How does pitch analysis work?

Whenever we hear speech, we also tend to perceive some kind of "melody" in it – even if the speaker is not actually singing! The pitch pattern within an utterance or spoken expression can affect the meaning of the expression in question. In Phonetics and Linguistics, the word ***intonation*** is generally used in order to emphasize the role of pitch patterns in a language. In some languages, the meanings of individual words can be changed according to specific pitch patterns or pitch levels, and this kind of linguistic pitch patterns are referred to as ***tones***.

The pitch variation in speech is regarded as part of speech ***prosody***. Prosody refers to those properties in speech that extend over more than one speech sound segment - that is why they are sometimes also called ***suprasegmental*** features. In general, prosodic features also include stress and (pitch) accent, voice quality and speech rate.

How the larynx and the vocal folds work in speech production

Regarding the pitch and melody in speech, the most important part of the human sound producing organs is the ***larynx***, where the ***vocal folds*** are located (physiologically, they are not really "vocal cords", although you might have heard the somewhat deprecated term). The ***glottis*** refers to the opening between the vocal folds. The vocal folds can be brought together so as to vibrate as the speaker is breathing. This concerted activity is called ***phonation***.

During voiced sounds, the speaker's vocal folds are vibrating, whereas during voiceless sounds, the vocal folds are kept apart, i.e., they are not used for producing sound. If you put your finger on your "Adam's apple", in front of your neck, while speaking, you can feel a slight vibration during vowels and voiced consonants. (The vocal folds are located right behind the "Adam's apple" that is actually part of the *thyroid cartilage* that protects your larynx.)

During regular voicing, i.e., in so-called ***modal phonation***, the vocal folds are closed while the speaker is breathing out. The pressure of the air under the closed glottis makes the vocal folds burst open and return to their closed position at a fast pace. Thus, the speaker is not actively "swinging" the vocal folds but just letting them "roll" in the airflow. If it were possible to listen to the sound created within the larynx, it would be a sort of buzzing noise. The mechanism is somewhat analogous to making a "rolling" sound with your lips (which would be a *bilabial trill* in phonetic terms).

How does pitch analysis from speech relate to laryngeal function?

If you are interested in the frequency of vibration of the vocal folds (the ***glottal frequency***) during voiced sounds, you can try to analyze the ***fundamental frequency*** of the voiced parts of the speech sample. Fundamental frequency is measured in hertz (Hz): one hertz refers to one cycle per second (in the glottis, one open-close cycle of the vocal folds). Since the fundamental frequency calculated from voiced sounds is related to the perceived pitch of speech, fundamental frequency analysis can be applied in intonation research, for example.

The fundamental frequency analysis in Praat is called ***Pitch*** analysis, because the particular method is specially adapted for studying human speech. Pitch analysis is based on the assumption that, in the audio signal, it is possible to locate successive nearly identical

periods whose durations correspond to the durations of vocal fold vibrations that are possible and likely for the speaker in question. When you zoom in on a voiced speech sound in the audio sample so that you can see about half a second of speech at a time, the individual periods should become clearly visible in the sound waveform.

The following (silent) video introduces the structure and functions of the larynx:

https://youtu.be/kfkFTw3sBXQ

The speaker may produce sounds other than modal phonation on his/her larynx. He/she can **whisper**, for example, in which case the vocal folds are held together as well, but a very small gap is left between the arytenoids for the air to flow through.
This position of the arytenoid cartilages and the vocal folds can be seen in the previous video, starting from 0:22.

Thus, when whispering, the vocal folds do not necessarily vibrate very much. Instead, a "hissing" noise is created in the narrow constriction between the arytenoids. So, the noise replaces voicing during whisper.

Even in whispered speech, the perceived pitch varies to some extent. However, the implementation mechanism of the pitch contour is different from that during modal phonation, as it is not based on periodicity. For this reason, the pitch level or pitch variation during whisper cannot be measured by the same method as during normal voicing.

