II. Technical phoniatrics

Kitzing, Peter

Published in:
STL-QPSR (Speech Transmission Laboratory. Quarterly Progress and Status Reports)

1977

B 10

II. TECHNICAL PHONIATRICS

A. ELECTROGLOTTOGRAPH AND CONTACT MICROPHONE FOR MEASURING VOCAL PITCH

A. Askenfelt, J. Gauffin, P. Kitzing x) and J. Sundberg

Abstract

The signal from a contact microphone or an electroglottograph is often more suited for fundamental frequency measurements with simple analogue circuits than the speech signal itself. In this paper a comparison is made between using a contact microphone placed on the neck below the larynx or an electroglottograph for measurement of fundamental frequency in connected speech. The advantages and drawbacks of the two methods are discussed.

Introduction

Two different strategies are generally used for the purpose of measuring the fundamental frequency in speech: low-pass filtering plus zero-crossing detection or double-peak-picking. Both these strategies raise certain demands on the input signal. For zero-crossing measurement the signal wave-form must not have more than one positive or negative zero-crossing in each fundamental period. For the double-peak-picking method there must be only one prominent negative and/or positive peak in each period. These requirements are often hard to fulfill when the speech signal is recorded by a microphone in front of the speaker's mouth.

An electroglottograph offers a signal mirroring the opening and closing of the vocal folds. This is obtained by measuring the variations in a high-frequency current between two surface electrodes placed on each side of the neck at the level of the glottis. The variation in the current is caused by the difference in the electrical impedance of the tissues when the glottis is opened or closed and thus corresponds to the fundamental frequency of phonation. The resulting wave-form generally meets the demands of fundamental frequency detectors, and electroglottography is known as an excellent method for measurements of the fundamental frequency in speech (Fant et al 1966; Fourcin 1974).

However, in some subjects the signal amplitude from the electroglottography may be too low, so that no reliable fundamental frequency measurements can be made. Such problems seem to occur particularly

---

x) Allmänna Sjukhuset, Malmö, Sweden

when the carrier frequency is comparatively low (in the region below 1 MHz) and the current from the electroglottograph is too weak.

An accelerometer microphone, henceforth called contact microphone, which is fastened to the neck below the larynx records a signal related to the sound in the trachea. The wave-form is reasonably independent of the articulation because of the high glottal impedance. Moreover, the trachea resonances are constant in frequency and more damped than those of the vocal tract. As a sequel of this, the wave-form of a contact microphone signal is suitable for fundamental frequency measurements and has therefore often been used for this purpose.

As demonstrated above there are two different methods for obtaining a signal suited for measurement of the fundamental frequency in speech. The purpose of the present investigation was to evaluate the reliability of such measurements based on signals from an electroglottograph and from a contact microphone.

Equipment

The electroglottograph used was constructed by Rundkvist & Kitzing. It has a conventional design (Le Cluse 1977) with a rather low carrier frequency of 200 kHz and an electrode voltage of .5 V. It contains a bistable circuit which is triggered by the signal from the electrodes. The resulting square-wave signal which has a constant amplitude and the same fundamental frequency as the electroglottogram is available at one of the outputs of the device.

The contact microphone (Special Instrument type BC-2) is a simple accelerometer containing a piezo-electric cheramic disc as the pick-up unit. The disc is enclosed in a metal container of 15 mm in diameter and 5 mm in thickness. It weighs about 20 grams. The contact microphone is run with an amplifier with high input impedance (5 M Ω), which is built into the connection wire in order to minimize hum.

The fundamental frequency detector (Fonema type 00063) is of the dual peak sensing type (cf. Larsson 1977). The positive and negative peaks in the input signal wave-form trigger a bistable circuit, so that a square-wave output is obtained having the fundamental frequency of the vibrations sensed by the contact microphone.

Experiment

Simultaneous recordings were made with the two devices described. The electroglottograph electrodes were attached to a fork held against the speaker's thyroid cartilage. The contact microphone was taped to the neck about 2 cm below the cricoid cartilage. The square-wave signal from the electroglottograph and the signal from the contact microphone were recorded on separate channels on tape. The signal from an ordinary microphone located 20 cm in front of the speaker's mouth was recorded on a third channel on the same tape, cf. Fig. II-A-1.

