

Parameterisation Methods of the Glottal Flow Estimated by Inverse Filtering

Paavo Alku

Helsinki University of Technology

Finland

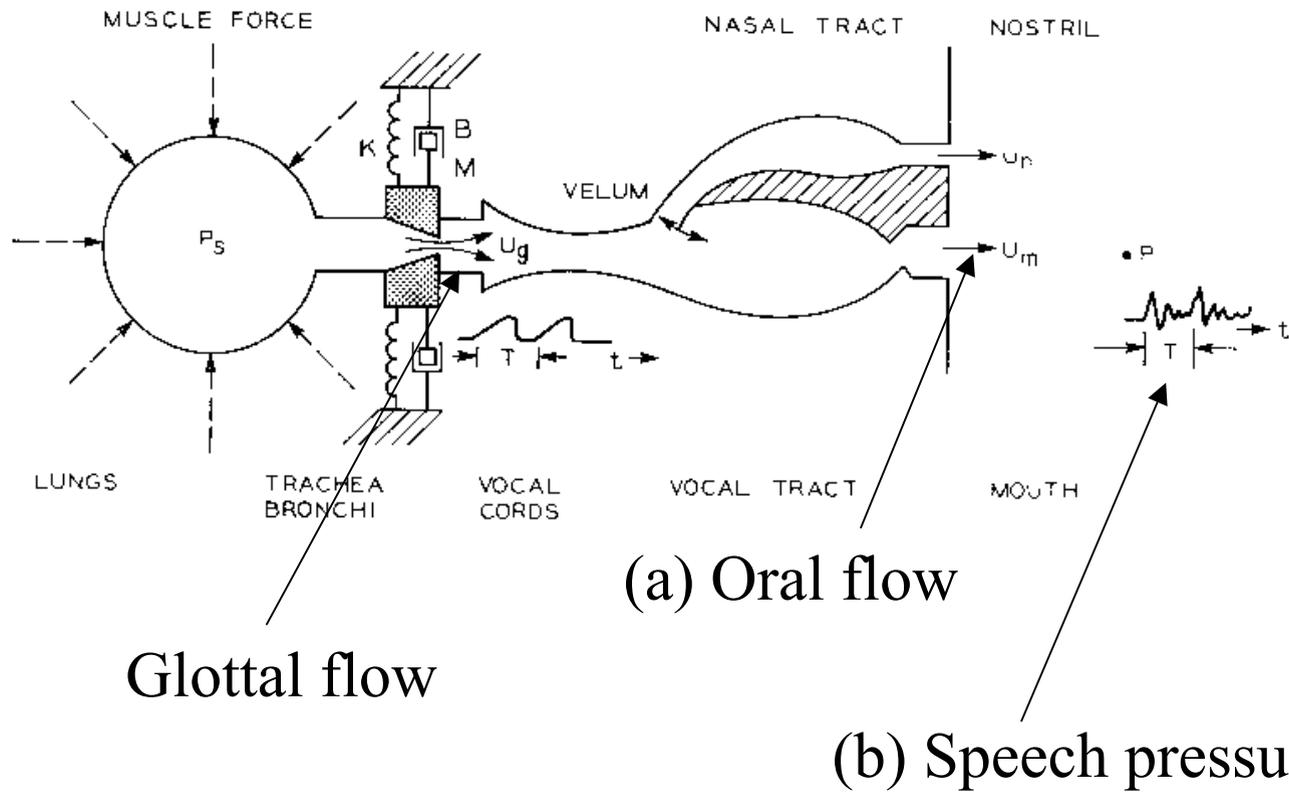
Contents

1. Introduction
2. Time-domain methods
3. Frequency-domain methods
4. Conclusions

1. Introduction

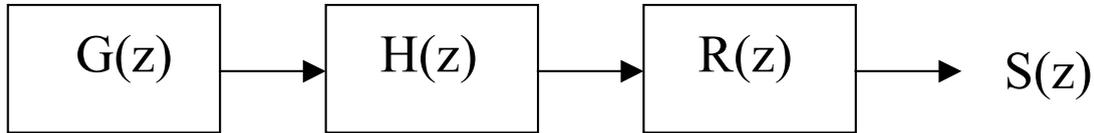
Inverse filtering:

Estimation of the glottal flow either from (a) the oral flow or from (b) the speech pressure waveform recorded in free field.



Idea of inverse filtering

1) Model of voice production



$G(z)$ = Glottal flow (the goal of the study !!)

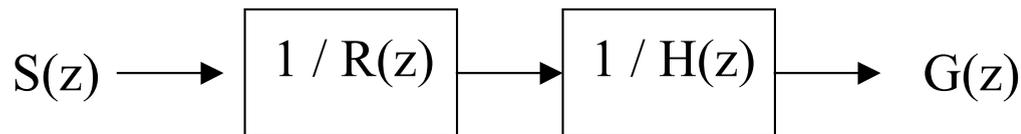
$H(z)$ = Vocal tract

$R(z)$ = Lip radiation

$S(z)$ = Speech signal (the only input !!)

