No articulation tests have been made but numerous observations of the performance on speech, ranging from rehearsals in the empty auditorium to addresses before a capacity audience, indicate that hearing is entirely satisfactory anywhere in the house and with an audience of any size.

The room seems to be entirely free from the excessive low-frequency reverberation and boom so often associated with large enclosures. The consultants have recommended the use of a suitable box set to provide better playing conditions for groups on the stage, and to make microphone positions less critical.

---

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOLUME 29, NUMBER 1 JANUARY, 1957

# Information Conveyed by Vowels

PETER LADEFOGED, *Phonetics Department, University of Edinburgh, Scotland*

AND

D. E. BROADBENT, *Medical Research Council Applied Psychology Unit, Cambridge, England*
(Received June 12, 1956)

Most speech sounds may be said to convey three kinds of information: linguistic information which enables the listener to identify the words that are being used; socio-linguistic information, which enables him to appreciate something about the background of the speaker; and personal information which helps to identify the speaker. An experiment has been carried out which shows that the linguistic information conveyed by a vowel sound does not depend on the absolute values of its formant frequencies, but on the relationship between the formant frequencies for that vowel and the formant frequencies of other vowels pronounced by that speaker. Six versions of the sentence *Please say what this word is* were synthesized on a Parametric Artificial Talking device. Four test words of the form *b*-(vowel)-*t* were also synthesized. It is shown that the identification of the test word depends on the formant structure of the introductory sentence. Some psychological implications of this experiment are discussed, and hypotheses are put forward concerning the ways in which all three kinds of information are conveyed by vowels.

IN recent years a great deal of research has been directed towards the specification of the "information-bearing elements of speech."[1] It seems that at the moment much of this research is hampered through lack of consideration of the kinds of information that are conveyed by speech. For the sake of convenience in exposition we may consider this information to be of three kinds. Firstly, when we listen to a person talking, we can receive information about what he is saying; in other words, we can appreciate the linguistic significance of the utterance. Secondly, in addition to the information we receive as a result of considering an utterance in terms of a linguistic system, we also receive information of a different kind about the general background of the speaker; thus we can usually infer something about a speaker's place of origin and his social status from his accent. This kind of information may be termed socio-linguistic; it is conveyed by the features of a person's speech which he acquires through the influence of the particular groups of which he is (or was) a member. Lastly there is the kind of information conveyed by the idiosyncratic features of a person's speech. These, like the group and linguistic features, may be part of an individual's learned speech behavior; but, unlike the other features, idiosyncratic features may also be due to anatomical and physiological considerations, such as the particular shape of the vocal cavities. The information which these features convey may be termed personal information. The relations between these three kinds of information are summarized in Fig. 1.

It is possible to arrange experimental situations which will elicit responses with respect to each of these three kinds of information. Thus one can ask a subject: Were these two sounds pronounced by the same speaker? (personal information); or: Is there any difference of accent between these two speakers? (socio-linguistic information); or: Do these two utterances consist of the same words used in the same way? (linguistic information). It is also possible to arrange a situation where the socio-linguistic information and the linguistic information will be assessed concurrently. These two kinds of information taken together are sometimes said to be equivalent to the phonetic value of a sound.[2] This point of view, however, is disputed by others who believe that "The phonetic

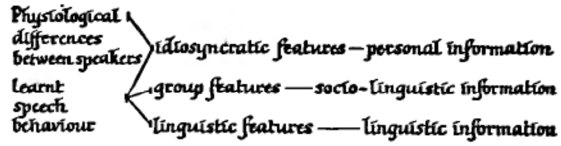


FIG. 1. Differences in utterances and the information that they convey.

[1] G. E. Peterson, J. Acoust. Soc. Am. 24, 629–637 (1952).

