Proceedings of the Institute of Acoustics

RESURRECTION OF SPEECH INTELLIGIBILITY TESTS IN AUDIOLOGY

R D WRIGHT

Alfred Peters Ltd, Industrial Estate, Hunmanby, N Yorkshire UK

1. SUMMARY

This paper describes how hearing aids are selected, and the general lack of use of speech materials in the selection process. The history and current status of speech tests is reviewed, in general and within audiology. Finally, equipment and procedures are described which are designed to make speech tests sufficiently fast and accurate to be useful in hearing aid fitting.

Academic papers are as boring as they are unread. To be more interesting and readable, here is the simplified content of this paper, amplified and, I hope, justified in the sections below:

A. Nobody knows how well a hearing aid works on speech,
B. because nobody bothers to measure speech intelligibility,
C. because standard speech tests are slow and unreliable.
D. Speech tests can be faster and somewhat more reliable;
E. it requires a computer, speech materials with random access (meaning digitised), closed response materials, and automatic scoring, which nobody in a clinic ever has;
F. unless, of course, it is our new AP600 speech audiometer.

2. HEARING AID FITTING

It is common pracitice in the UK for a person to be fitted with a hearing aid without ever having any objective measurement of the value of the aid for speech communication. For example, Green [1] states "... speech audiometry is a relative rarity in rehabilitation ...". Persons from a communication engineering or general speech processing background may find this situation surprising, particularly as the principal motivation for a person with a moderate hearing loss to seek a hearing aid is to improve speech reception.

In general communication engineering, communication channels are tested for speech intelligibility, following procedures which have been under development at least since the 1920's [2]. However a hearing impaired person usually receives an aid which is 'evaluated' only on detection thresholds for pure tones, not on speech.

An aid is evaluated using pure tones is simply because, in the clinic, a person's 'hearing' is measured on detection thresholds for pure tones (an audiogram). Quotes are used in the previous sentence to remind us that hearing is more than threshold detection of (relativvely long) pure tones. Having defined hearing loss in terms of pure tone detection, it is a short step to defining remediation also in terms of pure tone detection, and speech need not ever come into the audiological picture.

Indeed, there are elaborate strategies for specifying the characteristics of a hearing aid entirely from pure tone data (for example, reference [3]; there are many others, including various computer programmes). These strategies form the basis for the prescriptive approach to hearing aid fitting: measure an audiogram, apply some equations, and the hearing aid prescription pops out. This approach can be carried a step further: the hearing aid is fitted, and the person is tested again (using loudspeakers, because audiometer headphones con't sensibly be used over a hearing aid). This re-test is simply to determine how the aid affects pure tone detection. In the clinic, checking aided thresholds is considered being very thorough, going well beyond just determining the prescription and handing out the aid which comes closest to the prescribed performance. In a few clinics, where time and money allow, measurements of the amplification of the hearing aid are made while the aid is actually being worn (in-the-ear hearing aid testing). An aid which produces the gain which the prescription specifies is then described (in all seriousness) as an aid which is providing 'full benefit'. Actual benefit as defined in terms which are not obviously circular should involve consideration of something rather more useful than pure tone detection, but often does not. If 'real-world' benefit is seriously considered, it is often assessed by questionnaire rather than by intelligibility tests.

3. SPEECH INTELLIGIBILITY TESTS

The basic idea of measuring speech intelligibility is deceptively simple: test how well a person recognises what is being said. In practice, speech can be anything from a wordlist to a paragraph, and the results are highly dependent on the material used (especially its redundancy or predictability), how the test was conducted and how it was scored.

Although redundancy is vital to speech communication, intelligibility tests tend to use materials which largely eliminate redundancy. The principal method is use of lists of words spoken in isolation, eliminating sytactic, semantic and pragmatic constraints (though NOT phonological constraints).

There are two large areas of problems associated with these 'sim ple' word list tests: (1) linguistic complications; (2) learning effects.

