

Published in final edited form as:

Stammering Res. 2004 July 1; 1(2): 130–242.

Facilities to assist people to research into stammered speech

Peter Howell¹ and Mark Huckvale²

¹Department of Psychology, University College London, Gower St., London WC1E 6BT
p.howell@ucl.ac.uk

²Department of Phonetics and Linguistics, University College London, Gower St., London WC1E 6BT
m.huckvale@phon.ucl.ac.uk

Abstract

The purpose of this article is to indicate how access can be obtained, through *Stammering Research*, to audio recordings and transcriptions of spontaneous speech data from speakers who stammer. Selections of the first author's data are available in several formats. We describe where to obtain free software for manipulation and analysis of the data in their respective formats. Papers reporting analyses of these data are invited as submissions to this section of *Stammering Research*. It is intended that subsequent analyses that employ these data will be published in *Stammering Research* on an on-going basis. Plans are outlined to provide similar data from young speakers (ones developing fluently and ones who stammer), follow-up data from speakers who stammer, data from speakers who stammer who do not speak English and from speakers who have other speech disorders, for comparison, all through the pages of *Stammering Research*. The invitation is extended to those promulgating evidence-based practice approaches (see the *Journal of Fluency Disorders*, volume 28, number 4 which is a special issue devoted to this topic) and anyone with other interesting data related to stammering to prepare them in a form that can be made accessible to others via *Stammering Research*.

Keywords

UCL Psychology speech group database <http://www.speech.psychol.ucl.ac.uk/>; SFS <http://www.phon.ucl.ac.uk/resource/sfs/>; CHILDES <http://childes.psy.cmu.edu/>; PRAAT <http://www.praat.org/>; the Wellcome Trust <http://wellcome.ac.uk>

1. Introduction

Over the last 20 years, spontaneous speech data from speakers who stammer have been collected by the Speech Group in the Psychology Department of University College London (UCL). These data have mainly been collected through research funding from the Wellcome trust who, as a matter of policy, encourage public access to science. A proportion of these data has been transcribed. All these materials will be shared with the research community with the intention of encouraging research into stammering. In the future, we plan to make other types of data available provided by our own, and hopefully other, groups. Some background considerations about the choice of approach to make these data available are given in section 2 of this article.

Section 3 includes a description of the data we hold, and indicates which subset we are currently making available. The transcribed data are available in CHAT, PRAAT TextGrid and (in some cases) as SFS annotation items aligned against the audio records. The audio data have been prepared in WAV, SFS and MP3 formats. Undoubtedly there are other packages available too and authors or users of those systems are welcome to prepare the data

so that they can be processed with them. Users can submit articles that make the case for including other formats. If such articles are accepted after peer review, they will be invited to prepare the data in that format and it will be included in the archive.

Providing samples of data from speakers who stammer is an important step in encouraging research into stammering. However, people wishing to do research also need to have facilities to manipulate these data (for analyses of speech characteristics, to use these data in perceptual assessments and so on). The formats that are provided allow readers who are familiar with CLAN, PRAAT and SFS programs to investigate these data. Section 4 gives tutorial material and some illustrative analyses using the SFS software suite. SFS is used by the UCL speech group for the reasons also given in section 4. It is necessary to emphasize that exclusive use of SFS is *not* being advocated; CHILDES and PRAAT each have facilities that are not available in SFS (and vice versa). Thus, different software packages should be chosen according to what analyses end-users wish to perform. As indicated in the previous paragraph, some preparation has been done on the data so that each of these software packages can process the data. Future issues of *Stammering Research* will include details and demonstration of some of the capabilities of these other software suites. A general feature that commends SFS, CHILDES and PRAAT, is that they are all available for free. Details of how to access these software are included.

In addition to data and software for analyzing these data, researchers need an outlet for their findings. It is suggested that researchers consider *Stammering Research* as such an outlet. This will allow an archive of analyses to be built up on data available through the journal, provide a forum to make corrections to the data, and offer a repository that can be used to access new software that are useful for analysis of data like these.

