, even though this bias is not reflected by the sample (four are married, four unmarried). Nevertheless, listeners appear to perform at better than chance levels in determining marital status.

Mean confidence across speakers for these judgements was 3.48 (range 3-4.04) - which can be interpreted as approaching reasonable

confidence. Participants seemed more confident of their ratings of the 3rd ,5th and 7th speakers (mean confidence 4.04,3.75 and 3.67 respectively), all of whom were almost unanimously rated as unmarried. Speakers 3, 5 and 7 were, respectively, a 20yr old male, a 26yr old female and 22yr old female. This suggests age may be a factor in the judgement of marital status. However, this hypothesis is difficult to test as the four youngest speakers were all unmarried, and the oldest four all married.

Full details of statistical tests used in this section appear in appendix XIV .

Car Ownership

	Car Owner	*Non-Car Owner*	*TOTAL*
Rated as Car Owner	96 (81%)	66 (56%)	162
Rated as Non-Car Owner	23 (19%)	52 (44%)	75

Table VII: Contingency table for judgements of speaker car ownership across eight speakers and thirty-one participants.

Table VII gives equivocal evidence for the judgement of this variable. There was, however, a significant difference between the number of times individual listeners rated a car owner as owning a car, compared to the number of times a non-car owner was rated as such (Z=3.365, p=.0008). This held true when only the four oldest (Z=3.72 p=.0002), or the four youngest (Z=2.32 p=.021), speakers were considered.

However, all speakers appear considerably more likely to be rated as a car owners than not (see above table).

Listeners seem to have some ability to judge car ownership, but the large bias to ownership overall makes it difficult to claim much accuracy for this judgement.

Full details of statistical tests used in this section appear in appendix XIV .

Area of Origin

As 'area of origin' is a categorical variable, the most illuminating illustration of the responses was considered to be a contingency table, such as the one presented in fig. I.

As can be seen, speakers from the South, South-west or the Midlands are recognised as coming from that general area. The speaker from Wales is rated as not Southern, Scottish or foreign.. The speaker from abroad is almost unanimously identified as foreign. Taking into account that the vast majority of raters were from the south of England, these judgements are perhaps best described as a three-category decision, the categories being: a) from this (broad) region, b) not from this region but from this country, c) not from this country.

