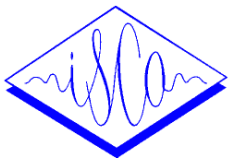

Speech Input/Output Assessment
and Speech Databases
Noordwijkerhout, the Netherlands
September 20-23, 1989

DATABASES INCORPORATING SPONTANEOUS SPEECH FROM FLUENT AND DISFLUENT SPEAKERS

Peter Howell, Michael Johnson, Karima Kadi-Hanifi, Pippa Bark,
Patricia Hanke, Celia Bonnett & Trudie Wingfield.

Department of Psychology, University College London,
Gower St., London WC1E 6BT,
England.

ABSTRACT

Some major concerns in the speech pathology field are (1) what are the defining characteristics of particular types of disfluency, (2) do normally fluent speakers show the same disfluencies as speakers with particular speech disorders, (3) how can changes in the speech of these speakers, both with regard to responses to therapy and developmental changes in the disorder, be assessed? The databases which we have set up in order to begin to answer these questions and an illustrative analysis are described.

1. INTRODUCTION

The questions that our research group is trying to answer are (1) what types of disfluencies occur in various forms of disordered speech, (2) for a particular disorder what determines whereabouts these disfluencies occur within the stream of speech and (3) for each type of disfluent speech, what is the relative frequency of the various types of disfluency? It is intended that information on these three principal issues be used to specify the defining characteristics of various disorders. This, in itself, is not a trivial problem as attested to by the lively debate about whether stuttering and normal nonfluency are canonically different or not (see Wingate's, (1988) critical assessment of Wendell Johnson's position. Johnson was primarily responsible for the view that there is no difference between stuttering and normally fluent speech).

Many of the subsidiary questions which we wish to deal with require that analysis be performed on spontaneous speech. One thing which we want to characterize is the severity of a speaker's disorder. If grammatical factors determine the occurence of disfluency-types and if linguistic complexity varies between speakers, then the characterization of severity would be unsatisfactory if it depended upon standard read material which does not have the same complexity as that which the speaker would choose to use. The characterization of disfluency in spontaneous speech is necessary when studying all the interesting developmental questions: for example, to assess whether changes in the frequency of disfluency-types and the points where the disfluencies occur in the speech are due to changes in the progress of the disorder or developmental changes in the speaker's speech. A final concern has been with whether characteristics of the disfluencies in a particular disorder

depend on the language in which it is spoken: clearly languages differ in the inventory of phonemes, prosodic structure etc., so they offer a convenient way of assessing how the distribution of disfluencies is determined by these factors.

Surprisingly, there has been relatively little attention to whether disfluencies which occur in the speech of normally fluent speakers occur in similar locations as those in disfluent speech. The exceptions to this are that it is known that pauses tend to occur at clause boundaries and two reports which claim to show that the constraints which determine where stutterings occur in stuttered speech also apply to normal speech (Chaney, 1969; Silverman & Williams, 1967). Thus, the control analyses of normal speech should provide much-needed information concerning where and why the forward flow of spontaneous speech by normal speakers is interrupted.

2. SPEECH RECORDINGS

Databases of speech recordings have been made. The number of speakers and their ages, the length of the recordings, the type of speech (read/spontaneous conversation), language of the speaker, type of disorder the speaker has if any and whether anything special was done at the time the recordings were made (e.g., speaking along with a metronome) are summarised in Table I.

Table I. Existing databases

No.	No of speakers	Length	Type	Language	
Stuttered databases:					
1	10	2 x 613 words	Read	English	
Once normally and once under delayed auditory feedback (DAF)					
2	10	2 x 613 words	Read	English	
Once normally and once with a metronome running at 2 Hz.					
3	10	2 x 613 words	Read	English	
Speakers read the passage under normal and with feedback shifted down an octave.					
4	33	1/2 hr (av)	Spon	English	
21 Adults and 12 Children					
5	6	1/4 hr (av)	Spon	Spanish	Children
Autistic databases:					
6	5	1/2 hr (av)	Both	Spanish	Children
7	5	1/2 hr (av)	Both	English	Children
Normal control databases:					
8	5	1/2 hr (av)	Both	Spanish	Children
9	5	1/2 hr (av)	Both	English	Children
10	68+	5,000 words per pair of speakers	Spon	English	Survey of English Usage see Svartvik & Quirk (1980)

In addition, recordings of 12 normal speakers differing in the clarity of their speech have been recorded with clarity of the speech confirmed by perceptual tests. The latter are currently being encoded.

