PITCH DETERMINATION OF SPEECH SIGNALS - A SURVEY

Wolfgang J. Hess

Lehrstuhl für Datenverarbeitung, TU München
Postfach 202420, D-8000 München 2, W. Germany

In this short survey, the numerous pitch determination algorithms (PDAs), devices, and methods are grouped into two major classes: time-domain PDAs and short-term analysis PDAs. The short-term analysis PDAs leave the time domain by some short-term transform specific to the individual method applied. They provide a sequence of average estimates of fundamental frequency FO or fundamental period TO emerging from the short-term intervals (frames). Opposed to this,the time-do-

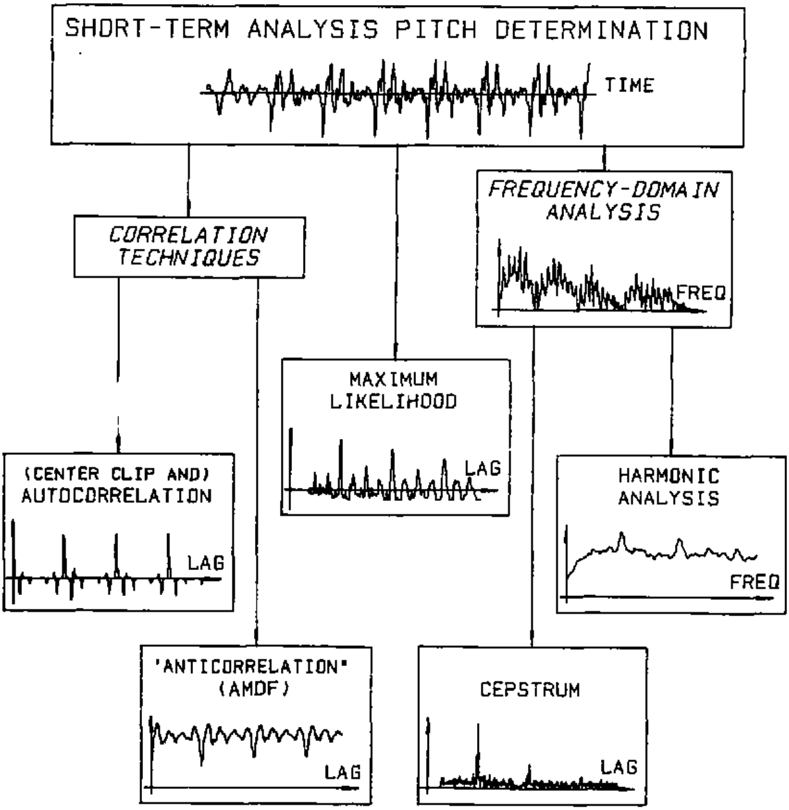


Fig. 1

main PDAs extract the information of periodicity from the signal directly or after filtering; these PDAs are usually able to track individual fundamental periods.

Fig. 1 presents an overview of the prevailing short-term analysis PDAs. These PDAs are further grouped according to the individual short-term transformation applied.The sequence of operations is quite similar for all these PDAs. After an optional preprocessing step (which can be a moderate low-pass filter or an adaptive center clipper),the signal is subdivided into short segments (frames) whose typical duration ranges between 30 and 50 ms. This value is chosen because most of the short-term transforms (with the only exception of the AMDF, see below) require several pitch periods within a single frame. After the formation of the frames, the short-term transform is performed on every frame individually. This transform is intended to perform in a way that it concentrates all the available information on periodicity into a single principal peak which is then detected by a peak detector and labeled as the pitch estimate for that frame. For each transform, the figure shows a typical example (computed from the 50 ms frame displayed in the upper part) of the short-term function from which the estimate of pitch is derived. Not all the known short-term transformations, however, behave in the desired way. Among those which do, we find the well-known short-term autocorrelation technique (Rabiner 1977) whose performance is greatly improved when the signal is preprocessed by center clipping. Its counterpart, the average magnitude difference function AMDF (Ross et al. 1974) is defined by

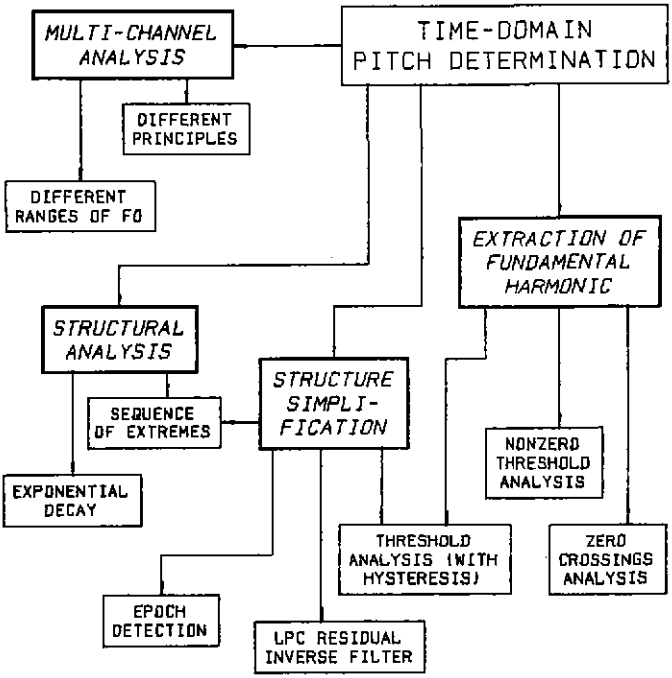
$$\mathrm{AMDF}(l) = \quad | a(n) - a(n+l) |$$