See over for Fig. I

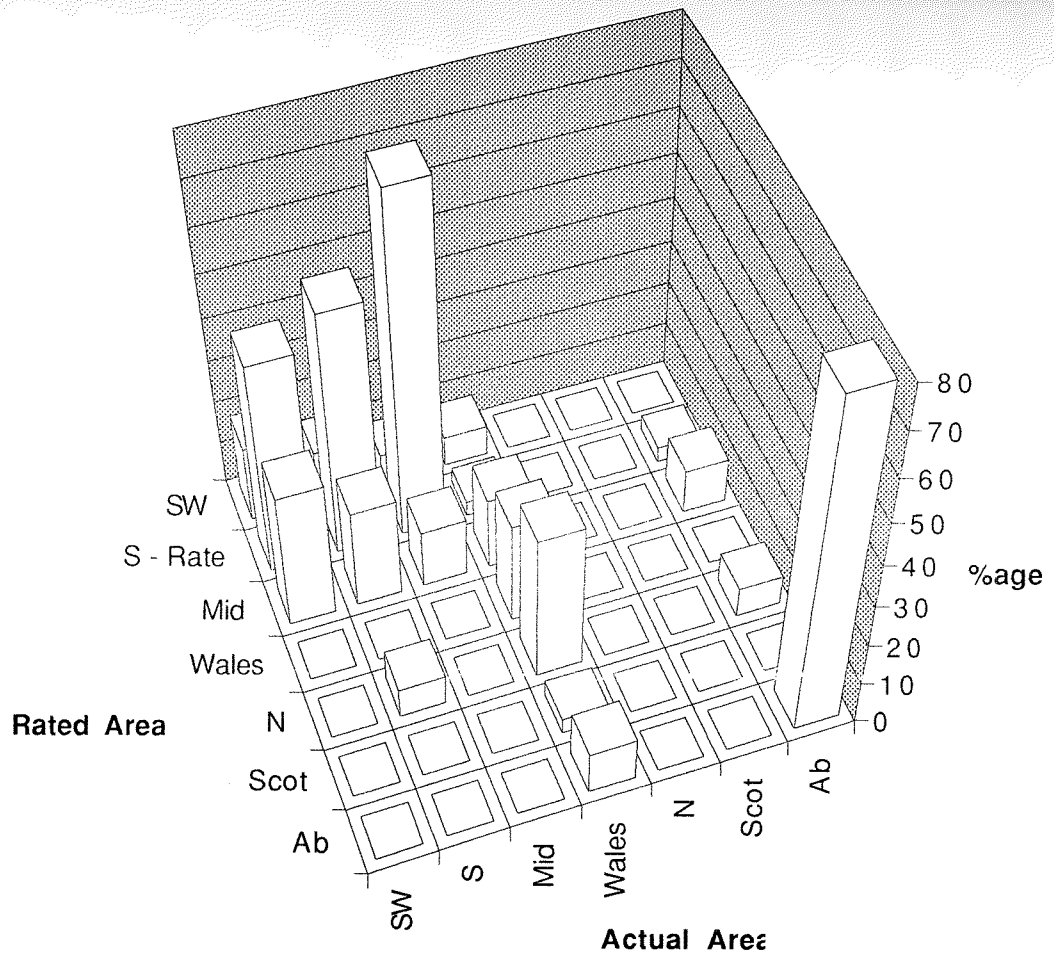

Fig I: Area of Origin versus Rated Area of Origin Contingency Table (%age frequencies used)

Preliminary Analysis Of Ordinal Variables

All (non-confidence rating) variables that had an ordinal structure were considered. Speakers were ranked for each variable and placed in rank order. If, for a particular variable, more than one speaker produced the same value (eg both speakers earned £12,000), then the individual participant's ratings were averaged across those speakers. Hence, the data analysed contained only one data point per participant per value of the variable. The logic behind this reduction was that the rating of different speakers identical on a particular variable represented multiple ratings of the same value of the same variable. Page's L tests were then performed on each of the variables. The full results of all these tests, and a fuller description of the method, can be found in appendix XII.

With the exception of 'frequency of social outings', all tests were highly significant ($p<.0001$). In other words, rated judgements of the variables showed a significant trend with the increasing actual values of those variables. In all cases, an examination of the ranked means indicates that these trends were in the positive direction. Hence, as the value of the variables increases, the rating of that variable also tends to increase. This finding holds true for *age, height, earnings, eye colour, hair colour, hair length and typical others for social outing.* It was decided that all these variables merited further analysis. For clarity, these analyse will be considered in two broad sections:

Further Analysis Of Age, Height and Earnings

Age

The Pages' L test showed there was a trend, but are people's ratings of age *accurate?* Fig. II suggests that they are. The graph presents rated age as a function of actual age. An interesting result is that, as actual age increases, the *range* of age judgements increases quite substantially.

Although fig II provides some interesting information, it is a little hard to visualise the participants' accuracy. Fig. III presents the data in an alternative form, and also provides a best straight-line fit.

The equation of the straight line is as follows:

Rated age (yrs) =.7 x Actual Age (yrs) + 5.83 yrs $r^2 = .65$

To put this into words, the rated age is well predicted by the actual age, although, in the range studied, all ratings tend to be under-estimates, and these under-estimates become more gross as the actual age rises. The equation predicts that a 20 year old would be rated as a 20 year old, but a 60 year old would probably be rated about 48. Taking figures II and III together, it would seem that age can be judged with some accuracy, but all judgements tend to be slightly low, and the accuracy declines, and the variance increases, with older voices. This hypothesis is supported by listener confidence ratings - participants become less confident as the speakers become older (ψ_L= -11.05, p=.026, see appendix XVI).