[2] Peter Ladefoged, Lingua 5, 113–127 (1956).

value of a speech sound is independent of language and meaning."[1]

In this article an experiment is discussed which is concerned with those features of vowel quality which convey linguistic information. This has led to some consideration of both the group and the idiosyncratic features of vowel quality, and tentative conclusions are reached concerning the ways in which all three kinds of information are conveyed.

### DEVELOPMENT OF EXPERIMENTAL MATERIAL

It has been supposed for almost a century now that the variations in the formants, or regions of the auditory spectrum in which there is a relatively large amount of spectral energy, are responsible for most of the information conveyed by vowels. However, there is as yet no general agreement on the precise properties of the formants which convey the information. There are differences of opinion as to whether the value of a given vowel depends on the absolute values of certain properties of its formants, or whether it depends on the relation between these values and the values for other vowels pronounced by the same speaker. As a proponent of the first view, we may instance G. E. Peterson, who analyzed a group of matched vowels obtained by recording the sounds produced by speakers trying to imitate the phonetic quality of the vowels in two reference words. He came to the general conclusion that "front vowels could be rather readily identified by observing the positions in frequency of the peaks of the first three formants." [1] Contrasting with this view is the theory propounded by M. Joos,[3] to the effect that the phonetic quality of a vowel depends on the relationship between the formant frequencies for that vowel and the formant frequencies of other vowels pronounced by that speaker.

A necessary part of Joos' theory is that whenever a listener to speech has to identify a vowel without the benefit of any clues from the context, he utilizes whatever knowledge he has of the speaker's formant frequencies in other words. Even when the vowel which the listener is considering is quite unlike any that he has ever heard that speaker produce before, he nevertheless focuses his attention not on the absolute values of the frequencies of the formants, but on the relations between those frequencies and the general ranges of frequencies which seem to be characteristic of the speaker. Thus unknown vowels are identified in terms of the way in which their acoustic structure fits into the pattern of sounds that the listener has been able to observe.

This part of the theorem has now been verified in an experimental situation. It has been found that subjects hearing a test word immediately after hearing a specified introductory sentence are greatly influenced in their identification of the test word by the range of the formant frequencies in the introductory sentence.

In order to carry out this experiment it was first of all necessary to obtain introductory sentences which were identical except in the ranges of their formants. This cannot of course be done by recording different people saying the same sentence, because the utterances are bound to differ in many ways. Accordingly it was decided to use synthesized speech, which can be precisely controlled in all respects. The particular instrument used for the purpose was the Edinburgh University Phonetics Department's copy of the Parametric Artificial Talking Device[4] developed at the Ministry of Supply Signals Research and Development Establishment. The essential parts of the device are a generator producing a pulse corresponding to the larynx pulse which serves to excite the vocal tract; four formant generators which respond to the pulse excitation; and a generator which will produce noise corresponding to the excitation in fricative sounds. This instrument will synthesize speech which can be specified in terms of six variables, but which nevertheless sounds so natural that recordings of some sentences are always confused with recordings of normal speech. The six variables which are normally specified are the intensity and frequency of the pulse excitation, the frequencies of the lowest three formants, and the intensity of the fricative noise. In order to set up the synthesizer so that it will produce an utterance, information depicting these variables as functions of time is painted on a glass slide. The slide is then scanned by a mechanism which produces six controlling voltages which vary with time. The voltages control the appropriate generators of the synthesizer so that a sequence of speech-like sounds is produced.

As well as the factors which are specified by the information painted on the glass slide, it is also possible to vary other factors, such as the frequency of the fourth formant, and the amplitudes and damping constants of all four formants; but no provision is made for controlling these factors as functions of time, and they were not in fact varied in the course of the experiment. In addition, it is possible to alter the frequency range over which each of the formant generators is operating. It was this facility that was used to produce the necessary variations in the introductory sentences.

Six versions of the sentence *please say what this word is* were synthesized with the PAT device. This sentence was chosen as a suitable introductory context because the formant frequencies of the sounds vary over a wide range. Formant one varies between the low value necessary to produce the /i/ in *Please* to the high value required for the /ɐ/ in *what*; and formant two varies

[3] M. Joos, *Acoustic Phonetics*, Supplement to Language 24 (1948).