This is an extensive document. It has been written as far as is possible so that the different sections can be read in isolation. The body of the text gives the description of what data and software are available and also reports studies using these data that we hope have some intrinsic interest to readers. The material in the Appendices will need to be consulted when a reader has some specific need. Appendix A will be used when the user wishes to select some of the data in whatever format they need. Appendix B describes the machine-readable transcription convention that UCL Psychology's Speech Group has adopted which users may choose to use or convert to one they are more familiar with. Appendices C and D give worked exercises using SFS utilities in two important areas for manipulation of these data (transcription and formant analysis). Appendix E gives some details of applications of a Hidden Markov model software suite to recognition of dysfluencies. The background knowledge that is assumed is given at the start of each Appendix. These appendices were, in part, written as tutorial material for this article though they also serve as important sources for general users of SFS. Also, applications in the body of this paper draw on the information provided in these appendices. For all these reasons, it is appropriate that they be included in this document. It should be apparent that these facilities are only part of what SFS offers. There is extensive other software in and various options for people wishing to write their own scripts (in MATLAB, C, and in speech measurement language, SML which is a high level scripting language for manipulating information in SFS files, for details of SML see section 1.5 of the SFS manual at <http://www.phon.ucl.ac.uk/resource/sfs/>). The CHILDES and PRAAT systems have various similar options that, it is hoped, will be described in future articles.

2. Background to release of data on speakers who stammer and software for its analysis

Speech data are expensive to collect, speech is time-consuming to analyze, the range of analyses that one group can do is limited (e.g. because of time constraints). Individuals who want to start a research program into stammering often have the daunting tasks of obtaining expertise, equipment, software and an administrative structure (to locate participants, obtain ethical permission etc.) before they can conduct their research. If they are interested (like us) in developmental issues associated with stammering, they may have to collect longitudinal data for several years.

There are pitfalls when attempting to make recordings. Clinics and schools are not ideal recording environments. (Many recordings that speech therapists have offered us in the past have been too poor in quality for analysis.) Special skills need to be acquired for eliciting speech from young speakers and also some children who stammer.

UCL Psychology Department's speech group has extensive audio data on speakers who stammer of appropriate quality for linguistic and acoustic analysis, which it is prepared to supply to the wider community thereby circumventing these problems in 'getting started' in research into stammering. Generally speaking, three issues need to be addressed so the data can be used. These are 1) conventions for data preparation need to be specified, 2) information needs to be supplied about how to use available software to manipulate data in these formats, and 3) information has to be provided about where and how other similarly formatted data can be deposited or accessed.

The way in which three different systems address these issues is discussed. The three systems selected for consideration are MacWhinney's CHILDES system, Boersma's PRAAT system and Huckvale's SFS system. Each of these systems was developed for different purposes and each has advantages that make it more useful for some purposes than are the others. This is implicitly acknowledged as, for instance, there is provision in the CHILDES system to access PRAAT software to take advantage of the facilities for speech analysis provided by PRAAT. The components of each system and the purpose for which they were developed are briefly indicated.

CHILDES

The CHILDES project was developed specifically for research into child language. It is masterminded by Brian MacWhinney and has three separate components that address the above three issues: CHAT (Codes for the Human Analysis of Transcripts) provides the data conventions, CLAN (Computerized Language ANalysis) is the software package and CHILDES (CHIld Language Data Exchange System) is the repository for the data. The three components together are referred to as the CHILDES project (MacWhinney, 1995). CHILDES was developed for higher level linguistic analysis. This makes sense in terms of the target populations for which the system was developed. Lengthy recording sessions are often necessary to obtain even limited samples of speech from very young children. Consequently, it would not make sense to archive the entire audio recordings in these cases (a lot of the data so stored would be taken up by interlocutors and sections where the child says nothing). The system has been developed so that audio data can be logged and there are links to audio analysis software (e.g. PRAAT). This is particularly useful when samples from older speakers are employed. Most of the data that are available through CHILDES are from fluent speakers. Further details can be found on the CHILDES home page at <http://childes.psy.cmu.edu>.

PRAAT

PRAAT was developed by Paul Boersma and is specifically an audio data analysis tool and PRAAT has the advantage that it is simple to use. It reads all kinds of standard audio format files including WAV. It runs on Unix, Mac and Windows platforms. Transcription data are in the form of a TextGrid which can easily be created within the PRAAT system or imported via a suitably formatted text file. The software incorporates sub-tools for neural net analysis, speech synthesis and some statistical functions, and a scripting facility is available which allows users to write specialized functions. There is a PRAAT user group who share data (as well as software). PRAAT is close to SFS in the facilities it provides, but each of them have specialized processing software. For example, PRAAT does sophisticated analyses of harmonicity while SFS incorporates software for dealing with ancillary signals provided from a laryngograph. The PRAAT Home Page is at <http://www.praat.org> and the manual is accessed within the software.