See over for fig. II&III

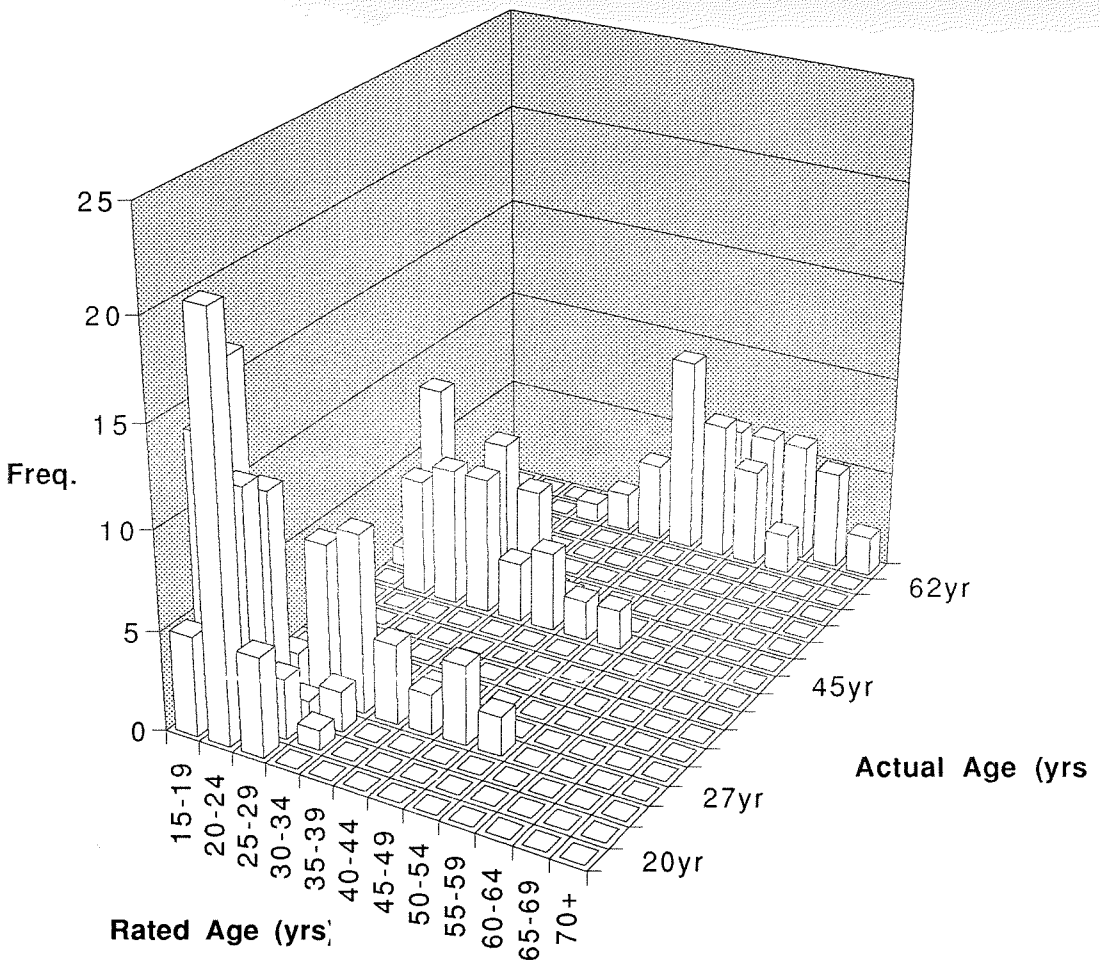

Fig. I: Frequency of Judgements of Rated Age with Increasing Actual Age

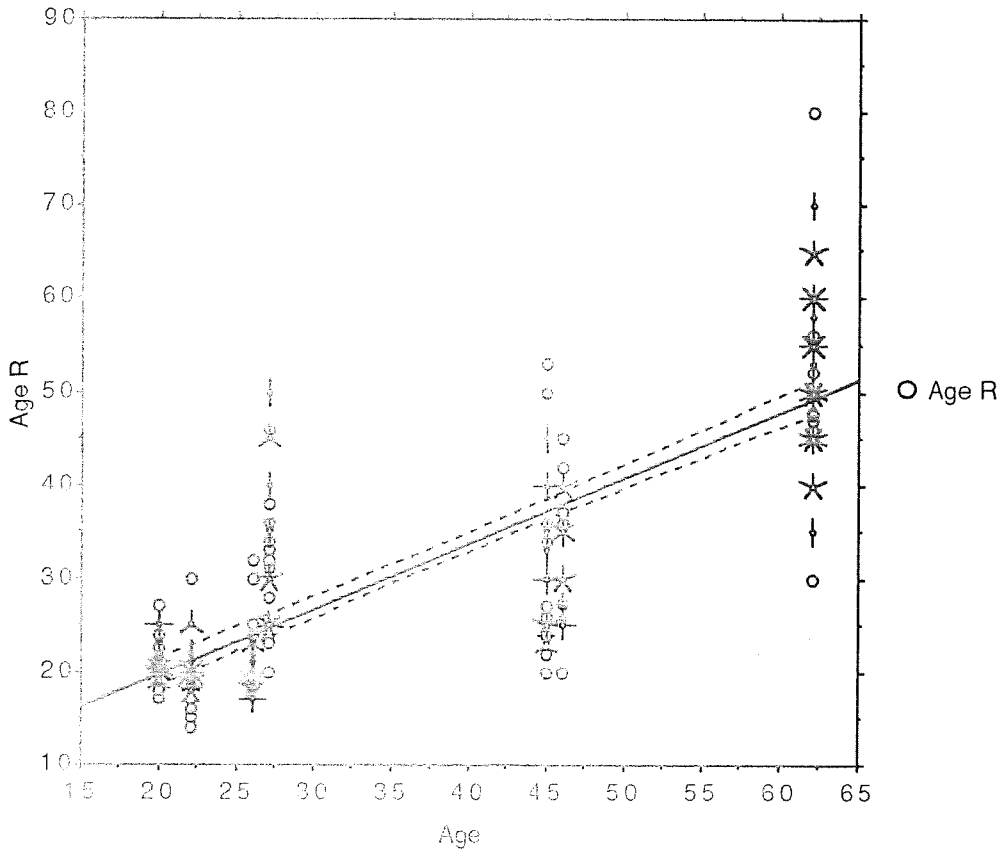

Fig. III: Scatterplot of Age Ratings (yrs) against actual age (years) with best fit & 95% confidence interval lines