[4] W. Lawrence, "The synthesis of speech from signals which have a low information rate," in *Communication Theory*, W. Jackson, editor (Butterworths Scientific Publications, London, 1955), Chap. 34.

TABLE I. Differences in the six versions of the introductory sentence: *Please say what this word is.*

| Sentence version | Differences from sentence 1 | Frequency range in cps Formant 1 | Formant 2 |
|---|---|---|---|
| 1 | ... | 275–500 | 600–2500 |
| 2 | F. 1. down | 200–380 | 600–2500 |
| 3 | F. 1. up | 380–660 | 600–2500 |
| 4 | F. 2. down | 275–500 | 400–2100 |
| 5 | F. 2. up | 275–500 | 800–2900 |
| 6 | F. 1. down F. 2. up | 200–380 | 800–2900 |

between the high value in the /i/ of *please* and the low value at the beginning of the /w/ in *word.*

In making all six versions of the introductory sentence the synthesizer was controlled by a single slide. Consequently the versions were identical with one another except for the variations which were introduced in the ranges over which the formant generators operated. The variations are summarized in Table I, which shows the highest and lowest values both of formant one and of formant two that actually occurred in each version of this sentence.

It is interesting to note at this point that despite the great acoustic differences between the versions they were all readily identifiable as the same sentence. Moreover, all the trained phoneticians who listened to the different versions agreed that the variations which had been introduced did not appear to make any significant difference in either the linguistic or the socio-linguistic information which was being conveyed. With the exception of version six, which did sound rather unnatural and could not be judged as a sample of normal speech, all the different versions sounded like the same sentence pronounced by people who had the same accent but differed in their personal characteristics.

In addition to these introductory sentences, four test words were synthesized. Each of these was of the form *b*-(vowel)-*t*. The formant frequencies for the middle of the vowel in each of these words are shown in Table II. The vowels in each of these test words were of comparatively short duration.

## TEST PROCEDURE

A short listening test was devised with the aid of recordings of the material which has been described in the previous section. This test was taken by sixty subjects. The first part of the test consisted of recordings of the test words *A*, *B*, *C*, and *D* arranged in a random order. There were ten items in this part of the test. Subjects were told that they would hear ten words, each of which might be either *bit*, *bet*, *bat* or *but*. They were instructed to tick the appropriate word on the answer sheets with which they had been provided. The means of the responses in respect of each test word are shown in Table III.

TABLE II. The frequencies of the first two formants in the four test words.

| Test word | Frequency in cps Formant one | Formant two |
|---|---|---|
| *A* | 375 | 1700 |
| *B* | 450 | 1700 |
| *C* | 575 | 1700 |
| *D* | 600 | 1300 |

Between each of the first five words in the listening test there was a short pause during which subjects were requested to count aloud from one to ten. This was done in an attempt to prevent the identification of a test word being unduly influenced by the auditory memory of the preceding word. The efficacy of this procedure is discussed in a subsequent section.

In the second part of the recording the test words occurred immediately after the various versions of the introductory sentence. Subjects were given the following written instructions:

> You will now hear a voice saying *Please say what this word is.* This will be followed immediately by one of the words: *bit*, *bet*, *bat*, *but*. Please tick the appropriate word on the answer sheet below. There are twelve test sentences in this part of the recording; after answering in respect of each, there will be a short pause, during which you will be requested to count aloud, slowly, from one to ten.

TABLE III. Means of the responses of 60 subjects for the ten words in the first part of the listening test.

| Test word | Number of subjects identified as: *bit* | *bet* | *bat* | *but* |
|---|---|---|---|---|
| *A* | 52 | 8 | ... | ... |
| *B* | 14 | 46 | ... | ... |
| *C* | ... | 27 | 33 | ... |
| *D* | ... | 1 | 14 | 45 |

The twelve items were arranged so that the predicted responses occurred in a random order. The results of this part of the test are shown in Fig. 2.