SFS

SFS stands for Speech Filing System. It provides an integrated method of dealing with different sources of information about speech sounds. The raw audio record is at the core of the system. There are options so that a number of other analyses (manual or computational) on the same speech data can be displayed for inspection. Transcriptions can be manually entered in any format (Appendix B gives the Joint Speech Research Unit format, JSRU, that the UCL Speech Group uses). The filing system provides the system that integrates analyses from these several sources for visual or statistical inspection. The integration of these sources of information is the attraction, though there are utilities that allow the audio recordings to be uploaded or dumped in WAV or other standard formats and, similarly, TXT files can be dumped or uploaded.

For the Speech Group's work, the SFS facility that allows, *inter alia*, audio data and aligned transcriptions to be concurrently displayed has been particularly useful in the development of Howell's EXPLAN theory of spontaneous speech control (Howell, 2002, 2004; Howell & Au-Yeung, 2002). This theory maintains that motor execution of one segment takes place concurrent with the planning of the following segment and that fluency may break down when execution time on one segment does not allow sufficient time to complete planning of the following segment. SFS displays of the audio waveform and the associated transcription provide the information necessary for evaluation of predictions of this theory. Thus, the audio item provides an indication of the time required for execution of the current segment, the annotation item indicates the structure of the following segment that can be used to ascertain how complex the word is to plan (Dworzynski & Howell, 2004; Howell & Au-Yeung, 1995a; Howell, Au-Yeung & Sackin, 2000). These two factors can be examined jointly to determine whether they lead to fluency breakdown.

The software provided by SFS includes many of the same facilities as PRAAT. SFS can display selected analyses of the same stretch of speech aligned in time. Appendix A of this article gives details of how to access an extensive SFS database (i.e. the data on speakers who stammer). SFS was developed on a Unix system, and certain advanced software features require use of the Unix command line interpreter (such as those in Appendix E). The SFS Home Page can be found at <http://www.phon.ucl.ac.uk/resource/sfs/>

Although our work employs SFS extensively, other users may prefer to work with one of these other systems if they have specific needs of the facilities provided. Basic versions of the files have been supplied that allow users of these other systems to get started. For CHILDES users, files have been converted according to TEXTIN (MacWhinney, 1995, p. 158), and links to WAV files provided. Other information that could be added to the CHAT

files (speaker, age etc.) are available in the Access file described in Appendix A. The files have also been prepared as PRAAT TextGrids (and WAV files) at Paul Boersma's request. He intends to inform his email list of PRAAT users about the possibility of analysing and reporting results on these data. It is also recognised that some users may just need to listen to the files or read the transcriptions. The transcriptions can be examined with current word processing packages. Users who do not intend to carry out acoustic analyses probably do not want high-fidelity WAV files as these are cumbersome to access. For this reason, MP3 files have been made available which are much shorter files. Though there is some data loss, the audio files are still of good quality. Most of the remainder of this article refers to the SFS versions of the files and associated analysis software.

Access to UCL Psychology Department Speech Group's data files—This article is intended to set the ball rolling to stimulate research in this area. Part of UCL's data on speakers who stammer will be made available. We are not just going to supply the audio data, but also, where available, orthographic and phonetic transcriptions. Some of the latter are also aligned against the audio records.

Appendix A indicates how the data in the various formats can be accessed. The data will also be distributed in an alternative medium that some users may find more convenient. CDs have been made of different subsets of the data set (described in Appendix A). The listening center at UCL's Phonetics and Linguistics Department holds copies of these CDs (the data cannot be customized for individual clients so if you need selections of data appearing on two or more CDs, you will have to purchase the CDs concerned). CDs can be purchased for a cost of around £10 each (including p&p) for those who prefer their data in this format. To comply with the European Union's data protection act, we have ensured that speakers cannot be identified from the recordings.

Transcription of spontaneous speech is a difficult task and it is unlikely to lead to 100% agreement between transcribers, and the transcription procedure has been designed to be more detailed at some points in utterances (around dysfluencies) than others (the rest of the speech). The reliability estimates that have been made for a selection of the transcriptions indicate that there is satisfactory agreement between trained transcribers (Howell, Au-Yeung & Sackin, 1999, 2000). However, this is the first time these transcriptions have been open to public scrutiny and although we have been rigorous in preparing the transcriptions, there will inevitably be some errors. Users should notify by email to psychol-stammer@ucl.ac.uk any errors they locate so that these can be corrected in subsequent release versions of the data. This 'public' correction procedure is preferred to one where audio files are not available, so errors are not visible. In addition, the recordings where there are audio files alone can be used by anyone who wishes to start from scratch (using the current or any other transcription scheme). We consider it imperative that new and improved transcriptions should be made available for scrutiny in the same manner as those we have provided.