2) Solving the vocal tract transfer function $H(z)$

3) Computation of $G(z)$ by **cancelling the effect of the vocal tract.**



Analysis of the glottal source consists of two stages:

(a) Inverse filtering

(b) Parameterisation =

Expressing the most important features of the obtained glottal flows in compressed numerical forms.

Parameterisation methods developed are discussed

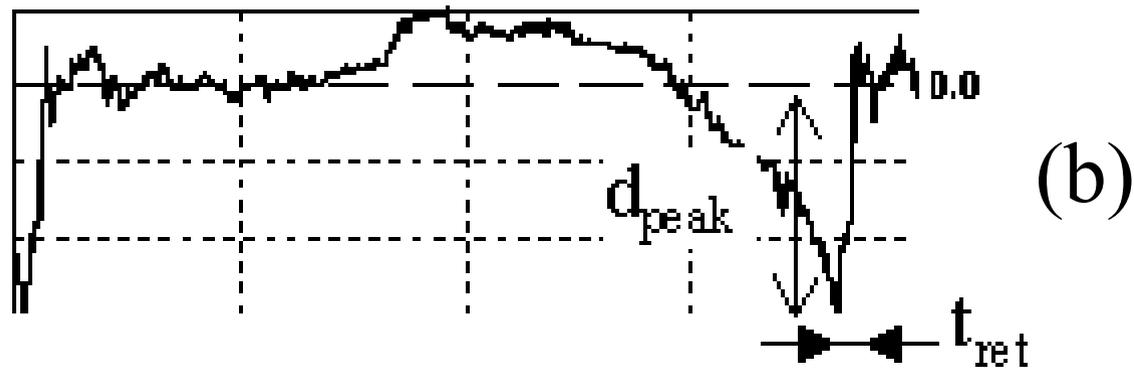
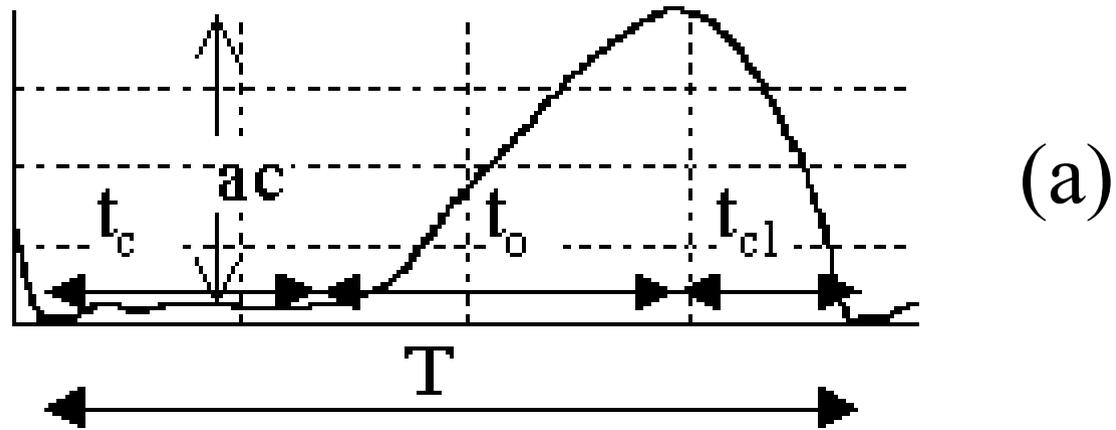
Assumptions:

- Input information is either the glottal flow or its derivative obtained by inverse filtering
- No other information signals (e.g., subglottal pressure) are to be used
- Either oral flow or pressure signal is used in inverse filtering

2. Time-domain methods

Extraction of critical time-instants
or amplitude values directly from the time-
domain waveforms of the flow or its
derivative

Glottal flow (a) and its derivative (b)



2.1 Time-based parameters

(a) “Classical”:

- Open quotient (OQ) = $(t_o + t_{cl}) / T$
- Speed quotient (SQ) = t_o / t_{cl}
- Closing quotient (ClQ) = t_{cl} / T

(b) “Similar”:

- Closed quotient (CQ) = $t_c / T = 1 - OQ$

- Return quotient (RQ) = t_{ret} / T

OQ, ClQ, and SQ have been used widely in the analysis of the glottal source behaviour:

- in different loudness and pitch conditions (e.g., Holmberg *et al.*, JASA, 1988; Sulter, Wit, JASA, 1996)
- in assessment of vocal disorders (e.g., Hillman *et al.*, JSHR, 1989)

- in measuring effects of vocal loading in employees (such as teachers) with intensive voice use (e.g., Vilkman *et al.*, 1997)
- in studying effects of the glottal flow characteristics on speech perception (e.g., Childers and Lee, JASA, 1988).
- in studying the glottal flow characteristics of singing voices (e.g., Sundberg *et al.*, J. Voice, 1993)

- (Very) general trends when intensity is raised:

OQ: decreases

C1Q: decreases

SQ: first increases, then decreases

- Typical values (males, normal loudness, Holmberg *et al.*, 1988):