Which characteristics of speech can be affected by laryngeal function?

The pitch observed in the speech is largely based on the oscillation frequency of the speaker's vocal folds for voiced sounds. Thus, the speaker is able to influence the pitch of his/her speech by regulating the settings of his/her larynx during speech.The oscillation frequency of the vocal folds changes mostly depending on how tight or loose they are. The tension of the vocal folds is regulated by the speaker especially by turning the arytenoid cartilages attached to the back of the vocal folds.

In addition to the pitch of the voice, the speaker can also control the manner in which the vocal folds vibrate. The position of the larynx and the muscles around it has an overall effect on the phonation and thus on the sound quality observed in speech. For example, breathy or creaky voice can be due to an individual habit. On the other hand, in some languages, different voice qualities can also be used for distinguishing the meanings of words.

The voice can be influenced and trained by the speaker to some extent, but the typical pitch of the speaker as well as the range of sounds depend largely on physical characteristics. Long vocal folds vibrate more slowly and can therefore produce lower sounds than short vocal folds. There are, of course, a lot of individual differences. Moreover, the voice may change with age, and health conditions are also easily reflected on the voice.

Pitch analysis does not always provide useful information

In summary, the results of pitch analysis can be interpreted most easily if the speaker speaks in his or her “ordinary voice” and does not have voice-related speech disorders.

Some speakers tend to lower their voices e.g. at the end of their statements (even by some force) so that the vibration of the vocal folds becomes intermittent. In this case, the voice may sound "pressed" and "creaky". During creak, the durations of successive glottal periods may vary so much that a consistent measurement cannot be obtained from the recorded audio signal by conventional pitch analysis methods. In some cases, the fundamental frequency of the creaky voice may simply drop below the lower limit of the minimum pitch setting for that speaker, so that portion will be automatically excluded from the analysis result. Creaky voice can sometimes be created in a special way where the vocal folds are vibrating in their entire length while a shorter part of the vocal folds is simultaneously vibrating at a faster rate. Such irregularities can interfere with pitch detection, and care should be taken with the results.

If there are sudden "jumps" up or down or very long interruptions in the pitch curve, you should always take a closer look at the sound sample, even if you think the voice sounds uniform and normal. It can be useful to check the pitch analysis settings and consider the role of potential voice quality changes.

Fundamental frequency and perceived pitch in speech

The fundamental frequency calculated from voiced portions in speech has an indirect relation to the perceived pitch of the speech. In general, the higher the fundamental frequency, the higher the perceived pitch of the voice. However, people are most sensitive to frequency differences between sounds whose frequencies are below a couple of thousand hertz, whereas the auditory discrimination ability is not as good with sounds in higher frequencies. For this reason, frequency differences reported on the Hertz scale do not directly reflect the magnitude of the perceived pitch difference, and they may not be useful in auditory perception studies and speech research. As a solution, various auditory pitch scales have been developed for describing human perception, such as the ***mel*** scale and the ***semitone*** scale.

Perceptual pitch scales are always relative rather than absolute. For example, it is not possible to report the pitch of a particular sound as "20 semitones". Instead, it is necessary to state the reference pitch that was used, e.g. “20 semitones above 100 Hz” or "20 ST (re 100 Hz)".

Thus, in order to convert an absolute frequency (expressed in hertz) to semitone scale, a ***reference value*** must be selected to which all the Hertz values are compared. A pitch value reported in semitone scale is describes the pitch difference that listeners would perceive if they were to compare a sound with that frequency/pitch to the selected reference frequency/pitch. Thus, when using the semitone scale, negative (minus sign) values may also occur: a negative number refers to a fundamental frequency below the reference value. The reference frequency of 100 Hz (***semitones re 100 Hz***) is used by default in Praat. Other reference values are also possible, so you need to clearly indicate which reference frequency is used when reporting your results.