Access to UCL's Phonetics Department's SFS software—SFS can be obtained from <http://www.phon.ucl.ac.uk/> under Research Resources.

We do not have the resources to offer software support. We believe that providing these facilities offers a forum for scientific collaboration and exchange of ideas, which is the ethos behind *Stammering Research*. Copyright to the data is held by Howell and copyright to the software is held by Huckvale. The data and software are freely available to anyone for research and teaching purposes. If the data and/or software are used in publications, these etc., users have to a) notify Howell (p.howell@ucl.ac.uk), b) acknowledge the source in any publication by referencing this article, c) include an acknowledgement that data collection was supported by the Wellcome Trust.

Outlet—It is intended that publications reporting analyses of these data will appear in *Stammering Research*. The prospect of extending research and assisting beginners to get started in research was made possible with the advent of Internet publications. In the main, the Internet has failed to deliver these possibilities to date because, where e-journals have appeared, they have usually been electronic versions of the printed journals that were available previously and have not provided access to data sources. *Stammering Research* welcomes submissions of articles for consideration that report analyses of these data, comparison between these data and previously published findings and so on. Submissions are invited at any time. There are no restrictions about what these analyses can be nor who may submit their work: Acoustic, articulatory, phonetic, phonological, prosodic and syntactic analyses would all be appropriate. The reports could also cover type/token, qualitative, transactional analyses. As implied, they can be compared with fluent speech or with samples of speech from people with other disorders or just analyses reporting on characteristics of stammered speech. Authors should be prepared to return analyses and scripts back to the archive (email submissions to psychol-stammer@ucl.ac.uk).

The SFS system can also be used for preparing material for perceptual tests. For instance, the software and the data could be used to replicate the classic Kully and Boberg (1988) study that showed that interclinic agreement in the identification of fluent and stammered syllables was poor. They can also be used to check some of the claims made by Cordes-Bothe and Ingham in support of time interval analysis as a means of assessing stammered speech. In this connection, it would be particularly useful to have TI sections of these freely-available materials judged by some members of these authors' expert panel, as the judgments of this panel have previously been used as benchmarks for other data about which intervals are, and are not, stammered.

It is hoped that these data will be of some lasting value to the research community (in the areas of stammering research and speech in general). In order to gauge whether there is a call for a facility like this, we have prepared a limited selection of our data set at present (more will follow if this proves popular). As stated above, a section of *Stammering Research* has been set up which is devoted to analyses that include (though is not necessarily restricted to) these data.

3. Description of the data

A complete description of the UCL archive of data from speakers who stammer is given and then, the subset in the initial release is described. The complete data set currently includes 249 speakers who are categorized into five classes depending on the range of ages over which they have been recorded. There is also a holding class for young children we are still seeing but cannot project how long they will be available for recording. The data within each of these classes have undergone different amounts of preparation levels. This document describes the version one release of data from the first class where recordings are only available over a limited age range and where there is no possibility of obtaining more recordings. Samples are a) available as audio alone (as SFS files), b) with orthographic transcriptions (separate TXT file), c) with phonetic transcriptions (again separate TXT files) and d) where phonetic transcriptions are aligned against audio waveforms (available as SFS files). (CHILDES and PRAAT versions of the files are also available.) The alignment step under d) has a final check at the point where the transcriptions are aligned against the audio waveforms so these represent our highest level of data preparation. A full description of all classes of data we hold and current level of preparation is given below. The procedure for revising data in later revisions (corrections and generally useful ancillary analyses) is to send these in (as indicated in section 2). Depending on demand, other data classes will be released in phases as work is completed.

We record all speakers who stammer who volunteer. For our current project work, we are particularly interested in speakers in the age range eight to teenage. Pre-teen speakers who stammer have a good chance of recovery. Consequently, we wish to follow up children who stammer and controls over this period, examine their speech and see whether different paths of fluency development are followed by those children who persist and those who recover. Ideally we want a minimum of three samples in this period (one in the age range 8-10 years, one between 10 and 12 years and one at teenage). We have complete sets of such recordings for 24 of our speakers.