## DISCUSSION OF RESULTS

It will be seen from Fig. 2 that subjects are undoubtedly influenced in their identification of the test word by the auditory context in which it occurs. Thus word *A* is identified as *bit* by 87% of the subjects when it is preceded by version one of the introductory sentence; but as *bet* by 90% of the subjects when it is preceded by version two in which the first formant varies over a lower range. All that remains to be shown is that the influence of the introductory sentence is in accordance with the theory put forward by Joos concerning the relative nature of this aspect of vowel quality.

The relations between the formant structures of the vowels in a number of words can be conveniently represented by means of a formant chart which shows the frequency of the first formant at a time in the word when the formant structure is changing at a minimum rate plotted against the frequency of the second formant at the same time. In order to provide a basis for discussion, some of the vowels of one of the authors (P.L.) are shown in this form in Fig. 3. The symbols used are /ɪ/ as in *bit*, /ɛ/ as in *bet*, /a/ as in *bat*, /ʌ/ as *but*, /i/ as in *please*, /e/ as in *say*, /ɒ/ as in *what*, and /ɜ/ as in *word*. The axes in this and the subsequent diagrams have been arranged so that these acoustic charts can be easily compared with the vowel diagrams used by phoneticians. The scale used throughout is the Koenig scale.

The pattern formed by the vowels shown in Fig. 3 may be taken as a representation of one of the kinds of relationships which can occur. Bearing this in mind, we may now consider the relationships between the vowels in each of the six versions of the introductory sentence and the test words with which they were designed to be associated. Figure 4 presents these data; solid points lettered *A*, *B*, *C*, and *D* represent the test words, and the open circles indicate the vowels in the different versions of the introductory sentence.

| Test Word | Version | Formant Relation | Number of subjects identified as bit | bet | bat | but |
|---|---|---|---|---|---|---|
| A | 1 | = | 53 | 7 | | |
| | 2 | F.1. down | 4 | 54 | 2 | |
| B | 1 | = | 5 | 55 | | |
| | 2 | F.1. down | 1 | 57 | 2 | |
| | 3 | F.1. up | 58 | 2 | | |
| | 6 | F.1. down F.2 up | 7 | 46 | 6 | 1 |
| C | 1 | | | 25 | 35 | |
| | 3 | F.1. up | | 48 | 12 | |
| | 5 | F.2. up | | 23 | 37 | |
| D | 1 | = | | | 11 | 49 |
| | 4 | F.2. down | | 1 | 36 | 23 |

FIG. 2. Means of the responses of sixty subjects identifying the test words *A*, *B*, *C*, and *D* preceded by different versions of the introductory sentence.

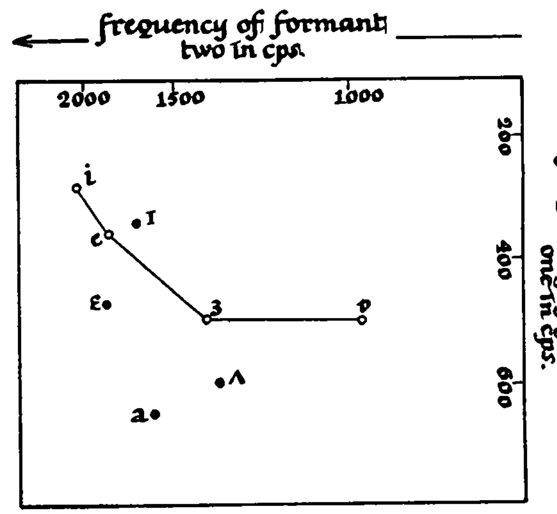


FIG. 3. The formant structure of some of the vowels of one of the authors (P.L.).