Analysis

A computer program was developed for the analysis. The input signal is the square-wave signal from the fundamental frequency detector. As the square-wave signal from the electroglottograph was distorted by the tape recorder, it had to be restored by means of a flip-flop circuit, before it was sent into the computer. The rectified and smoothed amplitude of the contact microphone signal was simultaneously fed to the computer on a separate line. For the purpose of inspecting the results of the measurement, the program plots the fundamental frequency as a function of time together with the amplitude signal mentioned, cf. Fig. II-A-2.

When no signal or noise is fed into a pitch detector it generates a more or less random output. Therefore, it is essential to exclude the parts of the signal corresponding to unvoiced sounds and pauses. As the amplitude of the contact microphone signal drops substantially when voicing stops, this amplitude was used as a gating criterion. The operator sets an amplitude threshold, and all parts of the fundamental frequency curve, which are associated with intensities below this threshold, are eliminated from all subsequent calculations performed by the computer.

For the fundamental frequency curve a non-linear smoothing algorithm is available (Rabiner et al 1975). The algorithm uses running medians for the smoothing process. The resulting fundamental frequency value, $\overline{f}_n$, of an N-point median smoother, where N is an odd integer number, represents the centermost value of a sequence of the N adjacent samples:

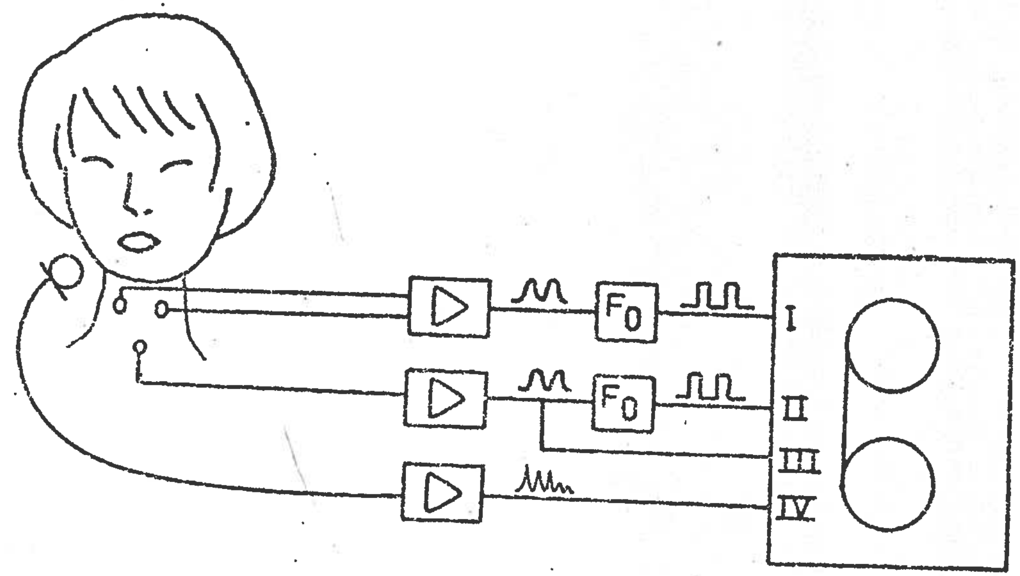

Fig. II-A-1. Schematic diagram of the data-recording procedure.