If you want to find out how many semitones away the two measured frequencies are from each other, you must first convert each frequency to the semitone scale according to the same reference frequency, for example “-1.5 ST re 100 Hz” (this means the pitch at 1.5 semitones below 100 Hz!) and “5.5 ST re 100 Hz” (5.5 semitones above 100 Hz). The difference in pitch can then be calculated, which in this case would be 7 ST, regardless of the reference frequency to which the original frequencies were compared.

It is good to note that the difference between two notes reported in semitone scale cannot be converted to hertz scale. When reporting key figures related to speech pitch, it is often useful to calculate the statistics both in Hertz and in semitones.

The pitch differences reported in the computational semitone scale are easy to interpret if you are familiar with musical intervals. For example, a pitch difference of 12 semitones should perceptually correspond to one octave, and a pitch difference of one semitone should correspond to the difference between two adjacent piano keys. Pitch values reported in semitone scale will usually be small numbers that are easier to read.

How to perform pitch analysis in Praat?

There are usually two ways to analyze sound samples in Praat:

- in the editor window, where the analysis graphs can be viewed together with the audio sample, browsed and zoomed in and out; or
- via the Object window, where new analysis objects can be created from the selected Sound object.

Pitch analysis can also be performed in the aforementioned ways in Praat.

Within the editor

In the Sound and TextGrid editor windows in Praat, you can show or hide the pitch curve by selecting the **Show pitch** command in the Pitch menu. The pitch contour is shown in blue in the editor window. Likewise, the display range for the pitch curve is shown in blue numbers on the right side of the editor window. When you click on a point in the waveform or analysis displays, red grid lines will appear at the selected point. Again, the pitch value at the selected time (that is, at the red vertical line) is displayed in blue numbers at the corresponding point on the right side of the editor window. If you need to copy the exact f0 value to another program, choose **Get pitch** from the **Pitch** menu. The measured value appears in the Info window, and you can select the text and copy and paste it somewhere else.

Pitch analysis via the Object window

In the Object window, a new Pitch object can be created by first selecting the desired sound object and pressing the **Analyze periodicity: To Pitch...** button in the dynamic menu. The parameters *Pitch floor* and *Pitch ceiling* should be supplied according to the speaker in question, in the same manner as the Pitch range that is defined when analysing pitch in the editor windows (see the explanation of the settings below). Measurements can be taken from the resulting new Pitch object by clicking on the **Query** button or the contour can be plotted into the Picture window with **Draw** commands. The contents of the object can be viewed in the Pitch editor (select the Pitch object and press **Edit**). The selected Pitch object can also be saved using the **Save** menu commands.

Pitch analysis settings

In the editor window, the settings for pitch analysis can be changed in the Pitch menu, under Pitch settings... If the analysis is calculated in the Object window, supply the corresponding settings in the form dialog box before pressing OK.

Good to know: In practice, the pitch curve is computed by moving a small time frame over the original sound in small time steps. At each step, the pitch algorithm tries to detect a pitch value from the audio clip inside the frame. If a result is found that matches the given criteria, the pitch contour will include the point calculated for that window. In some frames, the analysis algorithm does not find sufficient periodicity. In that case, there is a break in the pitch curve at that time, since the analysis algorithm interprets the sound contained in that time window as "unvoiced".

In order for the analysis to be as successful as possible, it is important to **check the minimum and maximum frequency in the settings:** in the editor, these will be the *Pitch range (Hz)* and the two numbers following it, and in the Object window the same settings are called *Pitch floor (Hz)* and *Pitch ceiling (Hz)*, respectively. The problem is that the fundamental frequency ranges used by different speakers can vary a lot, and so the parameters often need to be adjusted according to the voice that is being studied.

The minimum frequency, or *Pitch floor (Hz)*, determines the lowest possible frequency that can be included in the Pitch contour. The lower limit determines the duration of the time frame that is inspected at each time step. The size of the analysis frame will be three times the duration of the longest possible period, according to the given lower limit. For example, if the pitch floor is set to 75 Hz, the time window that is moved over the sound sample will be 3/75 = 0.04 seconds long.

