Basic Notes on Audio Recording for Phonetics

Dafydd Gibbon, 03, 2017-02-14

Contents

1 General considerations

There are many things to consider when recording speech for phonetic and linguistic analysis, not just the recorder. The most important question concerns the criteria with which you intend to analyse the recording: phonemes, pitch patterns, duration, linguistic concepts such as stress and accent. I will list the most important practical points here. Note: these notes are informal, selective, and based on my own experience. Others will have variant views on a number of points.

There are many handbooks which give much more detailed information on different practical aspects of phonetics:

Gibbon, Dafydd, Roger Moore and Richard Winski, eds. 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter (2 editions: single volume and four volumes).

Hardcastle, William J., John Laver and Fiona E. Gibbon, eds. 2010. *Handbook of Phonetic Sciences, 2nd Edition*. London: Wiley-Blackwells,

International Phonetic Assoc., 1999. *Handbook of the International Phonetic Association*. Cambridge: University Press.

2 Pre-recording phase

First, decide which acoustic features of the speech signal are important (e.g. pitch, amplitude, formants). For example, there are different linguistic and phonetic definitions of 'stress'. Stress has been described as phonological, as a factor in speech production, as a prominently perceived item in an utterance, with some phonetic correlates (e.g. pitch pattern, syllable duration, rhythm). In general, stress has to be seen as a top-down, cognitively constructed feature of production and perception, not as a simple phonetic phenomenon.

Some aspects of recording a speechsignal are relevant for all purposes and include many of the following:

1. Speech styles to be recorded, for example:
    1. prompted ‘laboratory’ speech for phonetic experiments;
    2. task-oriented speech;
    3. dialogue (interview, conversation, etc.) - note that dialogues for phonetic analysis require stereo recording.
2. Technical choices:
    1. General advice: Garbage in – garbage out! Go for the highest quality devices and formats possible – realistically, compromises are usually needed for financial reasons – and remember that the data, like other resources such as tools, should conform to standard formats and specifications to ensure that the data will be reusable (i.e. available for other than the original use and for yourself and other researchers) and interoperable (i.e. usable with other devices and software).
    2. Microphone (to capture the range of frequencies required): many good brands are available e.g. Sennheiser, Shure, Røde; a high quality headset microphone may be

preferable for some purposes. A Lavalier microphone for attaching to clothing may have somewhat lower audio quality.

3. Choice of audio file format (e.g. unc ompressed WAV for full quality recording, vs. MP3 or WMA compressed formats for lower quality but smaller files); however, in general uncompressed WAV format is preferable for phonetic analysis with Praat.
4. Choice of recorder with choice of WAV and MP3 formats (e.g. Zoom H2N Handy Recorder, the most popular in general, or Roland R-05, popular among musicians); a recorder with only MP3 or WMA formats may be suitable for some purposes, and for the analysis of stress it may be suitable if you are looking onlyat syllable duration and fundamental frequency, or if neither you nor anyone else will use the data again (but remember that in general data are time-consuming and expensive to collect and therefore should be reusable).
5. Choice of recorder setting:
    1. Input gain (microphone gain, input volume):
        1. no Automatic Gain Control (AGC): if you use it, you will not be able to use measurements of amplitude or intensity since AGC changes these automatically to keep the column constant (also resulting in changes of noise level),
        2. set the colume control as high as possible, but avoiding peak overdrive (i.e. the peaks should be just below the maximum on the volume meter).
    2. Sampling frequency:
        1. sampling frequency of at least 2*f*, where *f* is the highest frequency to be recorded (Nyquist sampling theorem);
        2. high quality: human hearing reaches about 20kHz, so high quality recordings must use at least 40kHz sampling frequency (e.g. the 48kHz DAT standard or the 44.1kHz CD standard);
        3. medium quality: 40kHz is not strictly necessary for most phonetic purposes, as the relevant frequencies (fundamental frequency below about 600Hz (for child speech, vowel formant frequencies below 6kHz) are generally below 10kHz, so 20kHz is sufficient, and in practice 16kHz and 22.05kHz are often used to keep file sizes down;
        4. oversampling: oversampling uses a sampling rate which is much higher than the Nyquist frequency in order to avoid various noise effects; for phonetics, the best compromise frequency is a Nyquist frequency of 40kHz, which is achieved with the standard sampling frequencies of 48kHz or 44.1kHz;
        5. sampling trivia: in case you are curious about why the strange-looking number 44.1kHz was chosen as a standard: it is the product of the squares of the first 4 prime numbers above 1, which permit efficient ‘down-sampling’ to 16 different lower frequencies, basically by dividing by combinations of these numbers:

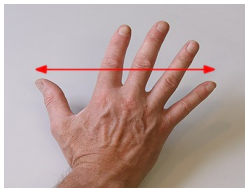
            $2^2 \times 3^2 \times 5^2 \times 7^2 = 44100\text{Hz}$

3 Recording phase

There are several points to be considered during the recording phase, in particular the scenario and environment, which will depend on the speech style being recorded:

