Basic Notes on Audio Recording for Phonetics

Dafydd Gibbon, 2016-05-09

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 provide much more detailed information on different practical and theoretical aspects of recording for phonetics and speech technology:

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 definitions of 'stress': it is not a simple acoustic phonetic property but complex: phonological, a factor in speech production, or a prominently perceived item in an utterance. Some factors are valid for all purposes and include many of the following:

1. Speech styles to be recorded:
    1. Prompted ‘laboratory’ speech for phonetic experiments.
    2. Task-oriented speech.
    3. Dialogue (interview, conversation): dialogues require one microphone per person for high quality analysis.
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).

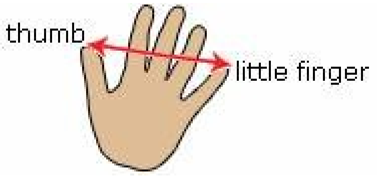
2. Microphone (to capture the range of frequencies required): many good brands are available e.g. Sennheiser, Shure, Røde; a high quality headset microphone may be preferable for some purposes. A Lavalier microphone for attaching to clothing may have somewhat lower audio quality.
3. Choice of audio file format (e.g. WAV for full quality recording, vs. MP3 or WMA compressed formats for lower quality but smaller files); however, in general I advise using WAV format 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, R-26 or other Roland devices, 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 sufficient for some purposes, e.g. syllable duration and fundamental frequency, and if neither you nor anyone else will use the data again (in general, data should be reusable, though).
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 input gain 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. quality should be the highest possible - in general record with the highest frequency and resolution available and practical – high quality can always be downgraded later if necessary:
            1. 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),
            2. medium quality: 40kHz is not strictly necessary for most phonetic purposes, as the relevant frequencies (fundamental frequency below about 600Hz (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,
            3. 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, with the standard sampling frequencies of 48kHz or 44.1kHz,
            4. sampling curiosity: in case you are curious about why the strange-looking number 44.1kHz was chosen as a standard: the product of the squares of the first 4 prime numbers >1, permitting efficient ‘down-sampling’ to 16 different lower frequencies, by dividing by combinations of these numbers:

                2^2 x 3^2 x 5^2 x 7^2 = 44100Hz

3 Recording phase

There are several points to be considered during the recording phase, in particular the scenario 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, possibly slightly to the left or the right of the speaker,
        2. A useful measure for the minimum distance (but not quite enough) is the span, the distance between outstretched thumb and little finger, between about 18...22cm for adults – so place the microphone slightly further away than this.

           Source of image: http://www.icoachmath.com/image_md/Hand-Span3.jpg
        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 receive appropriate instructions.
    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:
    1. 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 list of any problems which occurred.

4 Post-recording phase

There are many kinds of post-recording activity:

1. Systematic filing of metadata in a database (or temporarily in a word processor table),
2. Systematic 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. Cutting 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/
4. 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.:

      engexp05_2016-05-09.txt, engexp05_2016-05-09.doc, engexp05_2016-05-09.odt
   2. Text format: keep a plain text (ASCII) version of the transcription (not possible for some orthographies), and print it on paper.
5. Annotation:
   1. For all kinds of phonetic and discourse analysis it is necessary to annotate the recording with software such as Praat (the most popular), Annotation Pro, Transcriber or WaveSurfer, or ELAN (more for video than for phonetics).
   2. A selection of useful annotation software:
      1. Praat: a very versatile ‘phonetic workbench’ with high quality measurements of acoustic parameters, graphics for publications, speech synthesis, scripting (for automatising analysis), for Windows, Mac, Linux:

         http://www.fon.hum.uva.nl/praat/
      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 manual post-editing), Windows, Mac, Linux:

         http://www.sppas.org/
   3. Phonetic analysis of annotations:
      1. A wide selection of Praat scripts for ‘data mining’ of annotated speech recordings:
         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 CSV formats:

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