OQ=0.6

C1Q=0.2

SQ=1.8

Extraction of the classical time-based parameters might be **problematic** due to, for example, artifacts in inverse filtering or gradual opening of the vocal folds

=> The true critical time-instants are replaced by artificial instants when the flow crosses a certain level (e.g., 50 % of the max-min difference)

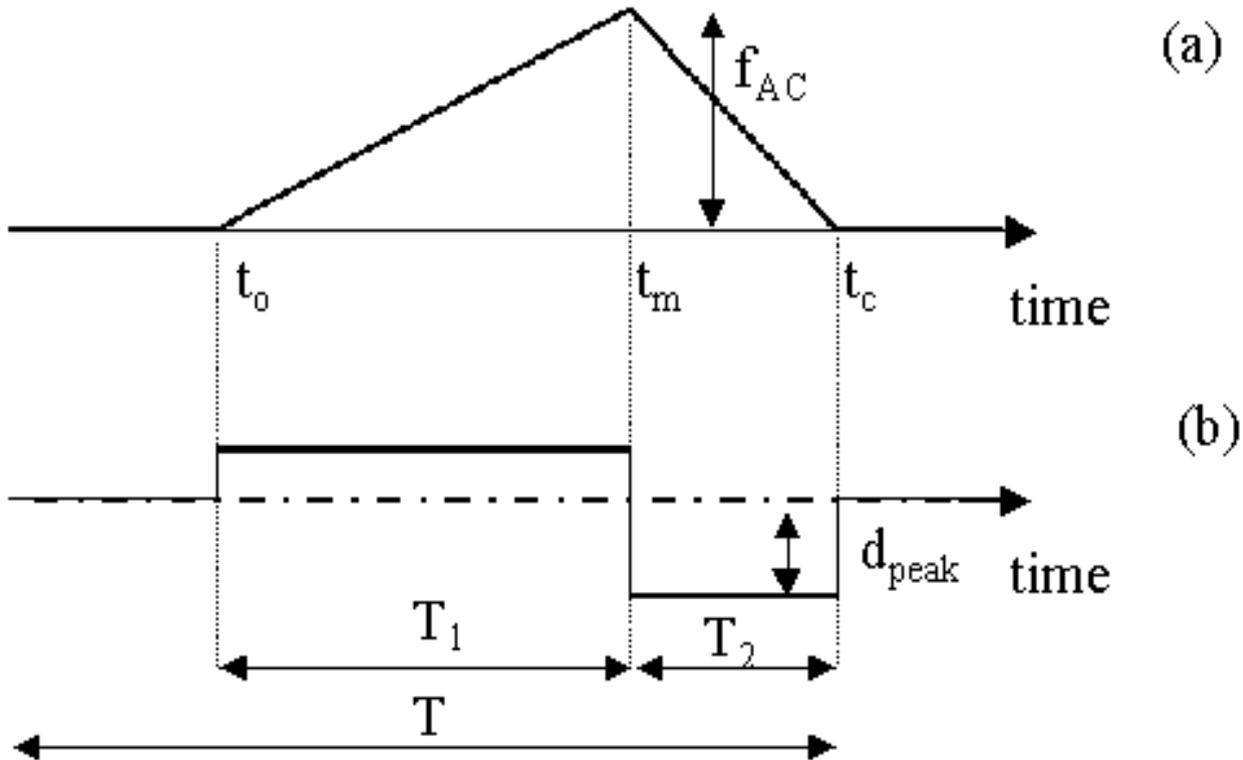
(Dromey *et al.*, J. Voice 1992)

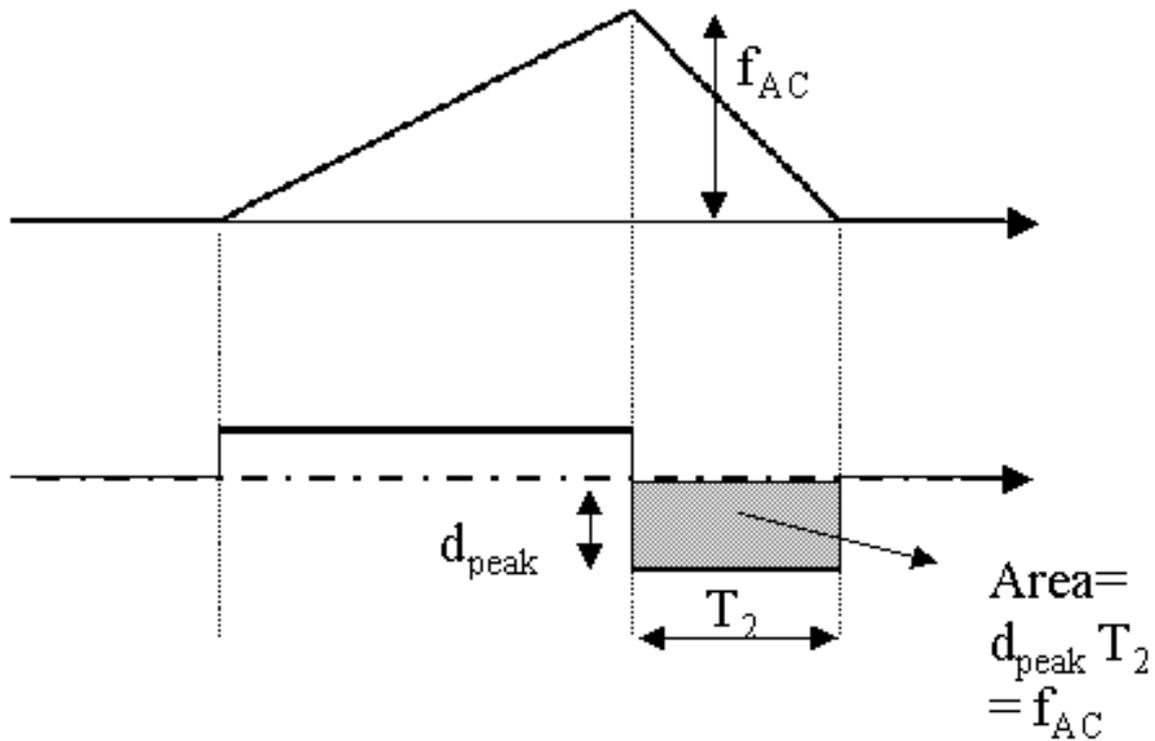
Another remedy: Development of robust methods to the time-based parameterisation of the glottal flow