Height

Much the same questions were asked about height as were asked about age. Fig.IV suggests that participants demonstrated some accuracy in rating height, although the tallest people were generally under-estimated.

Once, again, the question of accuracy is best answered by a scatterplot. The points on Fig. V are derived - height ratings are the mean height ratings for a particular speaker, actual heights are the height in inches of the speakers. The equation of the straight line (y=.19x-10.52) is rather hard to interpret because of the differing units for x and y. However, inserting a few numbers shows that there appears to be a certain amount of centering (heights of short people are slightly over-estimated and heights of tall people slightly underestimated). Significantly, the straight line is a reasonably good approximation of the relationship (r^2=.556), suggesting reasonable accuracy overall.

See over for Fig. IV&V

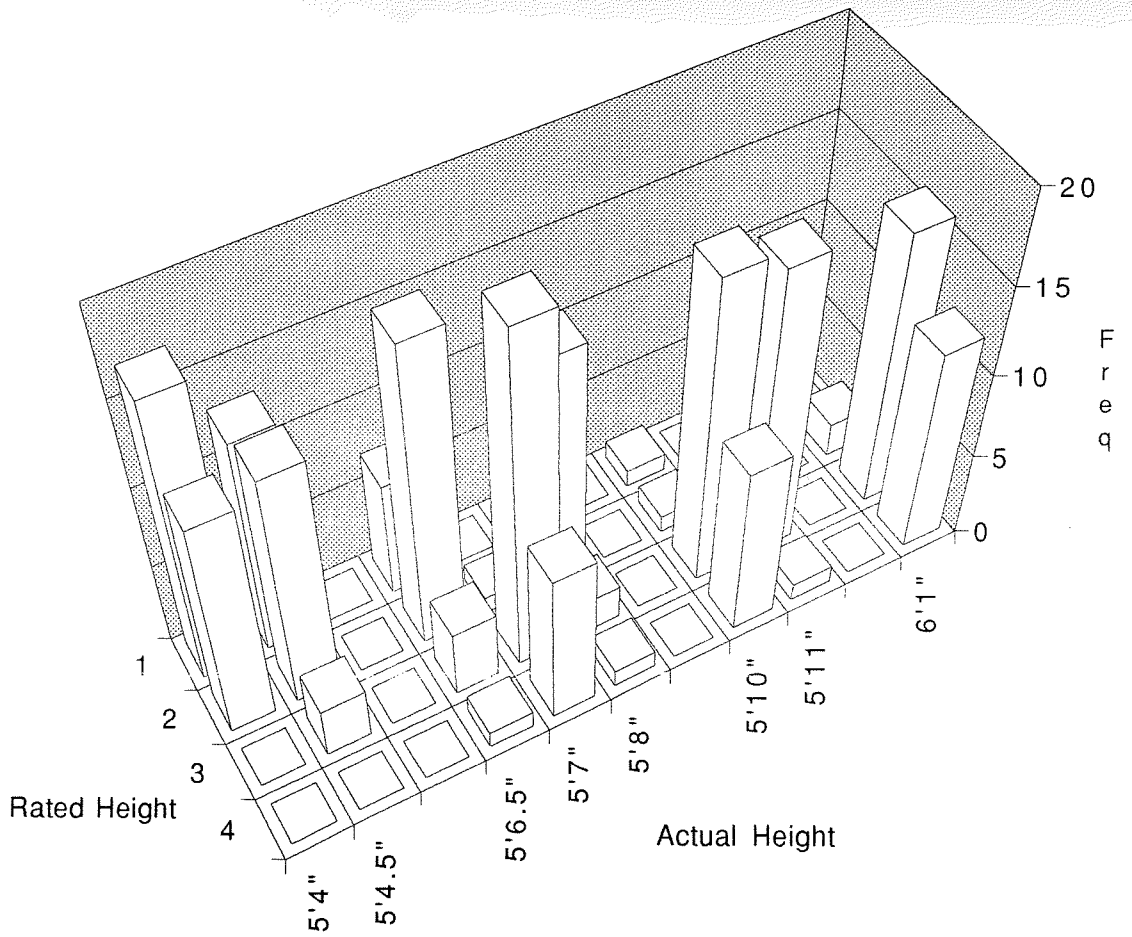

Fig IV: Frequency of Judgements of Rated Height (arbitary units) with Actual Height (inches)