1. Scenario environment for recording:
    1. Avoid echo as far as possible:
        1. echo is caused by hard walls and floors:
        2. so an environment with soft furnishings (curtains, carpets, cushions) is preferable, if recording is done outside a studio;
        3. and a studio should have sound-proofed floor, walls and ceiling.
    2. Avoid noise:

1. Place the microphone at least 25cm (10”) from the speaker in order to avoid breath noise, preferably slightly to the left or the right of the speaker,
2. A useful reference measure for the minimum distance is the *span*, that is, the distance between outstretched thumb and little finger. The span is usually between about 18cm and 22cm for adults, and only a little less than 25cm :



http://www.cpsisc-elearning.com.au/learningobjects/measurement/content/01_measurment/02_page.htm

3. When recording outside use a sheltered place to avoid wind and other noises; note that a wind muff (wind shield, ‘dead cat’) may filter out high frequencies, though this may not be important for some purposes (e.g. for news reporting).

2. Speaker:
    1. The speaker should be given instructions which are appropriate for the task.
    2. The speaker should be asked to give permission for the recording to be used for scientific purposes. The permission can be in writing, but in any case should be included in the recording as recorded metadata.
    3. The recording should include other metadata, including the date, the place, the speaker(s), other participants such as audience and those making the recording.
    4. Speakers should take a sip of water every 5-10 minutes to avoid drying out the vocal folds and thereby changing the voice quality.
3. Protocol:

    A protocol of the recording session should be kept, containing:
    1. the same metadata information as on the recording (time, place, participants);
    2. instructions given to speaker(s);
    3. permissions of speaker(s);
    4. file names and any other aspects of the recordings;
    5. a unique identifier indicating where the resource can be found;
    6. a list of any problems which occurred.

4 Post-recording phase

There are many kinds of post-recording activity, following levels of abstraction from data management:

1. Data management:
    1. systematic filing of metadata in a database (or temporarily in a word processor table),
    2. systematic file labelling of recorded audio files with a project (or language, etc.) name, serial number, and date, e.g.: ‘englinterview_05_2016-05-09.wav’.
    3. sutting of recordings: usually necessary in order to systematise items for analysis – but always keep the original recordings (cutting can be done using the general audio tool Audacity or the phonetic workbench Praat):

        http://www.audacityteam.org/

        http://www.fon.hum.uva.nl/praat/
2. Transcription:
    1. Transcription types vary according to needs (e.g. IPA or SAMPA phonetic vs. orthographic vs. discourse analytic transcription), labelled to match the recording, e.g.:

        englinterview_05_2016-05-09.txt

        englinterview_05_2016-05-09.doc

        englinterview_05_2016-05-09.odt

2. Text format: it is always a good idea to keep a plain text (ASCII, e.g. SAMPA) version of the transcription where possible (this may not be possible for some orthographies), and in any case to store a printed hard copy of the text.
3. For publication, transcriptions should use a Unicode font, not an *ad hoc* special purpose font.
4. For easy input and computational processing, the SAMPA alphabet is convenient. There are tools on the internet for converting SAMPA transcriptions to IPA Unicode.

3. Annotation:
    1. For all kinds of phonetic and discourse analysis it is necessary to annotate the recording using specialised software such as Praat (the most popular and in general the most useful), Annotation Pro, Transcriber or WaveSurfer, or ELAN (for video, not so good for phonetic purposes).
    2. The annotation file names should be matched to the recording files names. For example (for Praat annotations):

        englinterview_05_2016-05-09.wav

        englinterview_05_2016-05-09.TextGrid
    3. A selection of useful annotation software:
        1. Praat: the *de* facto standard tool for phonetic annotation and experimentation; Praat a very versatile 'phonetic workbench' with high quality measurements of acoustic parameters, graphics for publications, speech synthesis experimentation, scripting (for automatising analysis) with many scripts and other tools available on the internet, for Windows, Mac, Linux:

            http://www.fon.hum.uva.nl/praat/

            Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.
        2. Annotation Pro: with many of the features of Praat, compatible format, with facilities for perception experiments, Windows only:

            http://annotationpro.org/
        3. SPPAS: for large quantities of speech to be annotated it is worth considering automatic annotation (with correction by manual post-editing), Windows, Mac, Linux, implemeted in Python:

            http://www.sppas.org/

4. Phonetic and linguistic analysis of annotations:
    1. A wide selection of Praat scipts for automatic extraction of phonetic properties from annotated speech recordings, as in the following overlapping collections:
        1. http://www.helsinki.fi/~lennes/praat-scripts/
        2. http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/praat.html
    2. For both annotation and perceptual experiments, Annotation Pro is a useful tool:

        http://annotationpro.org/
    3. The TGA (Time Group Analyzer) online tool extracts many kinds of information from annotations in Praat TextGrid or Character Separated Value (CSV) formats: transcription text, timing statistics, graphs showing time relations:

        http://wwwhomes.uni-bielefeld.de/gibbon/TGA/