where a(i) represents the signal.The summation is performed over the whole frame which, in this case, is permitted to be quite short. The AMDF reveals a strong minimum at a lag equalling the period T0. The well known cepstrum analysis (Noll 1967) as well as harmonic analysis (Schroeder 1968,Terhardt 1979) are frequency--domain methods. Both are derived from the logarithmic power Fourier spectrum. The cepstrum PDA transforms this spectrum back into time domain, thus generating a large peak at T0. One of the several methods of harmonic analysis is spectral compression. The log power spectrum is compressed along the frequency axis by integer factors k=2, 3, 4 etc. Adding all these compressed spectra causes the harmonics to contribute coherently to the distinct peak at F0. The maximum likelihood PDA (Friedman 1977), finally, represents the mathematical procedure of detecting a periodic signal with unknown period T0 in a noisy environment.

Short-term analysis PDAs are unable to track individual periods since the phase relation between the signal and the short-term function is lost during the transform. This loss, however, makes these PDAs insensitive to phase distortion of the signal. Moreover, since they focus the whole information of periodicity into one single peak,they tend to be resistive against noise and signal degradation. The disadvantage of large computing effort more and more loses its importance because of the current progress in fast digital circuitry.

For the time-domain methods, the grouping provided in fig. 2 is somewhat more arbitrary. Leaving aside the multi-channel PDAs, however, we find most time-domain PDAs somewhere between two extremes: analysis of the temporal signal structure on the one hand and extraction of the fundamental harmonic on the other. In the typical single-channel PDA, we find a (linear or nonlinear) filter as the first step of signal processing. The effect necessary for this filter ranges from zero for structural analysis to extreme low-pass filtering for the PDAs which extract the fundamental waveform from the signal. Contrary to this,

Fig. 2


MULTI-CHANNEL ANALYSIS
TIME-DOMAIN PITCH DETERMINATION
DIFFERENT PRINCIPLES
DIFFERENT RANGES OF F0
EXTRACTION OF FUNDAMENTAL HARMONIC
STRUCTURAL ANALYSIS
STRUCTURE SIMPLI-FICATION
SEQUENCE OF EXTREMES
NONZERO THRESHOLD ANALYSIS
EXPONENTIAL DECAY
THRESHOLD ANALYSIS (WITH HYSTERESIS)
ZERO CROSSINGS ANALYSIS
EPOCH DETECTION
LPC RESIDUAL INVERSE FILTER

the subsequent period detection algorithms range from a simple zero crossings analyser for fundamental waveform extraction to the highly complicated procedures for structural analysis.Further details on these analysers are shown in figs. 3 and 4.

The structural analysers use the fact that the speech signal consists of several exponentially damped oscillations (fig. 3).The signal may thus be modeled either as a sequence of extremes, such as maxima, minima, or zero crossings (Tucker and Bates 1978), or by its envelope in form of a decaying exponential (Filip 1969). Just this last concept leads to a very simple analog circuit consisting of a diode,a capacitor, and two resistors. The disadvantage of this concept, however, is that it cannot cover the whole range of F0 possible for a multitude of speakers (at least 70 to 500 Hz) without producing significant errors.

At the other end of the scale, we find the PDAs which extract the fundamental waveform (fig. 4). Among these PDAs there are the oldest realizations, and a detailled survey of much of this work is presented by McKinney (1965).The amount of low-pass filtering necessary to realise this concept depends on the subsequent periodicity extractor: the simple zero crossings analysis extractor needs