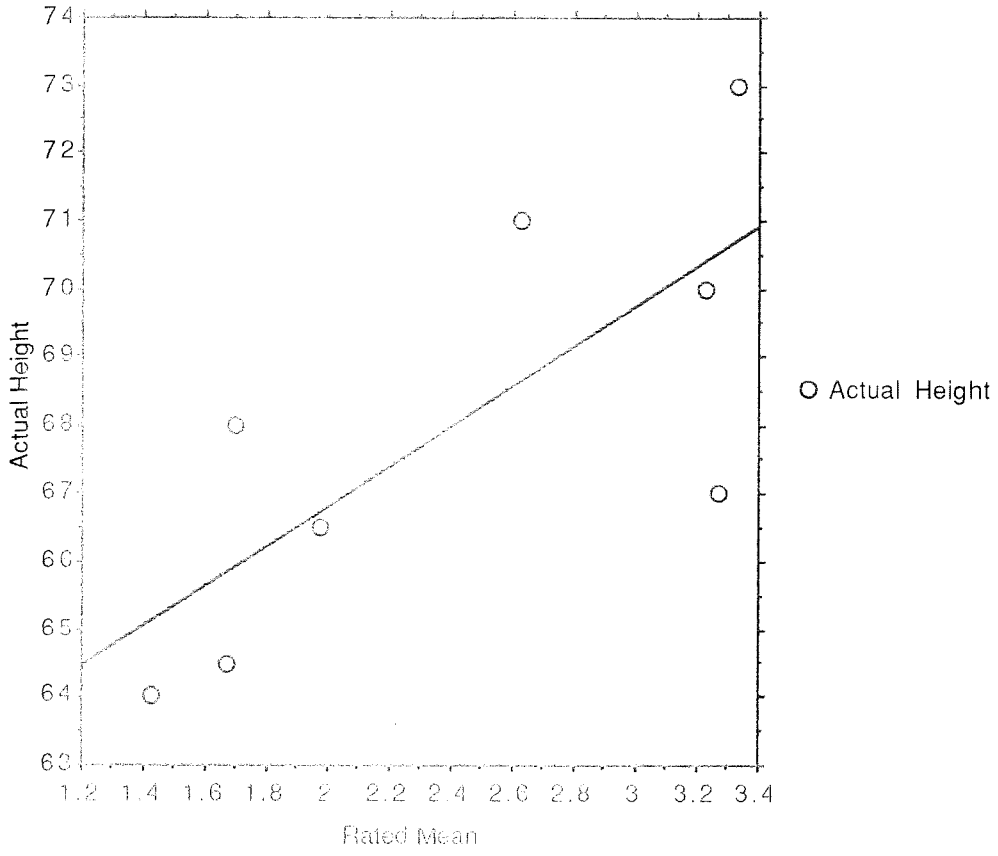

Fig. V: Scatterplot of mean height ratings (arbitary units)of eight speakers against their actual height, with best-fit line.

What underlies this accuracy, however? It is a commonly held assumption that women tend to be shorter than men, and the voice sample tends to reflect this . Mean male height was 70.25 inches (s.d. 2.5), mean female height 65.72 inches (s.d. 1.708). This difference was significant at the 5% level (t=2.97). Hence, an important question was whether height judgements could be made independent of gender judgements.

A further two Pages' L tests were performed - one on the female speakers only, and one on the male speakers only. Although the trend was still significant for female speakers (Pages L =48.5 Z=8.5 $p<.0001$), it was non-significant for male speakers (Pages L = -0.834, Z=0.79 p=0.20). It was noted that speaker 2, from Spain, seemed to be consistently under-estimated, but when he was omitted, there no significant difference between male speakers was found (X2 corrected for ties = 0.426, p=n.s.). The author would hypothesize that the voice can provide gender-independent information about height, but for males, gender is the information that is actually used. This hypothesis would merit further investigation.

Earnings

Attempts to do a similar analysis on earning were unfruitful. No line or curve fitting tried adequately explained the variance (the best was a straight line, $r^2=.17$). From the Page's L test we can assume listeners

have some ability to judge relative earnings. However, individual figures widely vary and are often wildly inaccurate (see the raw results in appendix XI). This lack of accuracy is reflected in the participants' lack of confidence in their judgements (mean 2.29, range 1.96-2.43).