=>

Normalized amplitude quotient
(Alku *et al.*, JASA, 2002)

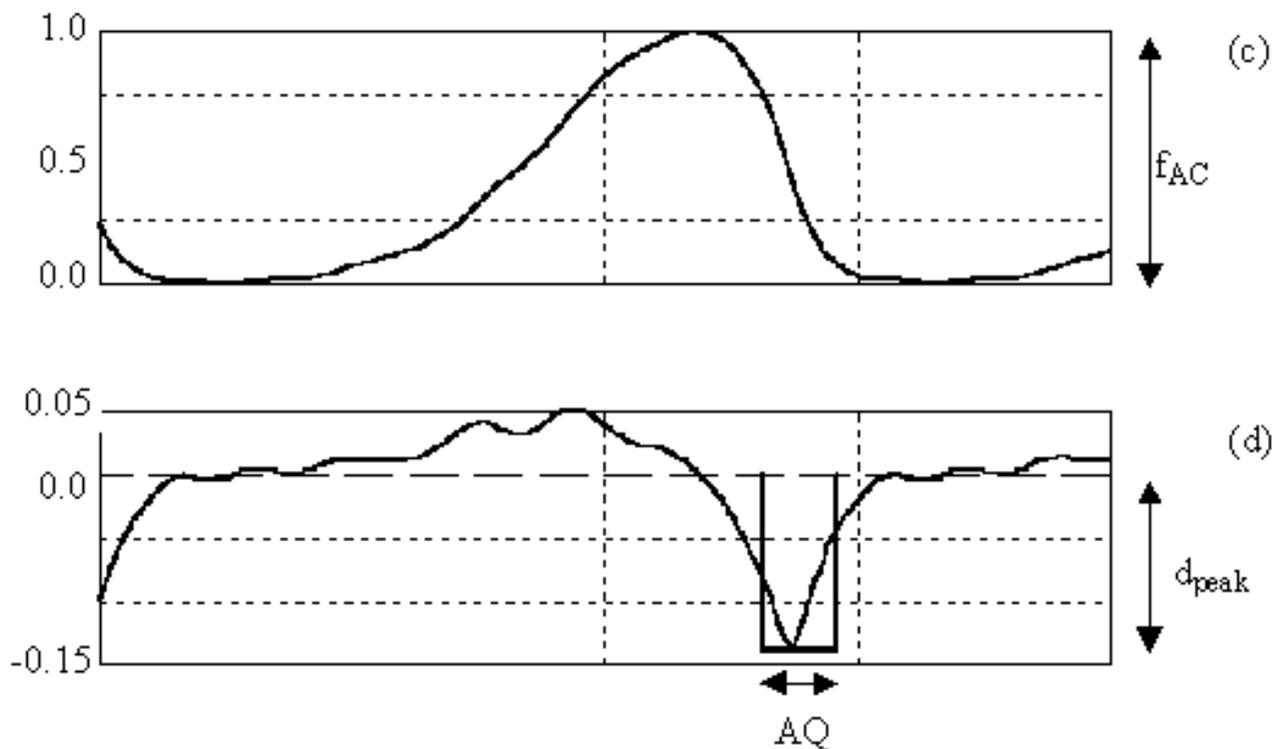
Triangular-shaped flow (a) and its derivative (b):