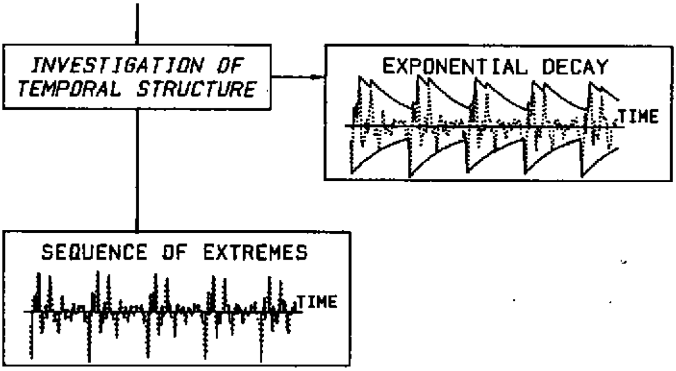


Fig. 3

the greatest degree of low-pass filtering,whereas the nonzero threshold analysis extractor, somewhat more complicated due to the need of level normalization or threshold adaptation, can tolerate the presence of higher harmonics to a certain extent. A further step towards tolerating the presence of higher harmonics is made when a hysteresis is incorporated into the threshold analyser. A single--channel PDA based on this concept,however, cannot cope with all types of signal and all ranges of FO. For instance, the concept of fundamental harmonic extraction requires the first harmonic to be present in the signal. If this does not apply, the first harmonic can be reconstructed by nonlinear distortion of the signal. In doing so, on the other hand, one risks to attenuate this very harmonic when it is strong in the original signal. The only way out, if one intends to process arbitrary signals by the relatively simple time-domain concept, is the implementation of a PDA which contains several channels,and which is provided with the facility to automatically switch between them. These channels may process subranges of FO with identical implementation (McKinney 1965), or they may represent different implementations, such as different ways of nonlinear distortion covering the whole range of FO in each channel (Hess 1979). These concepts, compared to the corresponding single-channel PDAs, give a considerable improvement in performance.

In spite of the many proposals and different principles, the problem of pitch determination in speech communication is still unsolved (Rabiner et al. 1976). Further research has to be done towards answering the question what kind of errors are most annoying from the perceptual point of view, and by what means a PDA can be optimised in order to produce a minimum number of these errors.

Selected References.

FILIP M. (1969): Envelope periodicity detection. JASA, vol. 45, pp. 719...732

FRIEDMAN D.H. (1977.1): Pseudo-maximum-likelihood pitch extraction. IEEE-T-ASSP, vol. 25, pp. 213...221

HESS W.J. (1979): Time-domain pitch period extraction of speech signals using three nonlinear digital filters. Proc. ICASSP-79, pp. 773...776

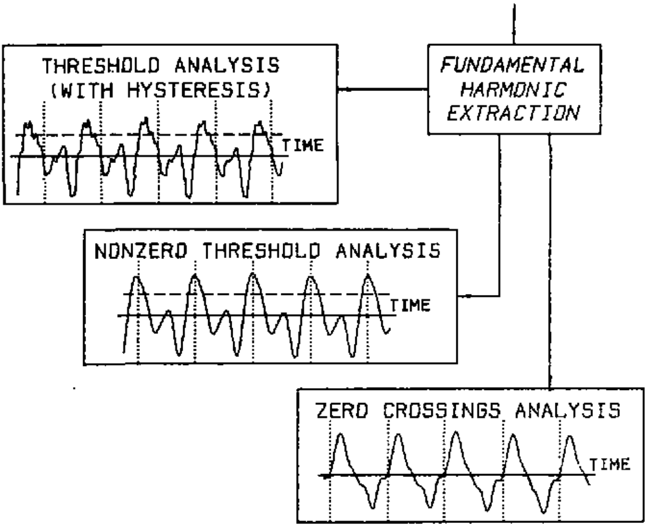

THRESHOLD ANALYSIS (WITH HYSTERESIS)
TIME
FUNDAMENTAL HARMONIC EXTRACTION
NONZERO THRESHOLD ANALYSIS
TIME
ZERO CROSSINGS ANALYSIS
TIME

Fig. 4

McKINNEY N.P. (1965): Laryngeal frequency analysis for linguistic research. Ann Arbor MI: Communic.Sciences Lab., Univ. of Michigan;Res.Rept. No. 14,340 pp.

NOLL A.M. (1967): Cepstrum Pitch Determination. JASA, vol. 41, pp. 293...309

RABINER L.R. (1977.1): On the use of autocorrelation analysis for pitch detection. IEEE-T-ASSP, vol. 25, pp. 24...33

RABINER L.R., CHENG M.J., ROSENBERG A.E., McGONEGAL A. (1976): A comparative study ... IEEE-T-ASSP, vol. 24,pp. 399...413

ROSS M.J., SHAFFER H.L., COHEN A., FREUDBERG R., MANLEY H.J. (1974): Average magnitude difference function pitch ... IEEE-T-ASSP, vol. 22,pp. 353... 361

SCHROEDER M.R. (1968): Period histogram and product spectrum: New methods for fundamental-frequency measurement. JASA, vol. 43, pp. 829...834

TERHARDT E. (1979): Calculating virtual pitch. Hearing Research, vol. 1

TUCKER W.H., BATES R.T.H. (1978): A pitch estimation algorithm for speech and music. IEEE-T-ASSP, vol. 26, pp. 597...604