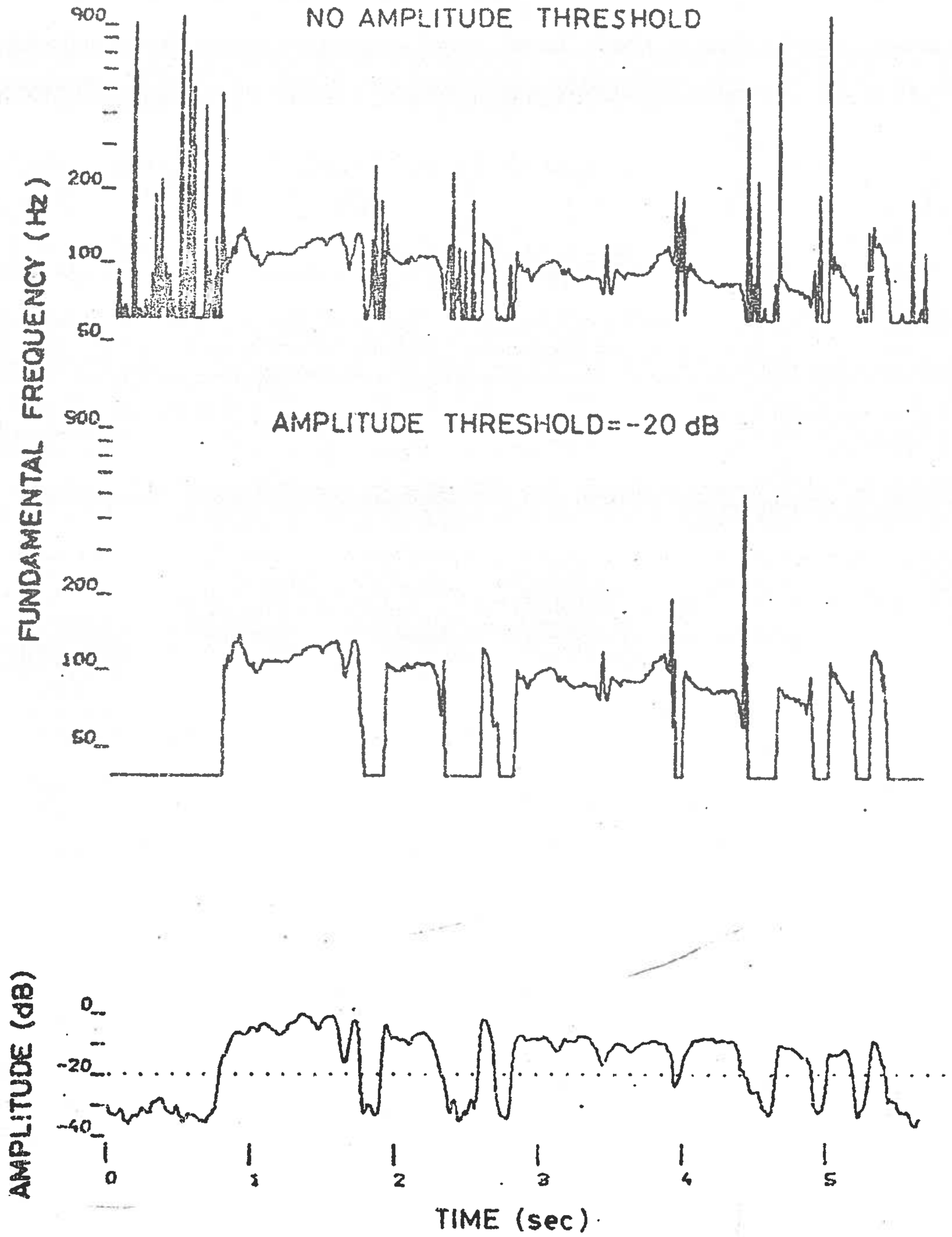

Fig. II-A-2. Fundamental frequency and amplitude curves plotted by the computer. The dotted line in the lowest curve represents the amplitude threshold set by the operator.

$$f_{n-\frac{N-1}{2}}, \ldots\ldots f_{n-1}, f_n, f_{n+1}, \ldots\ldots f_{n+\frac{N+1}{2}}. \quad (1)$$

Running medians have the property of preserving sharp discontinuities with durations exceeding a critical value, which depends on the size N of the median smoother. Discontinuities with shorter durations are eliminated. This smoothing is useful for eliminating some types of errors in the fundamental frequency curve, cf. Fig. II-A-3. When not otherwise stated, it was not used in the data presented below.

The fundamental frequency values are assorted in frequency classes 1 Hz wide. The resulting fundamental frequency histogram is plotted on the computer screen. By means of a cursor the operator defines lim 1 and lim 2 being the lower and upper limits of the relevant frequency range. Histogram contours may be smoothed by an algorithm. If so, each point F_n along the histogram contour is modified according to:

$$F_n' = \frac{F_{n-1}+2F_n+F_{n+1}}{4} \quad (2)$$

This smoothing can be repeated several times.

The computer program also calculates some statistics and presents the results in a new plot, cf. Fig. II-A-4. The statistical values are:

(1) the arithmetic mean of the distribution (MF_o),

(2) the mode of the distribution (F_{max}),

(3) the ratio between the parts of the distribution lying left and right of the mode (BIAS),

(4) the total sample time defined as the sum of the period times considered in the calculations (TIME).

Also, a triangular approximation of the distribution is presented. Its top is identical with the mode, and the areas on both sides equal the corresponding areas under the histogram contour.