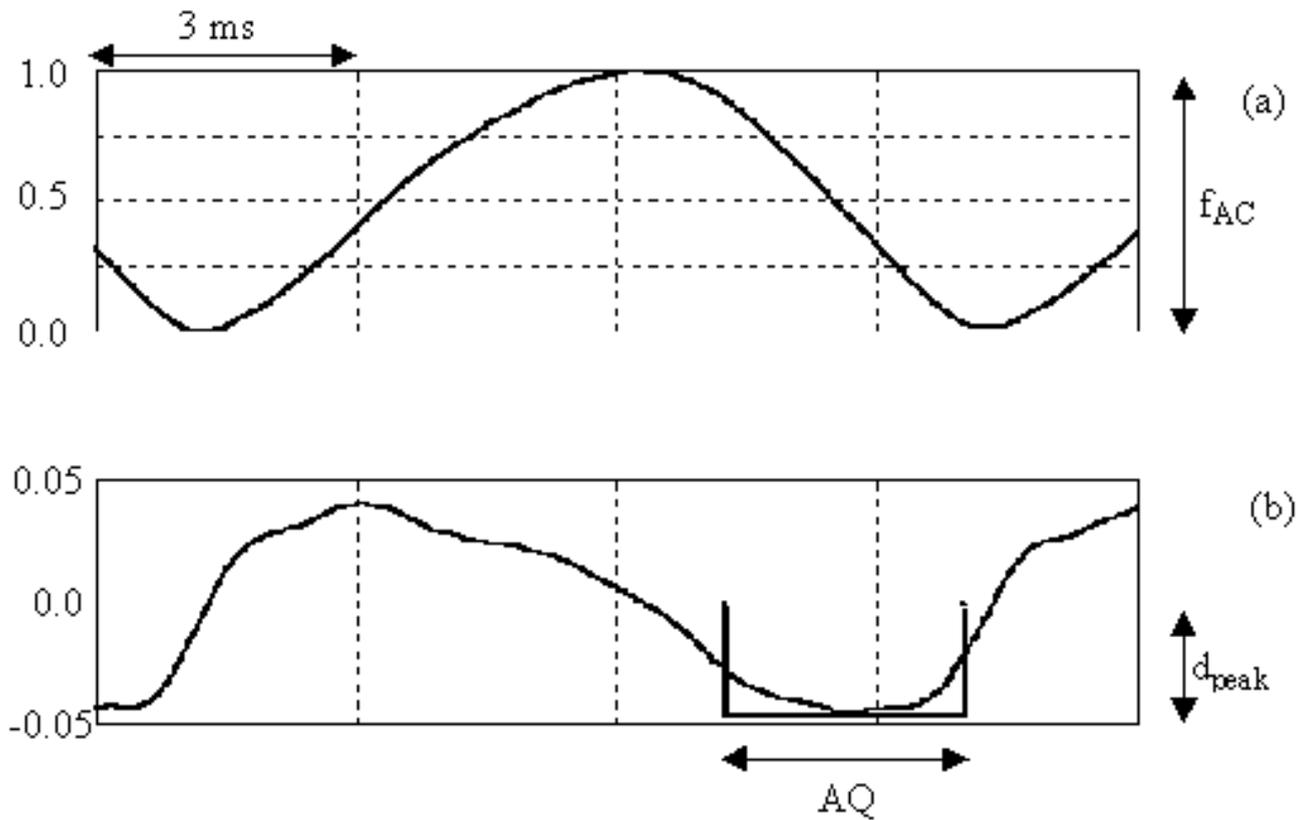


$$\Rightarrow CQ = T_2 / T = f_{AC} / (d_{peak} T) = AQ / T$$

Example of Amplitude Quotient (AQ) values for glottal pulses inverse filtered from real speech (normal phonation):



Example of Amplitude Quotient (AQ) values for glottal pulses inverse filtered from real speech (breathy phonation):