One might hypothesize that age and gender of the speakers are important factors in the judgement of earnings. However, there is still a significant trend when earnings of just males ($Z=4.92$, $p<.0001$), the four youngest speakers ($Z=4.91$ $p<.0001$) or the four oldest speakers ($Z=3.689$ $p=.0002$) are considered. However, the trend is non-significant when only female speakers are considered ($Z=0.512$ $p=0.30$). Of the speakers in the voice sample, males and females do not earn significantly different amounts (Male mean = £19.5k (s.d. 14.6) Female mean = £5.5k (s.d. 4.5) $T=1.768$ p=n.s.).

These results are a little hard to analyse, but the author would suggest that there is some age-independent information of earnings. This information can be picked up independent of gender but, for females, gender appears to be the source of information actually used.

Some Interesting Results from More Unusual Variables

Hair colour

The ordinal structure is shade (light to dark). Darker haired speakers tend to be rated as such. The trend is still significant if speaker 2 (from Spain and generally rated as foreign) is removed ($Z=4.05$ $p<.0001$).

However, if speakers with grey or greying hair are also removed, the trend becomes non-significant (X2 =4.097 p=.129). Hence, judgements of hair colour may be limited to grey/non-grey.

Hair length

An ordinal trend was found, but this became non-significant when males were removed from the sample (Pages L=-0.167 p=n.s.). However, female speakers in the voice sample do have significantly longer hair (T=2.449 p<.05)

Statistical Chaff

In addition, typical others for social outing and eye colour were examined. However, in the former case, the became non-significant when only married speakers were considered (X2=2.896 p=.2351) , even though the typical others of married and non-married persons did not differ significantly in the voice sample (T=1.85 p=n.s.). In the latter case, the trend became non-significant when the speaker from Spain was excluded (X2-3.182, p=0.204)

DISCUSSION

Summary Of Findings

Of all the variables studied, age, sex, height, area of origin, sports participation, marital status, gross earnings, hair colour and hair length

appear possible to judge to some degree. Judgements of age and height are reasonably accurate, and judgements of gender are basically faultless. Sports participation and marital status are judgeable at above chance levels.

The accuracy of other judgements is rather limited, particularly hair colour, hair length, gross earnings and area of origin. Hair colour appears to simply be a grey/non-grey distinction, which suggests age may be implicated, and judgements of hair length appear to be dependent on judgements of gender. Area of origin is probably a three category judgement along the lines of 'from listener's region/not from listener's region/not from listener's country'. Judgements of earnings are accurate only in purely relative terms and, although some information appears present for males, judgements of females' earnings appear to be determined by their gender.

Other factors also seem to be implicated in the judgements of sports participation, marital status and height. There does appear to be accurate, gender-independent information for height (as witnessed by the significant trend in the males-only test), but for males, information about gender dominates. Judgements of sports participation do not discriminate between the younger sports and non-sports players (but do between the older ones). The author suggests that listeners are treating this question as a judgement of general physical health, which would be consistent with these results. Lastly, age looks as if it is a factor in the

judgement of marital status, but this hypothesis was not testable here (all younger speakers were unmarried, all older speakers married).

Listeners also seem to have some ability to rate the accuracy of their judgements, as manifested in confidence ratings. In all cases, as accuracy of rating declines (or falls to zero), participants' confidence in their answers decline and comments about the difficulty of rating that variable become more common.

A Question Of Accuracy

An important issue to consider with these results is what we actually mean when we describe a judgement as accurate. From an ecological perspective, any information the organism picks-up that allows correct, reliable distinctions of a particular variable provides accurate information about that variable *and allows it to be accurately specified.* If judgements of age are accurate, and the majority of people over a certain age have grey hair, then age provides some accurate information about hair colour. In a society where most people that had gone grey dyed their hair back to its original colour (or people simply did not go grey), age would not provide accurate information about hair colour. However, and this is the essential part, the environment is highly constrained and all logical possibilities (eg old people do not tend to have grey hair) are not environmental probabilities. Hence, variables do not need to be uniquely specified to be accurately judgeable. Taking

this view of accuracy, the author considers that the voice provides information about all the above variables (ie sex, sports participation, marital status, area of origin, age, height, earnings, hair colour and hair length) with varying, but non-negligible, levels of accuracy.

Independently Specified Variables

Even given the above argument, the question of which information sources in the voice are independently specified is still an interesting one. Obviously, such a question can never be entirely answered empirically as there is always the possibility of excluding a variable from your study that would co-vary with the variables you *are* looking at. Nevertheless, we can consider which variables, of those studied, are specified independently *of each other*. From this study, the author tentatively suggests five such variables, which are as follows:

Social, bodily and chronological age - Age judgements are possible, both of chronological age, and of cultural (marital status) and physical (sports participation as general health) manifestations.