3. MACHINE ENCODING OF THE DATABASE

All of the data referred to except the Survey of English Usage data are held as analogue recordings on magnetic tape. The initial encoding of these data is as machine readable phonetic transcriptions and they are entered into a machine running Unix. Overall the transcriptions are phonemic but with a narrow phonetic transcription in the vicinity of a disfluency. The transcription conventions at the phonemic level are based on that employed at the Speech Research Unit at the Royal Signals and Radar Establishment, Malvern. This has been extended to code important aspects associated with disfluent speech (the disfluent sequences currently distinguished include prolongations, repetitions, pauses, phonetic mergers, excessive aspiration, interruptions and abandonment of a sequence) and to incorporate markers for factors known to influence the distribution of disfluencies (e.g., content/function word distinction and syllabification following a simplified version of Clements and Keyser, 1983). Timing and loudness are marked and the conventions for pitch markings are currently being upgraded to that described in Crystal (1969). (Further details available on request).

The transcribed material can be replayed through a text-to-speech system that has been modified to accept the above conventions. This is used (1) by the transcribers to check their transcriptions and (2) for the preparation of tests to validate whether listeners regard some component as a perceptually salient characteristic of the speech disorder. Tests are made Osberger and Levitt's (1979) transformation method. A demonstration tape is available.

Besides the transcription software, statistical tools for the manipulation of these data sets exist and it is envisaged that more will follow. Special mention should be made of the reliability assessments which are performed by transcribing stretches of speech twice and for all components of the transcriptions (prosodic, phonemic etc) a measure of inter-observer agreement (Cohen's Kappa) is calculated.

4. ILLUSTRATIVE ANALYSIS OF MATERIAL IN THIS DATABASE

It was mentioned earlier that one of the uses of the database concerns assessment of techniques employed in therapy. One question examined already is whether altering auditory feedback of a stutterer's voice reduces the incidence of disfluencies? Howell, Wingfield and Johnson (1988) compared the effects of stutterers speaking normally and either with a concurrent metronome click running at 2 Hz or along with DAF at a 200 ms delay. Various analyses were performed to establish what types of disfluencies occurred and where. In speech spoken without alteration to feedback, fewer disfluencies occurred on function than on content words. Disfluencies occurred at early positions in words and syllables. When subjected to the two forms of alteration, several changes in the pattern of disfluencies occurred. There was no change to the ratio of repetitions to prolongations under either type of feedback but with both the alterations, a phoneme was repeated fewer times than under normal feedback conditions. The alterations shifted the distribution of disfluencies causing them to tend to occur later in words and

syllables than under normal feedback conditions. Under the normal condition disfluencies occurred more often on content words whilst either alteration reduced this tendency. Effects specific to the alterations were that a 200-ms auditory feedback delay increased the number of disfluencies and pause length increased. Speaking along with a metronome reduced pause length. These analyses show that the effect of these types of alteration are complex and substantially different for the two alterations discussed. Thus, in answer to the initial question posed, these alterations do not reduce the incidence of disfluencies; neither is providing a speaker with an experience of speaking normally and yet both of them produces some improvement in fluency. The question which arises is whether alterations which do not have side effects on the speech produced can be devised.

ACKNOWELDGEMENTS

The initial funding for this work was provided by the Wellcome Trust (MJ, PB, TW and CB). The developmental work described has been supported by the Medical Research Council of the United Kingdom (KK-H). The work on autistics, retardates and their controls was supported by a grant from the Central Research Fund of the University of London (PatH). It is a pleasure to acknowledge support from each of these sources. Anyone interested in further details or who wishes to employ the databases should contact the first author.

REFERENCES

Chaney, C.F. (1969) Loci of disfluencies in the speech of stutterers. Journal of Speech and Hearing Research, 12, 667-668.

Clements, G.N. & Keyser, S.J. (1983) CV phonology: A generative theory of the syllable. Cambridge: MIT Press.

Crystal, D. (1969). Prosodic systems and intonation in English. London: Cambridge University Press.

Howell, P., Wingfield, T. & Johnson, M. (1988) Characteristics of the speech of stutterers during normal and altered auditory feedback. In W. A. Ainsworth & J. N. Holmes (Eds.), Proceedings Speech '88, Edinburgh: Institute of Acoustics, 3, 1069-1076.

Osberger, M.J. & Levitt, H. (1979) The effect of timing errors on the intelligibility of deaf children's speech. Journal of the Acoustical Society of America, 66, 1316-1324.

Silverman, F.H. & Williams, D.E. (1967) Loci of disfluencies in the speech of nonstutterers during oral reading. Journal of Speech and Hearing Research, 10, 790-794.

Svartvik, J. & Quirk, R. (1980). A corpus of English conversation. Lund: Gleerup.

Wingate, M. E. (1988). The structure of stuttering. New York: Springer-Verlag.