It will be seen from a comparison of Figs. 3 and 4 that when the test word *A* is associated with version one of the introductory sentence its relative position is similar to that of P.L.'s *bit*; and, in fact 87.5% of the subjects did identify it as *bit*. But when this word occurs in association with version two its relative position is more like that of P.L.'s *bet*; which accounts for the shift in identification whereby 90% of the subjects now consider it to be *bet*. Similar reasoning explains the change in identification of word *B* when it is associated with version one (92% *bet*) as opposed to version three (97% *bit*). But we must also note in connection with word *B* that when it was associated with versions two and six of the introductory sentence by far the majority of the subjects still identified it in the same way (i.e., as *bet*) as when it was associated with version one. The probable reason for this is that the relative position of the vowel /ɛ/ as in *bet* can be anywhere in a comparatively large area. As Daniel Jones[5] has noted: "The vowel (sc. /ɛ/) varies a good deal with different speakers." Presumably, therefore, the shifts in its relative position due to its being associated with versions two and six were not great enough to move it out of the part of the vowel pattern in which it is reasonable to expect to find a vowel of the /ɛ/ type.

The results shown in Fig. 2 indicate that there is a considerable amount of disagreement concerning the identification of word *C*. Some of the reasons for this can be appreciated from a comparison of the data presented in Figs. 3 and 4. Only when it is associated with version three of the introductory sentence does the vowel in this word have a relative position which is comparable with any of the relative positions of P.L.'s vowels. In these circumstances 80% of the subjects did identify it as the same word, *bet*. But when it occurs in

[5] Daniel Jones, *An Outline of English Phonetics* (W. Heffer, Cambridge, England, 1956).

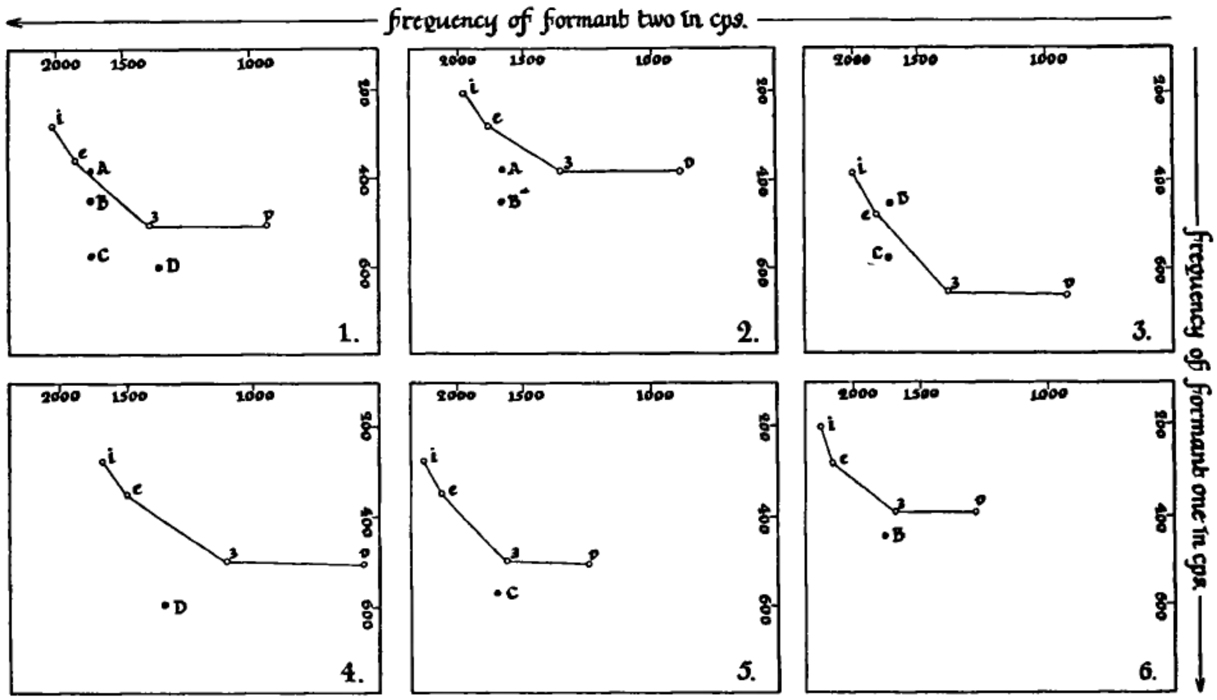


FIG. 4. The formant structure of the six versions of the introductory sentence and the test words that they were each designed to be associated with.