An example of the linguistic complication has already been given: a person tends to say the nearest known word. Items in the word list thus vary in difficulty depending upon the vocabulary of the listener, as well as for more acoustical reasons such as the spectral characteristics of the consitituent sounds. For a word to be 'heard and repeated' requires the proper functioning of the entire linguistic heirarchy of both perception and production. An error at any stage causes the word to be 'not heard', even if a better explanation would be 'word not known', or 'person has difficulty pronouncing the word', or even 'person conducting the test has difficulty recognising the speech sounds of the person being tested'.

The problem of learning effects is really one of response set. If a word list is re-ordered and repeated, a very large response set reduces to a very small one. Scores improve dramatically, and the test becomes a measure of the subjects memory and intelligence, rather than the subjects hearing.

One way to eliminate some of the linguistic complication, developed particularly for speech tests for children, is to have the subject point to a picture (one of a small set) rather than 'say what was heard'. The main linguistic complication is then the matter of vocabulary.

There are other differences between a 'say what your heard' test and a picture-pointing test. The response set is small, and totally defined. This means that memory is no longer an issue, so the test items can be used repeatedly. It also means that 'errors' are confined to a small number of possibilities, so errors can be classified. With careful test design, the errors can begin to shed light on just what sounds, or spectral components, or sound categories (distinctive features) are or are not involved in errors. The error analysis can be much more illuminating than the simple measure of tokens correctly responded to.

Word lists which define the response set are called closed-response tests, and have been used for more than 30 years in general assessment of communication systems, beginning with the Rhyme Test [4]. The analysis of errors for 'diagnostic' implications developed throughout the 1060's and 1970's, and the Diagnostic Rhyme Test [5] is now very commonly used for intelligibility testing purposes on both sides of the Atlantic. However, it has only rarely been used in audiological research, and to the best of my knowledge is never used in audiology clinics.

One diagnostic closed response test which has been widely used in audiological research is the FAAF test [6], designed at the UK Institute for Hearing Research, but again there is little if any clinical use.

4. SPEECH TESTS IN AUDIOLOGY

The standard approach in clinical audiology is to use open set word lists. In this respect there has been no progress in standard practice in 30 years. Indeed, some of the materials in use originated over 50 years ago. The Psycho Acoustic Laboratory of Harvard University published the PAL 20 PB-50 test materials in 1948 (also called the Harvard word lists). These were modified at the Central Institute for the Deaf in 1952, forming the CID W-2 and W-22 wordlists, which are not only still in use but remain the '... most extensively used monosyllabic (open set response) test in the United States'[7]. In the UK the same approach prevails, but using material developed in the UK. The materials currently on offer from the Royal National Throat Nose and Ear hospital in London are the Fry (1939, 1961), Booth-royd (1968) and MRC (1947) wordlists.

These speech materials are becoming so dated that the master tapes are wearing out, and there is a current crisis in the UK audilogical community over what to re-record, and on what format (and who pays the bill)[8].

An additional source of uncertainty in current audiological prac tice is the question of the purpose of speech tests. We have been discussing evaluation of a hearing aid, but in audiology there are several areas where speech tests might be used:

1. disability assessment: using speech and no hearing aid, to measure the effect of a hearing loss on the ability to discriminate speech.

2. treatment evaluation: testing aided vs unaided speech perception.

3. treatment comparisons: testing one aid vs another, or one parameter setting vs another. This aspect is becoming increasingly relevant as hearing aid parameters proliferate.

4. diagnosis: using a graph of speech discrimination score vs level to help determine the nature of the hearing loss.

The first three appear straightforward applications of speech test technology, but the fourth, the diagnostic use, is less expected. Curiously, it was this fourth category which motivated the bulk of clinical speech test for about twenty years, from about 1960. If word lists were played at various signal levels, a graph could be formed of recognition score vs signal level. In audiology this record is called a speech audiogram, and the shape of the result on the basic six-way differentiation required for diagnosis: 1- normal; 2- non-organic loss; 3- conductive loss; 4- sensory disorder; 5- peripheral-neural disorder; central auditory disorder.