Class 1, includes speakers who were either a) older than the maximum age of our target group when they were first seen (i.e. only seen after they have reached teenage), b) speakers who are in the target age ranges but who were only available at one target age because they live too far away from the laboratory or c) speakers who were in the required age range but with whom we have lost contact (most often because they have moved home and have not notified us of their new address and telephone number). Permission for data release cannot be obtained for speakers in class c). c) represents attrition of the sample, though there is no reason to suppose that this affects those speakers who persist differentially relative to those who recover. Class 2 consists of speakers who have not reached teenage who have been recorded at all target ages they have passed through that we are continuing to see (this class includes children who are under 8 years). Class 3 and 4 have not provided data at one of the first two target ages but attended at the third target age (i.e. they have reached teenage). For class 3, recordings are available at 8 and teenage, but not age 10-12 (often because the recording sessions clashed with school or family obligations and could not be rescheduled). For class 4, we have recordings for 10-12 and teenage. The lack of recordings at age 8-10 reflects the fact that these children were not seen at clinic until they were aged 10+. Class 5 are participants for whom we have at least three recordings at the designated ages, and we have continued to see most of these beyond the stipulated upper age. Many have supplied other forms of data. Appendix A describes an ACCESS file that is included in the data directory, which gives demographic information about the speakers.

Table 1 gives an indication of what is available and in what form. Not all data can be released at present (we have ethics permission, but are still waiting written consent by individuals or by clinics). Also, there are some data where the audio quality is not good enough for release. The numerator in each cell indicates what is being released and the denominator the total available. Thus 41/158 in the participants column indicates that data from 41 participants out of a total of 158 are being released.

4. Some uses for the data including illustrations of applications of SFS tools and concepts for assessing stammered speech

Data similar to those made available in Appendix A have been used to investigate a range of questions about stammering from many different perspectives. A comprehensive list of all studies conducted is beyond the scope of this article. Studies conducted by the UCL group range from acoustic analysis of articulatory features associated with stammering, to pragmatic analyses of speakers who stammer in conversation with others. At the acoustic level, the UCL group has examined how the phonation source operates in people who stammer (Howell, 1995; Howell & Williams, 1988, 1992; Howell & Young, 1990), whether the vowel in a series of repetitions is neutralized using formant frequency analysis methods similar to those described in Appendix D (Howell & Vause, 1986) and speech rate has been measured from digitized oscillograms (Howell, Au-Yeung & Pilgrim, 1999; Howell & Sackin, 2000). PRAAT offers acoustic analysis software that could extend our understanding of what happens to the voice when fluency breaks down. Phonetic and phonological analyses have been performed to assess whether these factors are implicated in

stammering (Dworzynski & Howell, 2004; Dworzynski, Howell & Natke, 2003; Howell & Au-Yeung, 1995a; Howell, Au-Yeung & Sackin, 2000). The change in pattern of stammering over development has been examined within prosodically-defined units in a variety of languages (Au-Yeung, Vallejo Gomez & Howell, 2003; Dworzynski, Howell, Au-Yeung & Rommel, 2004; Howell, Au-Yeung & Sackin, 1999) and different ways of defining these units (based on lexical or metrical properties) have been investigated for Spanish (Howell, in press). Various forms of syntactic analysis have been performed to establish whether syntactically complex utterances are more likely to be stammered than simpler ones (Howell & Au-Yeung, 1995b; Kadi-Hanifi & Howell, 1992). A pragmatic factor that has been examined is whether the speech of the interlocutor affects the speech of the person who stammers (Howell, Kapoor & Rustin, 1997). CHILDES offers a variety of techniques that extend the possibilities of examining other high order effects on stammering (including pragmatic ones). There is considerable scope for further phonetic, phonological, prosodic, syntactic and pragmatic analysis of these data and some suggestions follow.

Suggested studies

Perception of stutterings—The data that are supplied can be used for assessing the effect of different perceptual procedures on stammering assessment, for training therapists/pathologists on stammering assessments, for showing how heterogeneous stammering patterns can be within and across age groups. The materials could also be used to replicate the classic, but somewhat dated, Kully and Boberg (1988) study that showed judgements about the same sample of speech is judged differently by different clinics.