Physical and social gender - Gender judgements are possible, both as physical gender, gender typical characteristics (ie tending to be shorter), and as cultural manifestations (hair length and earnings). Hair length and earnings are described as cultural manifestations of gender because, in the society of the listeners, women *tend* to have longer hair, and to earn less.

Physical height - Judgements of this variable were not found to be entirely independent of judgements of gender. However, gender does provide some information about height, and there does appear to be some gender independent information about height.

Tenancy of Speaker in Listener's Geographical area - Only a crude 'this area/this country/abroad distinction' was found. However, one must consider that all participants spoke English and all read a standard passage, which may have obscured some of the more 'macro' cues to geographical area (colloquialisms, pauses and general accent).

Social Position - This variable is particularly tentatively put, but judgements of earnings might be best seen as general judgements of social position. In the listeners' society women do *tend* to hold lower social positions - this may explain gender being an important factor in the judgement of earnings.

The author makes no claim for the necessity of these five variables, and does not believe the present study provides enough information to state them definitively. They are stated only as hypotheses and it would be useful for future studies to investigate them. However, before going on to suggest ideas for further research, the author would like to point out some problems in the existing one.

Some Limitations

Like all studies, this one has some problems and some limitations. The most obvious of these is the very small number of voices that were judged - only eight. Further, all speakers, at the time of the study, lived within a few miles of Southampton University and were all connected, in one way or another, with it. Although the researcher attempted not to select speakers on grounds of being particularly informative, they still represent a highly non-random sub-set of the English speaking population (which, in turn, is a non-random sub-set of the speaking population).

Further, the small number of speakers combined with a large number of variables inevitably meant that some variables had a rather limited range, and that many were highly co-variant. This obviously casts some doubt on the conclusions made because some hypotheses could not be fully tested. In some cases the problem is solved by ecological significance; that is to say, the co-variance of some variables in the speaker population would be expected in the general population. However, in other cases, the absence of non-typical, but relatively common, combinations puts conclusions on unfirm ground. If an elderly woman with very short red hair, who is an avid squash player, doesn't own a car, comes from Spain, is 6'4" and has a Scottish accent seldom occurs in the general population then it does not matter that, say, age provides information about hair colour. However, a middle-aged, unmarried man may be a relatively common occurrence. If this is the

case, finding marital status judgements to be determined by age means that judgements of marital status may be ecologically inaccurate.

In addition, the results are limited by the listeners, who are generally young women from the south of England. Due to the lack of range in the listener sample, it is possible that the nature of some of the judgements are group-specific. For instance, the increased variance of age judgements with increasing speaker age may be due an increasing difference between speaker and listener age, rather than an increase in age *per se*.

Theoretical Questions and Research Issues

As was alluded to in the introduction, it would be possible to interpret the results of this study from a representational perspective. In fact, the participants' reports of how they approached the task could be seen to support the comparison of representations to some long-term store. Before making too much of this evidence however, one should note that the majority of participants had done, or were currently studying for, A-level psychology - an examination dominated by representational descriptions.

Providing evidence for one theoretical perspective over another was not the central aim of this study. Nevertheless, however you interpret the conclusions given here, this study has shown the importance and utility of using units of analysis other than the pure tone or the disconnected

phoneme. It has also shown that the construction of a theory from the study of supposedly elemental components is not the only way in which one can fruitfully proceed - it may also be useful and informative to look at the information organisms actually encounter in their everyday world.

This study also opens up some interesting theoretical questions. First of all, despite going to some lengths to consider accuracy in an ecological context, the author appreciates that the question of what is directly specified, and what is indirectly inferred, is an important one. If the process really is 'this person is about forty, people in my society that are forty are usually married, therefore this person is married' , then this is theoretically significant. One might investigate the question of direct vs. indirect for certain characteristics by examining the informational structure of the voice. This might be done by the systematically manipulating the recordings (by filtering, degrading, compressing, tone shifts etc) and studying the effects of such manipulations on judgement. To take an example from the introduction, if performance on gender judgements dropped when the fundamental frequency of the speech was removed then we might hypothesize that F_0 provides some information about gender. Alternatively, the study of non-typical cases (as mentioned above) might also be interesting. If characteristics could be judged even if they were non-typical in respect to some other variable (eg a middle age person that is unmarried), this would cast doubt on the assumption that such judgements are inferred from a small number of directly specified variables.

Related to the question of direct specification versus indirect inference is the distinction between what may be roughly called 'intrinsic' and 'optional' characteristics of the voice. Some characteristics, such as age and height, may be inevitably transmitted due physical differences in the biological speech apparatus. Others may be optional in the sense that they are under the control of the speaker - he or she may be able, for instance, to 'put on' a convincing Northern accent.