Because speech was used for diagnosis of hearing disorder, the original and more general use of speech materials has been rather lost track of in recent years, at least in the UK. Then, especially since about 1980, other procedures developed (such as brainstem potential measurements, and computer-aided tomography) which were much more powerful and definite diagnostic measures than were speech tests[9]. With the demise of speech tests for diagnosis has come the general decline of speech tests for any clinical audiological purposes [10].

5. HOW TO IMPROVE SPEECH TESTS

There have been several advances in test technique over past dec ades which could be applied to the problems of clinical speech audiometry. Five general areas can be mentioned:

1. Comparative rather than absolute measures.
2. Control of signal to noise ratio (SNR).
3. Use of closed response set materials.
4. Automatic recording of the subject's response.
5. Adaptive measurement techniques.

The first two points both relate to the problem of variability of speech reception test scores. Speech tests can be perverse: insensitive over large variations in SNR, and then suddenly oversensitive. If a person is tested 'aided and unaided', there will be gross differences in signal level within the subject's ear canal (the whole point of an aid is amplification), but SNR will be uncontrolled. This is a general problem in clinics: plenty of opportunity for changing levels (wittingly or not), but often no method to control SNR. The result is that absolute measures of speech reception ability can vary enormously, depending much more upon presentation level (and resultant but unknown SNR) than upon the patient's absolute level of disability, or the patient's absolute amount of 'aided benefit'.

The conclusion is that speech is difficult to use for absolute measurements as in points 1 and 2 in section 4, above. The road to accurate, repeatable results is to concentrate on relative measurements: on the change in speech scores occasioned by two different test conditions, all else being equal.

Further, rather then letting SNR be determined by whatever falls out from a combination of signal level and noise floor of the test suite or audiometer or hearing aid, SNR should be controlled directly by mixing noise with signal at the audiometer. This single step eliminates the main source of variation in scores. The graph of 'percent words correct' vs SNR has a slope of 6%/dB at the steepest point [11], for monosyllabic wordlists. Thus it only takes a 2 dB difference (barely audible to a person of normal hearing) between presentation levels to account for the 12.8% test-retest variability which has led some audiologists to abandon speech materials [12].
The importance of closed response lists was mentioned in section 3, above. Memory is eliminated as a source of variation, and errors can be tabulated and analysed. Additionally, the subject can simply press a button on an appropriate response panel (or box or screen) and the test becomes fully automated. The test system presents an item, the subject presses a button, the system knows what the selection was and can keep score (including error analysis) automatically, and the test proceeds without need for a clinician to note and record the subject's responses.

Further, the test system (if it has been programmed with the appropriate algorithm) can do something rather more clever than simply plod through a word list to determine percent correct, and then adjust the SNR and repeat the process until eventual it produces a speech audiogram. If the whole point of the speech audiogram is for comparative measures, and if the point of comparison can sensibly be defined as, for instance, the SNR which gives 50% (or 70% or whatever is required) correct responses, then the system can change the SNR after much less than a full wordlist. In the limit, the system can adapt after every item, increasing SNR for incorrect responses and decreasing it for correct ones, and simultaneously decreasing the SNR stepsize in proportion to how far current results are from the target performance.

These adaptive tests have become a standard tool in laboratory psychophysics [13], an only await the availability of the proper combination of test materials, audiometer and computer (as described in the next section) to perform similarly in the clinic.

6. IMPROVING AUDIOMETERS TO AID SPEECH TESTS

The elements required to improve speech tests in clinical audiology have already been discussed. The principle ingredients are:

1- digitised speech materials
2- with full random access
3- and automatic recording of patient responses
4- allowing use of adaptive test strategies.

Digitised speech materials are specified for several reasons: it is a necessary (though not sufficient) condition for random access, and it helps eliminate some acoustical variability (because signal quality will not vary according to, for example, the amount of wear on the recording, or the condition of the tape heads), and it simplifies control of SNR (because signal level can be unambiguously specified, and maintained as a constant).

Random access eliminates the need for multiple recordings of dif ferent 'orderings' of the same word list. In closed response tests the items can be repeated, but the order must vary or there will be the possibility for the subject to remember the order and improve her or his score by use of cognitive rather than auditory processing.