Studies on speech control in speakers who stutter—Some basic familiarity with acoustic phonetics is assumed to understand this section (for those requiring a refresher, see Ladefoged, 1975). The stammered sequence “kuh, kuh, Katy” contains a different sounding vowel (“uh” or as it is known more precisely “schwa”). Van Riper argued that a speaker who produces such a sequence had selected the wrong vowel at the start of this sequence and detected this by listening the sound of his or her voice (called feedback monitoring). As the speaker cannot produce “Katy” when the incorrect (“schwa”) vowel has been inserted, the speaker interrupts speech and tries again. Howell and Vause (1986) argued that the vowel in a sequence of repetitions might sound like schwa because the vowels are short and low in amplitude (by analogy with vowel reduction that occurs in rapidly spoken, or casual, speech where the vowels also sound like schwa even when some other vowel is intended). They tested this hypothesis by acoustic analyses that compared the vowels in a sequence of repetitions with the vowel after fluent release (Howell and Vause also conducted perceptual tests, see the preceding section, which are not discussed here). They found that the formants of the vowels in sequences of repetitions and after fluent release occurred at the appropriate frequencies (suggesting that the vowel in the sequence of repetitions had been correctly articulated). Thus, they concluded that van Riper’s feedback monitoring account of part word repetitions was not correct. The recordings of speaker 210 (at age 11 years 3 months) can be used to check Howell and Vause’s finding. At 20.4s, the speaker appears to say “guh-go”. The values of F1, F2 and F3 are similar in the “uh” and “o” sections indicating the vowels are similar (as Howell & Vause, 1986 reported).

Van Riper’s argument could be applied to consonants that are prolonged. Prolonged /s/. sound canonically like /s/. However there are different forms of /s/ that sound different which depend on what vowel follows. So, for example, an /s/ before an /i/ vowel sounds clearly different to an /s/ before an /u/ vowel. A possible explanation of /s/-prolongation could be that the wrong form of /s/ was selected and produced and when the speaker detected this, the transition to the following vowel could not be made leading to the speaker prolonging the /s/. This can be tested by seeing whether the /s/ in words that have an /i/

following is acoustically identical when prolonged compared with when it is spoken fluently (in the same way this was done when comparing the vowels in a sequence of part word repetitions to the intended vowel at fluent release in the previous study). Speaker 61 at age 14 years 8 months produced two prolonged /s/s before an /i/ vowel - one at the beginning of the word “CD” at 47 s and one at the beginning of “CCF” at 115 s. Though there are no fluent /s/s before /i/ in this recording, a recording was made a month later (14 years 9 months) in which the speaker says ‘CD’ twice (both times fluently) (these appear at 112.2 and 116 s in this file). Oscillograms, spectrograms and cross sections were taken of the fluent and dysfluent /s/s. The main feature is that the spectra peak at around 5kHz and this applies to fluent and dysfluent forms of /s/ before /i/. Based on the acoustic similarity and informal listening to these examples, speakers appears to be articulating the form of /s/ that would permit coarticulation with the intended vowel that follows. Thus, an account of prolongation based on selection of the inappropriate allophone of /s/ does not seem correct.

So far only continuant phones (vowels and fricatives) have been discussed. These are produced with articulatory positions that do not change over time. It is possible that speakers who stammer have problems in controlling speech timing which would be reflected in phones that require changes in articulation over the time of their production. Plosive stop consonants are one class of sounds where such timing problems might be manifest. Plosives start with a short period of broad band energy that marks sound onset (the burst). The plosives can be divided into voiced (/b, d, g/) and voiceless (/p, t, k/) forms where corresponding pairs (e.g. /b/and /p/) have the same place of articulation. The differences in voicing arise because speakers control the timing of articulatory gestures in distinct ways for these two classes of plosives. After burst onset, vocal fold vibration starts almost immediately for voiced plosive (voicing gives rise to the pitch epoch markers mentioned in the pitch synchronous analysis part of section 3 of Appendix D which appear as striations in broad band spectrograms). In contrast, voiceless plosives have a period, after the burst, during which the phone is aspirated before voicing starts. The time between burst onset and onset of voicing can be used as a simple measure of voice onset time (VOT) that characterizes the difference between voiced (short VOT) and voiceless (long VOT) plosives. If a speaker who stammers has problems initiating voicing in time, this would be reflected in longer VOTs for /d/s than /t/s and make /d/s sound something like /t/s.