One example shows the difference very clearly. Some time ago, it was common to use women to act the parts of small boys in radio plays. Listening to these recordings today, this strategy appears rather transparent - one can tell that the speaker is a woman, not a young boy. However, at the same time, we can tell that the speaker is *meant* to be a small boy. What appears to be happening is that the director and the actress are treating age and gender as optional characteristics, when they are actually intrinsic. In contrast to this, I am regularly surprised by the normal speaking voice of famous character actors[10]. Here, accent is being successfully treated as an optional characteristic. Physical, acoustical and anatomical investigations might throw more light on which characteristics are intrinsic, and which optional. It would be interesting to investigate which changes in general body dimensions (and their counterpart changes in the speech apparatus) led to detectable

[10] I am particularly referring here to the actor who plays Inspector Burnside in the 'The Bill". In contrast to his screen role, he speaks in a quiet and cultured manner.

changes in the voice, and which of these changes actually gave a lawful change in the perception of certain characteristics.

Studies of the abilities of various groups of people to fake certain speech characteristics would also be interesting. It is important to note here that, from an ecological perspective, evidence a characteristic can be faked is very positive. Fakeability implies the presence of something to fake ie the presence of reliable information on which people make judgements. A comparison of information structure between a real case (eg a Northern speaker) and a convincing replica (an actor) would allow researchers to identify the critical sources of information.

Further Study

The above section has already given some ideas for further research. To briefly recap, questions concerning the direct or indirect nature of judgements, the identification of intrinsic and optional characteristics, and the possibility of these characteristics being faked, would be interesting to study. Other possibilities for research arise from the fundamentally exploratory nature of this research. The present study was intended more to open up interesting questions than to provide any highly robust answers. This is a valid and useful way to start an investigation, but further studies should redress the balance to some extent.

The results suggest that there are up to five variables independently specified in the non-verbal information of speech. Further studies should investigate these variables in more detail, and with tighter experimental control. There are at least four areas worthy of further examination. First of all, the accuracy in judging these five variables should be further investigated - just how close, for instance, can listeners get to the actual age of the speaker? If some are later found to not be accurate, an interesting related question would be whether raters were at least *consistent* in their judgements.

The variables should also be investigated using different listener and speaker groups. It would be interesting to see to what extent these results are general or group specific (eg are age judgements generally less accurate with increasing age, or is this result best described by the difference between speaker and listener ages?). Also, as mentioned before, the effect of common non-typical cases should be looked at to increase our understanding of which variables are accurately specified in an ecological sense. Finally, the study and comparison of different types of speech episode, preferably ones more natural than a set passage recording, might be interesting. If the research interest was still non-verbal speech information one would obviously have to take some precautions when using more natural speech. This should not be too troublesome - researchers could, for instance, use speech in a language listeners were not familiar with and degrade recordings slightly so individual words are not actually distinguishable. This may mirror, to some extent, work currently being done on prosody (see introduction).

Implications and Applications

This study has shown that some judgements made about a speaker from voice recordings. This is a finding that should be investigated and extended in psycholinguistics. As mentioned in the introduction, there appears to be a tendency in this field to concentrate on the 'stereotype principle'. Perhaps some of the assumptions about the unjustified nature of judging a person from their voice should be reviewed. This study also sets some interesting new questions for acoustics and anatomical investigation ie how changes in the body's gross physical dimensions, speech apparatus, and possibly neural and nervous systems, lead to changes in speech production, and what these changes are.

Finally, although the study was not performed with any particular practical uses in mind, extensions of the work done so far can be seen to have many practical uses. Such information could be used to develop automatic, computerised systems to identify such characteristics. These may find uses in security ('voiceprints') , artificial intelligence and for more frivolous entertainment purposes. In the worlds of marketing and telesales, how a speaker's voice will be perceived, and the extent to which this is manipulable might be even more significant (and lucrative!) information. Applications may also be found in the casting and training of actors for radio, stage and screen. In these areas, an extension of the more qualitative, subject judgements studied in the first

section of this report may be useful because they are such a rich source of information.

Conclusion

This study has answered some questions, but has revealed many more and has probably created some of its own. What seems irrefutable is that speech can specify, non-verbally, reliable information about the speaker. It would seem likely that at least the five independently specified variables listed here are detectable with some accuracy. Some other variables may also exist. Even the ones already identified may differ from culture to culture and language to language. They may have also to be extensively revised in light of subsequent research. Nevertheless, what I hope I have done is open up an interesting, and relatively novel, area of research that others may be interested in continuing.