The pitch floor should not be lowered unless this is necessary in order to match a low-pitched voice. ***If the limit is set too low,*** the time frame will grow very long. This makes it hard to detect fast pitch changes that might be of interest for the researcher, as the curve will become too "smooth".

On the other hand, ***if the pitch floor is too high for the speaker's voice***, the time window will not be long enough to include three complete pitch periods. In this case, the pitch curve may sometimes include values that are one octave above the original pitch, or there might not be any pitch values at all in some of the low-pitched portions.

In each time frame, several pitch candidates are often detected during the first stage of the pitch analysis. The pitch ceiling value is needed in the latter stage in order to compute the "cheapest path" between the consecutive time windows. The aim of this process is to avoid sudden, probably erroneous pitch jumps on the final pitch contour.

The maximum frequency, or *Pitch ceiling (Hz)*, determines the upper frequency limit above which the fundamental frequency candidates are not considered.

The limits are thus set in such a way that the fundamental frequency of the speaker in question can be assumed to move between them. In practice, the default pitch floor is 75 Hz, which is often suitable for a low-pitched voice (for example, the average adult male speaker). For an even lower-pitched voice, the floor may be set at 50-60 Hz, or even less for creaky voice. The upper limit should then also be lowered to, e.g., 300 Hz. For a high-pitch voice (e.g., the average adult female), the lower limit can be set slightly higher, e.g., at 100 Hz, and the upper limit could be, for example, 500 Hz. If the speaker is a high-pitched child, you can try and set the lower limit around 200 Hz, and the upper limit should also be high enough. In any case, it is often worthwhile to experiment in the editor window in order to discover which parameters are likely to produce a plausible fundamental frequency curve.

There are further possibilities for fine-tuning the pitch analysis in Praat. However, it is best to get acquainted with the alternative methods via the internal manual in Praat, in case this becomes relevant for your work.

Note. If you only want to change the scale of the visible pitch curve in the editor window so that it fits the display, select **Advanced pitch settings...** from the **Pitch** menu and set the appropriate lower and upper limits for the *View range*. This setting will not affect the operation of the analysis or the shape of the curve, only the location and height of the graph displayed on the screen.

Why are there gaps in the pitch curve?

During unvoiced sounds, the vocal folds do not vibrate against each other, so for such sounds, pitch analysis will usually not detect much periodicity, and this results in gaps. For example, the unvoiced stop consonants **[k p t]** include a closure phase that tends to be almost completely silent, as the speaker's vocal folds are not vibrating and the airflow via the nasal and oral pathways is blocked for a brief moment. Of course, a longer break in the pitch contour may be due to an actual pause. In case of voice quality changes and irregularities, the voice detection may fail and breaks may appear in the curve. This problem may not be fixed even by changing the analysis settings. In case you are working with voice analysis especially, you can try selecting the alternative pitch detection method in Praat, i.e., the ***cross-correlation*** algorithm, and see whether that helps.

All speech-producing organs are physical parts of the human body and moving them takes time and consumes energy. This is worth remembering also when making pitch measurements from continuous speech. For example, changing the position of the larynx and vocal folds between a voiced and an unvoiced sound cannot in practice occur at lightning speed, and therefore "partially voiced" portions can often be observed between voiced and unvoiced sounds. When you look at ordinary everyday speech a little more closely, a consonant segment that is expected to be voiceless from the point of view of the sound system, may be at least partly or even completely voiced in real speech. In fast Finnish speech, glottal periods can often be observed in the sound waveform also during stop consonants [k], [p] or [t], and yet at the same time it can be heard and visually observed that there has been a stop-like closure at the corresponding place of articulation. In Finnish, voiceless stop consonants easily "catch" some voicing, especially when they occur between vowels or other voiced sounds. However, similarly to other ***coarticulation*** phenomena, the degree of voicing in real speech heavily depends on the context, on the speech situation and the speaker.