association with version one, where it occupies a relative position only slightly nearer the point in the pattern occupied by P.L.'s *bat* than his *bet*, it is not surprising that 58% of the subjects identify it as *bat* and 42% as *bet*. The results obtained from the association of this word with version five, however, are not so readily understandable. It might be expected that at least a small proportion of the subjects would identify this word as *but* in these circumstances. But in fact this did not happen, perhaps because this simplified treatment in terms of the frequencies of only two formants is not sufficient to account for the differences between these two words. On the other hand, tests with word *D* show that it is possible for the auditory context to influence the identification of a given test word so that it can be taken to be either *bat* or *but*. When this word was associated with version one, the majority of the subjects identified it as *but* (82%) as opposed to *bat* (18%); but in association with version four, in which the second formant was comparatively lower, then the results were *bet* (2%) *bat* (60%), and *but* (38%). Thus word *D* illustrates the fact that shifts in the range of the second formant in the introductory sentence can produce alterations in the identification of the test word which are of the same order as those produced by variations in the range of the first formant.

Taken all together, the results of this test seem to shown quite conclusively that, as Joos has said, the linguistic information conveyed by a given vowel is largely dependent on the relations between the frequencies of its formants and the frequencies of the formants of other vowels occurring in the same auditory context. It is, therefore, only of limited service to look for common points in the acoustic structure of equivalent vowels spoken by different speakers.

## PSYCHOLOGICAL IMPLICATIONS

It is obvious that this experiment provides a demonstration of perceptual constancy in the auditory field; that is an auditory phenomenon somewhat parallel to the visual case in which the response evoked by a stimulus is influenced by the stimuli with which it is closely associated. An example is the correct identification of the color of an object in widely differing illuminations. Consequently it is hoped that further investigation of the auditory phenomenon will provide data which are of general psychological interest.

There are many factors which have not been considered in any way in the course of the present experiment. For example, at the moment nothing is known about the length of the introductory sentence which is necessary in order to influence the identification of the test word. Nor do we know to what extent it is necessary to use an introductory sentence containing a wide variety of vowels which may serve as reference points. In addition, further evidence is required concerning the necessary degree of proximity between the auditory context and the stimulus word. A preliminary experiment has been reported[6] in which it is shown that if there is a ten-second silent interval

[6] Broadbent, Ladefoged, and Lawrence, Nature, **178**, 815–816 (1956).

between the introductory sentence and the test word, then the influence of the introductory sentence is significantly less. But the precise temporal limitations of the phenomenon have not yet been established.

Further research is also needed to determine the value of the procedure of counting between items in a test in order to weaken the auditory memory of the preceding words. Subjects were requested to count aloud between the first five items presented in the test. But despite this they were probably influenced in their identifications of the fifth word presented to them by their memory of the previous items. Both the first and fifth items were word $A$. When they first heard this word 45 subjects identified it as *bit* and 15 as *bet*; but when it occurred as the fifth item (i.e., after they had heard other test words) all 60 subjects identified word $A$ as *bit*. On the other hand, counting items had a significant effect on some occasions. Word $C$ was identified as *bet* by 37 subjects and as *bat* by 23 subjects when it occurred as the fourth item in the test, and could not directly be compared with word $B$; but when word $C$ occurred immediately after word $B$ it seems likely that the two words were judged together, since in these circumstances 10 subjects identified word $B$ as *bit* and 50 as *bet*, and 8 subjects identified word $C$ as *bet* and 52 as *bat*.