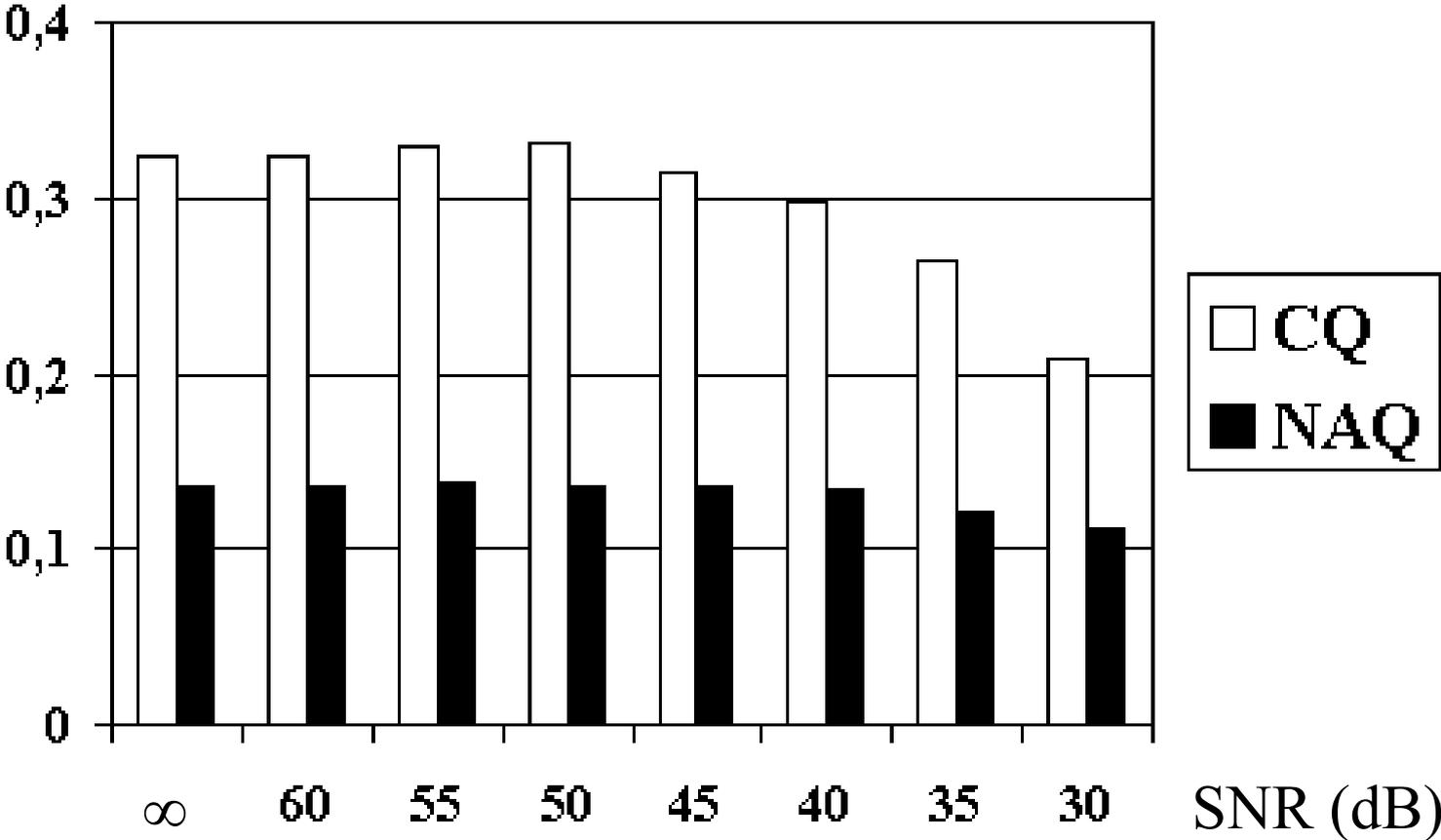


A time-based parameter to measure a sub-length of the glottal closing phase from two **amplitude-domain measures** is obtained:

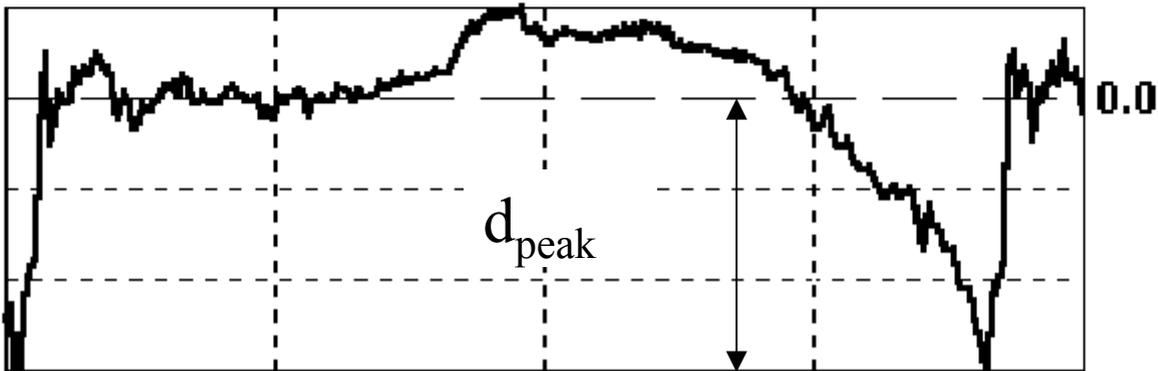
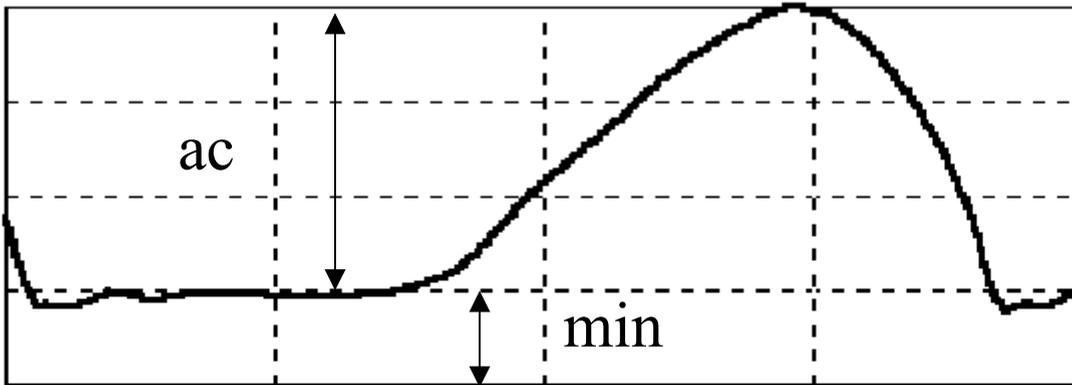
Normalized Amplitude Quotient:

$$\text{NAQ} = \text{ac} / (d_{\text{peak}} \cdot T)$$

Robustness of NAQ vs. ClQ



2.2 Amplitude-based parameters



Obtaining amplitude-domain parameters calls for using a calibrated flow mask.