Automatic recording of test responsis is fundamental to adaptive test strategies, as discussed in section 5, above. Another fundamental requirement is the need to control SNR with adaptive step sizes. This requirement means that the basic attenuation charactistics of the device should NOT be limited to the conventional 5 dB steps, but should be controllable down to something like 1 dB.

Finally, an 'automatic audiometer' can go on to make measurements unheard of in conventional audiometry, such as measuring not only whether the subject gets the right or wrong response, but how long it took to respond. Such reaction time measures can add considerably to the sensitivity (or, equivalently, the reliability) of the test results [14].

7. THE AP600 SPEECH AUDIOMETER

An audiometer fulfilling all the requirements for advancing the prospects of speech tests in clinical audiometry is under development at Alfred Peters Ltd. This project has funding from the UK Dept of Trade and Industry, which we gratefully acknowledge. The basic electronics were developed in the year up to Sept 1992, and in the current year we are concentrating on integration with a control computer, and software development.

The system will be able to use any required method for access to digitised data, from floppy discs to CD-ROM. This versatility is obtained by the simple expedient of building a 'PC-compatible' motherboard into the audiometer, so that any PC option (hardware or software) can be incorporated within the system.

There is no conventional audiometer front panel, but neither is there a computer keyboard. A special 'audiometer joystick' allows full control over all the convention (and unconventional) audiometric functions, using a visual display (graphical user interface) for interaction with the operator. The result is a system which uses CAA (Computer Aided Audiometry) in a clinic-friendly package to solve longstanding problems in use of speech materials in hearing aid fitting and assessment.

8. REFERENCES

[1] R GREEN, 'The Uses and Misuses of Speech Audiometry in Rehabilitation'; in M C Martin (Ed) SPEECH AUDIOMETRY, London: Taylor and Francis (1987)
[2] H FLETCHER and JC STEINBERG, 'Articulation Methods', Bell Systems Technical Journal, 8, 806-854 (1929)
[3] K BERGER, NS HAGBERG and RL RANE, PRESCRIPTION OF HEARING AIDS, Kent OH: Herald Publ Co (1977)
[4] G FAIRBANKS, 'Test of phonemic differentiation: The Rhyme Test', J Acous Soc Amer, 30, 596-600 (1958)
[5] WD VOIERS, 'Diagnostic Evaluation of Speech Intelligibility', in M HAWLEY (Ed) Speech Intelligibility and Speaker Recognition, Stroudsburg PA: Dowden Hutchinson and Ross (1977)
[6] JR FOSTER and MP HAGGARD, 'FAAF: An efficient analytical test of speech perception', Proc Inst of Acous, IA3, 9-12 (1979)
[7] A MARKIDES, 'Speech Tests of Hearing for Children'; cf [1]
[8] HC FULLER, National Physical Lab. (personal communcation)
[9] PIP EVANS, 'Speech Audiometry for Differential Diagnosis' [1]
[10] HC FULLER and IK MOSS, 'A survey of speech audiometry in the National Health Service', NPL Acoustics Report AC 105 (1985)
[11] GA MILLER and P NICELY, 'An analysis of perceptual confusions among some English consonants', J Acous Soc Amer, 27, 338-352 (1955)
[12] DM SCHWARTZ, 'Hearing aid selection methods: an enigma', in GA STUDEBAKER and FH BESS (Eds) Vanderbilt Hearing Aid Report, Monographs in Contemporary Audiology, Upper Darby PA (1982)
[13] H LEVITT and S RESNICK, 'Speech perception by the hearing impaired: methods of testing and development of new tests', in C LUDWIGEN and J BARFORD (Eds) Sensorineural Hearing Impaired and Hearing Aids. Scandinavian Audiology, Suppl 6, 105-130 (1978)
[14] R WRIGHT, M SPANNER and M MARTIN, 'Pilot Experiments with a Reaction Time Audiometer', Br Jr of Audiology, 15, 276-281 (1981)