## SOCIO-LINGUISTIC AND PERSONAL INFORMATION

All the responses demanded by the listening test which has been described above are specifically related to linguistic information. But, on the basis of this test, two points may be noted concerning the socio-linguistic and personal information conveyed by vowels. Firstly, as we have mentioned, there do not appear to be any differences in the socio-linguistic information conveyed by the different versions of the introductory sentence. It therefore seems to be a plausible hypothesis that socio-linguistic information does not depend on the absolute values of the formant frequencies, but is, like linguistic information, a matter of the relative formant structure of vowels. Secondly, there is tentative evidence that subjects belonging to different socio-linguistic groups gave different responses to some of the test material. Consideration of the precise criteria that were used in dividing subjects into groups in accordance with their accents is, unfortunately, outside the scope of this article. It must suffice to state that there were three main groups: in one there were seven subjects who had what is known as a Basic Scots vowel system[7]; in the second there were nineteen Scottish speakers who had vowel systems that had been slightly modified due to the influence of the English of England: and in the third there were nineteen subjects who were speakers of the form of English of England known as R.P. Table IV shows the responses of each of these three groups in respect of test word $D$ preceded by version one of the introductory sentence. It can be seen that there is a greater tendency among the Scottish speakers to favor the identification of this word as *bat*, presumably because in their speech the relative position of the vowel in *bat* is similar to that of the vowel in the test word. The relation between accent and type of response is only just statistically significant, using tau,[8] $p=0.05$, and further confirmation is desirable.

TABLE IV. Identifications of test word $D$ in association with version one of the introductory sentence by different groups of subjects.

| Number in group | Character of group | Identified as: *bat* | *but* |
|---|---|---|---|
| 7 | Scots | 3 | 4 |
| 19 | English influenced Scots | 4 | 15 |
| 19 | English (R.P.) | 1 | 18 |

On this basis it seems at least possible that both the linguistic and the socio-linguistic information conveyed by vowels depend largely on the relative positions of the formants. When we consider that a speaker has vowel sounds which are typical of a Scottish speaker (i.e., when we interpret the socio-linguistic information conveyed by his vowels), we probably do so by appreciating the relative formant structure of the vowels.

On the other hand, the personal information conveyed by vowels does seem to depend partly on the absolute values of the formant frequencies. Thus all the versions of the introductory sentence sounded as if they had been spoken by different voices. The reasons for this are best understood by reference to the articulatory processes involved in speech. The formants of a sound are essentially properties of the shape of the vocal tract. Consequently the ranges over which a speaker's formants can vary depend to a great extent on the size of his head. Because the ranges cannot be altered at will, they are not part of a speaker's learned speech behavior, and can therefore convey only personal information. Additional personal information is, of course, conveyed by the relative positions of some of a speaker's vowels, insofar as these are idiosyncratic features of his speech and not aspects which identify him as belonging to a particular group.

Finally, it is interesting to consider the current usages of the term "phonetic quality" in the light of the hypotheses that have been put forward above in connection with the ways in which all three kinds of information are conveyed. It seems that when phoneticians talk about the quality of a speech sound they do not always mean the same thing. When teaching the pronunciation of the vowels of a foreign language, for instance, they assess their pupils' utterances in terms of the linguistic and socio-linguistic information that they convey; they are therefore concerned with the relative formant structure only.

[7] D. Abercrombie and A. J. Aitken, *A Scots Phonetic Reader*, (to be published).

[8] J. W. Whitfield, Biometrika 34, 292–296 (1947).

But on other occasions when observing utterances, particularly in field-work situations, phoneticians seem to imply that vowel quality can be judged in relation to certain absolute standards. Many phoneticians believe that it is possible to describe the quality of a vowel in an isolated monosyllable spoken by an informant, even if they have no other information concerning the speaker.