Some speakers who stammer appear to have problems initiating voicing so this should be reflected in the VOT measure. Speaker 1100 shows this characteristic. For instance, 367 s into his audio file he shows multi part word syllable repetitions on the word ‘David’ (prior to ‘Hockney’) which sound devoiced (i.e. the /d/s sounds like /t/s). A /d/ realized as /t/ should have a longer VOT (close to a voiceless plosives) than that of a true /d/. Acoustic analysis supports this notion, as you will see if you measure the VOT of the /d/s in the part word repetitions of the attempts at ‘David’ (prior to ‘Hockney’) and compare them with the VOT of /d/ in a fluent word (e.g. the “don’t” that occurs at 80.7 s). You should also listen to the voiced sounds to confirm that they appear to be devoiced (i.e. the /d/ sounds like /t/).

Performance of HMM automatic dysfluency recognizer—Like fluent speech corpora, the data may also provide material for training automatic speech recognizers (Howell, Hamilton & Kyriacopoulos, 1986; Howell, Sackin & Glenn, 1997a, b; Noth, Niemann, Haderlein, Decher, Eysholdt, Rosanowski, & Wittenberg, submitted). Some hidden Markov model (HMM) utilities that have been used to construct a dysfluency recognizer are described in Appendix E. Performance of this recognizer on 5s intervals that experts agreed about for six test samples (0030_17y9m.1, 0061_14y8m.1, 0078_16y5m.1, 0095_7y7m.1, 0098_10y6m.1, 0138_13y3m.1, 0210_11y3m.1, 0234_9y9m.1) correctly classified 60% of intervals as stuttered/fluently. Though above chance, this is not particularly impressive. It does set a benchmark against which better dysfluency recognizers can be

developed and assessed. One obvious improvement would be to be more selective when training the phone models (in the benchmark version, these were obtained from fluent speakers, not speakers who stammer). In addition to providing training material specifically for automating dysfluency counts, these data could be used more generally to test the robustness of recognizers developed for fluent speech. As it is claimed that recognizers perform at high levels when material is fluent, to what extent do failures in these algorithms coincide with stuttered dysfluencies? These are just some of the topics that the group at UCL are addressing and doubtless there are numerous other topics that the data and software will be helpful in investigating.

The JSRU coding scheme is presented in Appendix B. Modifications for the application of this scheme for dealing with stammered speech and codes for properties that are important for studying stammering are given in section 6 of Appendix C. The material that has been prepared according to this scheme can be used for analyses similar to those described in the studies cited above. A tutorial on manipulating transcriptions in SFS (with a focus on aligning them against the audio waveforms) is given in Appendix C. As stated earlier, SFS may prove particularly useful when researchers want to visualize how disparate sources of information (e.g. duration and characterizations of phonetic difficulty) on different segmental units interact and lead to dysfluency. The aligned display convention should also prove useful when comparing the results of different analysis methods (such as those used for syntactic characterization of these materials) applied to the same stretch of data. A second SFS tutorial (Appendix D) covers basic aspects to do with acoustic analysis of speech that have been employed in some of the studies cited at the start of this section. Appendix E describes tools that could be used for developing HMM recognizers for dysfluent speech.

Appendices

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Wellcome Trust.

References

- Au-Yeung J, Vallejo Gomez I, Howell P. Exchange of disfluency from function words to content words with age in Spanish speakers who stutter. *Journal of Speech, Language and Hearing Research*. 2003; 46:754–765. [PubMed: 14697001]
- Dworzynski K, Howell P. Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders*. 2004; 29:149–173. [PubMed: 15178130]
- Dworzynski K, Howell P, Au-Yeung J, Rommel D. Stuttering on function and content words across age groups of German speakers who stutter. *Journal of Multilingual Communication Disorders*. 2004; 2:81–101. [PubMed: 18270544]
- Dworzynski K, Howell P, Natke U. Predicting stuttering from linguistic factors for German speakers in two age groups. *Journal of Fluency Disorders*. 2003; 28:95–113. [PubMed: 12809747]
- Howell, P. The acoustic properties of stuttered speech. In: Starkweather, CW.; Peters, HFM., editors. *Proceedings of the First World Congress on Fluency Disorders*; Nijmegen: Nijmegen University Press; 1995. p. 48-50.
- Howell, P. The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures. In: Fava, E., editor. *Pathology and therapy of speech disorders*. Amsterdam: John Benjamins; 2002. p. 95-118.