Typical values (Holmberg *et al.*, JASA, 1988):

$$a_c = 0.25 \text{ l/s}$$

$$a_{\min} = 0.10 \text{ l/s}$$

$$d_{\text{peak}} = 280 \text{ l} / \text{s}^2$$

Remarks on the amplitude-domain parameters:

ac : - larger for males than for females
- increases when intensity is raised

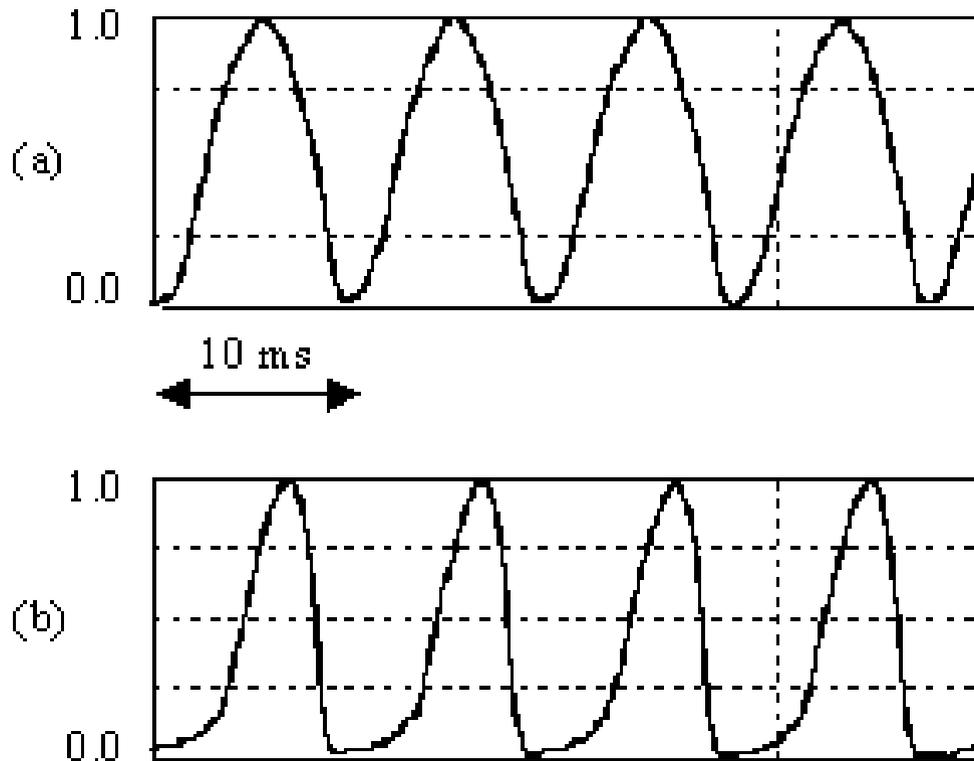
min: - typically non-zero
- decreases when intensity is raised