It is not yet known whether the cardinal vowels which serve as reference points for phoneticians are really precise points with fixed acoustic specifications, or whether they are only a set of vowels which have the same relative formant structure when pronounced by different phoneticians. But whatever their nature, it seems that the whole theory underlying the methods which are used in practice for describing vowels needs restating, so that it is made clear whether linguistic and socio-linguistic criteria are being used in making assessments of quality, or whether an attempt is being made to classify speech sounds in terms of purely auditory criteria.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the help of many of their colleagues. In particular, W. Lawrence, D. Abercrombie, and J. Anthony have been of great assistance.

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOLUME 29, NUMBER 1 JANUARY, 1957

# Realization of Prosodic Features in Whispered Speech*

W. MEYER-EPPLER
*Institut für Phonetik und Kommunikationsforschung der Universität, Bonn, Germany*
(Received September 26, 1956)

Experiments utilizing a visible-speech analyzer showed that changes of pitch in normal (voiced) speech are replaced in whispered speech by shifts of some formant regions accompanied by added noise between the higher formants.

## INTRODUCTION

It is a well-known fact that people can be understood without any difficulty when they whisper instead of speaking normally. This fact is not very strange if the formant frequencies of the vowels and the envelopes and spectra of the fricative and plosive sounds are considered to be the information-carrying elements of speech. It must be doubtful, however, whether full information can be carried by whispered speech in tonal languages like Chinese or many West-African languages where pitch is used to differentiate the meaning of various lexical items consisting of otherwise identical groups of phons.

Recently, Panconcelli-Calzia[1] and Giet[2] have dealt with the problem of whispering in tone languages. Whereas Panconcelli-Calzia found that it was difficult for Chinese-born subjects to understand whispered Chinese, Giet, who had lived in China for many years as a missionary, states that whispering in Chinese is as effective a means of verbal communication as normally spoken speech. According to Giet's arguments[3] there must exist some substitute for the missing pitch quality in whispered speech within the acoustical range. As Giet has already pointed out, there is no need to use tone languages for investigating substitutes in whispered speech. Similar results would be achieved by using any language where intonation belongs not to the *phonemic* but to the *prosodic* level, which term refers to features belonging to a sentence as a whole that are expressed by pitch and stress patterns. Intonation, e.g., may differentiate between a question and a statement.

## VOWELS "SUNG" WITHOUT VOICE

Some orienting investigations were undertaken with German vowels "sung" without voice. It is not difficult to produce the same whispered vowel on different pitch levels within a range of about a musical fifth (i.e., a frequency ratio of 2:3). Obviously this can only be done by changing the spectral structure of the vowels within the limits of recognizability. The subjects were asked to "sing" the first five tones of a diatonic scale (e.g.: *c*, *d*, *e*, *f*, and *g*) maintaining the quality of a given vowel as well as possible. The sounds were recorded on magnetic tape and analyzed by means of a visible-speech analyzer (Sona-Graph). The spectrograms of a test series using the German vowels [a] (as in *Tal*), [e] (as in *See*), [i] (as in *viel*), [o] (as in *Sohn*), and [u] (as in *Schuh*) are shown in Fig. 1. Whereas in the case of [a], [e], [i] and [o] the position of the first two formants remains unchanged, the third formant of [a] is shifted from its position near 2.5 kc to about 3 kc if higher pitch is intended; a similar shift is found at a weak fifth formant near 5 kc. In the case of [u] the

* This work was supported in part by Gesellschaft zur Förderung der Klangforschung, Cologne, and was presented at the Second ICA Congress, June, 1956.
[1] G. Panconcelli-Calzia, Lingua 4, 369–378 (1955).
[2] Franz Giet SVD, *Zur Tonität nordchinesischer Mundarten* (Verlag der Missionsdruckerei St. Gabriel, Wien-Mödling, 1950), pp. 95–97.
[3] Franz Giet, Lingua 5, 372–381 (1956).