- Howell P. Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders*. 2004; 39:122–139. [PubMed: 18259590]
- Howell P. Comparison of two ways of defining phonological words for assessing stuttering pattern changes with age in Spanish speakers who stutter. *Journal of Multilingual Communication Disorders*. in press.
- Howell P, Au-Yeung J. The association between stuttering, Brown's factors and phonological categories in child stutterers ranging in age between 2 and 12 years. *Journal of Fluency Disorders*. 1995a; 20:331–344.
- Howell P, Au-Yeung J. Syntactic determinants of stuttering in the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*. 1995b; 20:317–330.
- Howell, P.; Au-Yeung, J. The EXPLAN theory of fluency control and the diagnosis of stuttering. In: Fava, E., editor. *Pathology and therapy of speech disorders*. Amsterdam: John Benjamins; 2002. p. 75-94.
- Howell P, Au-Yeung J, Pilgrim L. Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter. *Journal of the Acoustical Society of America*. 1999; 105:481–490. [PubMed: 9921672]
- Howell P, Au-Yeung J, Sackin S. Exchange of stuttering from function words to content words with age. *Journal of Speech, Language and Hearing Research*. 1999; 42:345–354. [PubMed: 10229451]
- Howell P, Au-Yeung J, Sackin S. Internal structure of content words leading to lifespan differences in phonological difficulty in stuttering. *Journal of Fluency Disorders*. 2000; 25:1–20. [PubMed: 18259599]
- Howell, P.; Hamilton, A.; Kyriacopoulos, A. *Speech Input/Output: Techniques and Applications*. London: IEE Publications; 1986. Automatic detection of repetitions and prolongations in stuttered speech; p. 252-256.
- Howell, P.; Kapoor, A.; Rustin, L. The effects of formal and casual interview styles on stuttering incidence. In: Hulstijn, W.; Peters, HFM.; van Lieshout, PHHM., editors. *Speech Production: Motor Control, Brain Research and Fluency Disorders*. Amsterdam: Elsevier; 1997. p. 515-520.
- Howell P, Sackin S. Speech rate manipulation and its effects on fluency reversal in children who stutter. *Journal of Developmental and Physical Disabilities*. 2000; 12:291–315. [PubMed: 18259598]
- Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers. *Journal of Speech, Language and Hearing Research*. 1997a; 40:1073–1084. [PubMed: 9328878]
- Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. *Journal of Speech, Language and Hearing Research*. 1997b; 40:1085–1096. [PubMed: 9328879]
- Howell P, Vause L. Acoustic analysis and perception of vowels in stuttered speech. *Journal of the Acoustical Society of America*. 1986; 79:1571–1579. [PubMed: 3711457]
- Howell P, Williams M. The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. *Journal of the Acoustical Society of America*. 1988; 84:80–89. [PubMed: 3411057]
- Howell P, Williams M. Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. *Journal of the Acoustical Society of America*. 1992; 91:1697–1706. [PubMed: 1564205]
- Howell P, Young K. Analysis of periodic and aperiodic components during fluent and dysfluent phases of child and adult stutterers' speech. *Phonetica*. 1990; 47:238–243. [PubMed: 2130382]
- Kadi-Hanifi K, Howell P. Syntactic analysis of the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*. 1992; 17:151–170.
- Kully D, Boberg E. An investigation of interclinic agreement in the identification of fluent and stuttered syllables. *Journal of Fluency Disorders*. 1988; 13:309–318.
- Ladefoged, P. *A course in Phonetics*. New York: Harcourt, Brace, Jovanovich; 1975.

- MacWhinney, B. The CHILDES project. Hillsdale NJ: Lawrence Erlbaum; 1995.
- Noth, E.; Niemann, H.; Haderlein, T.; Decher, M.; Eysholdt, U.; Rosanowski, F.; Wittenberg, T.
Automatic stuttering recognition using Hidden Markov models. submitted
- Rosen, S.; Howell, P. Signals and Systems for Speech and Hearing. London and San Diego: Academic Press; 1991.

Summary of recordings available on UCL Department of Psychology Speech Group's data archive. Recordings are indicated according to class (see text) and type of file available

Table 1

	No of participants	No of files	No where orth avail	No where phon avail	No where phon aligned against audio
Class 1 -release 1	41/158	95/426	19/42	17/34	11/11
Class 2 - speakers who have not reached teenage, who have been recorded at all target ages they have passed through that we are continuing to see plus children who are < 8 years we are following up	7/ 21	13/57	5/7	1/1	2/2
Class 3 - available at 8-10yrs and 12yrs+	1/5	3/21	0/14	0/13	0/0
Class 4 - available at 10-12yrs and 12yrs+	7/41	20/184	6/57	6/54	0/1
Class 5 - recordings at the 3 target ages	5/24	7/142	1/63	0/44	3/69
Totals	249 61	830 138	183 31	147 24	82 16