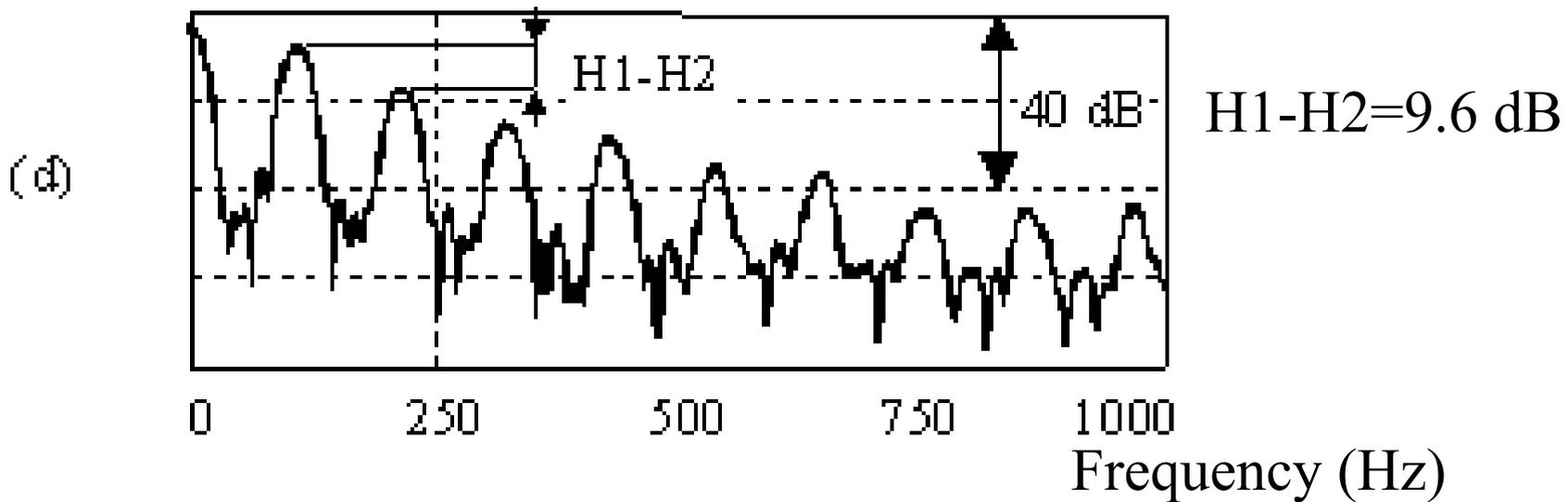
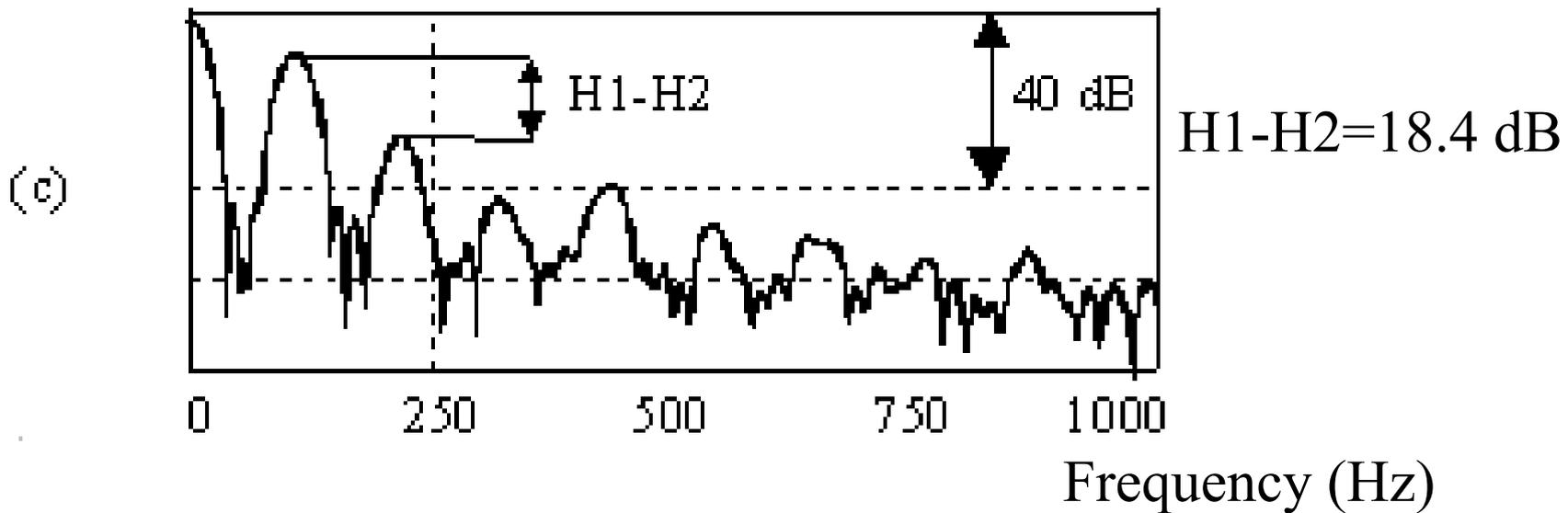
d_{peak} : - strong correlation with SPL

3. Frequency-domain methods

- Parameterisation is typically based on quantifying the decay of the voice source spectrum
- Decay of the spectrum is measured typically using the harmonics
- Parameters of all-pole models can also be used in quantifying the decay of the glottal source spectrum

An example of frequency-domain parameterisation method:
H1-H2 difference (Titze, Sundberg, JASA, 1992)





4. Conclusions

Analysis of the voice source

=> Inverse filtering

How to express the most important features of the obtained waveforms in a compressed way?

=> Parameterisation methods

Time-based parameters:

- Emphasis on measuring the glottal flow features **in the closing phase**

=> Closing Quotient (ClQ) or
Normalized Amplitude Quotient
(NAQ) “recommended” to be used

- NAQ also yields high robustness

Amplitude-based parameters:

- Negative peak amplitude of the flow derivative (d_{peak}) “recommended”
- Parameterisation of the voice source with absolute flow values obtained using the mask is not possible for all the application

Frequency domain methods:

- Measuring the decay of the source spectrum with spectral harmonics is reliable in “most cases”
- Problems:
 - 1) Extremely breathy voices => Upper harmonics noisy

2) Most of the known frequency-domain methods measure the absolute spectral decay

Glottal pulses with low F_0 are of larger dynamics in terms of the spectral decay than those with high F_0

=>

Comparison of glottal flows of female and male speakers calls for normalization of the spectral decay