Fundamental frequency histograms have been found useful for phoniatrical purposes (Kitzing et al 1975; Fritzell et al 1977). Therefore it is relevant to study the comparability of such histograms when derived from electroglottograph and contact microphone recordings made simultaneously. The entire material of 13 voices was thus compared using

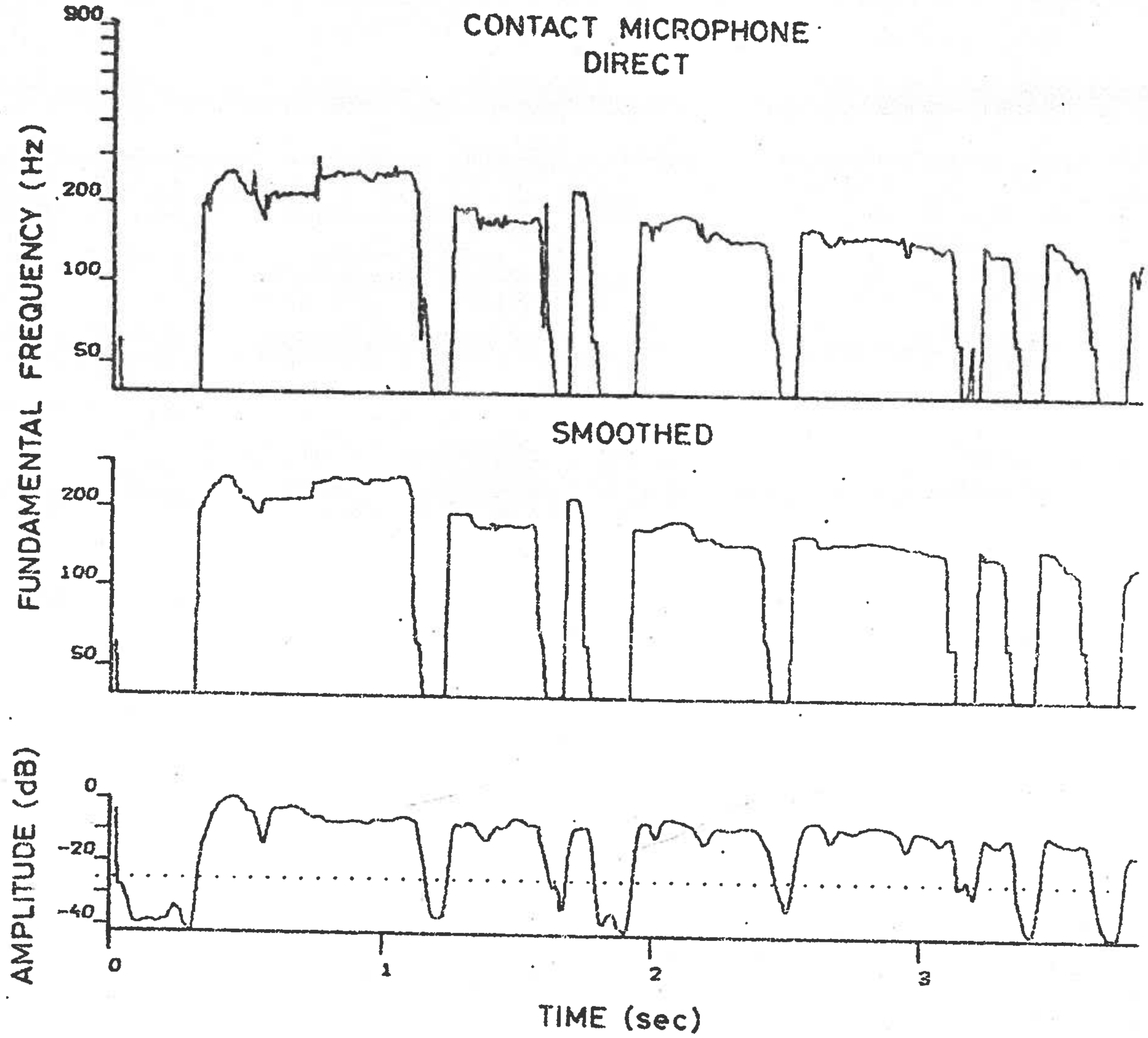

Fig. II-A-3. Fundamental frequency curves, before and after smoothing, together with the amplitude curve. All curves were plotted by the computer.

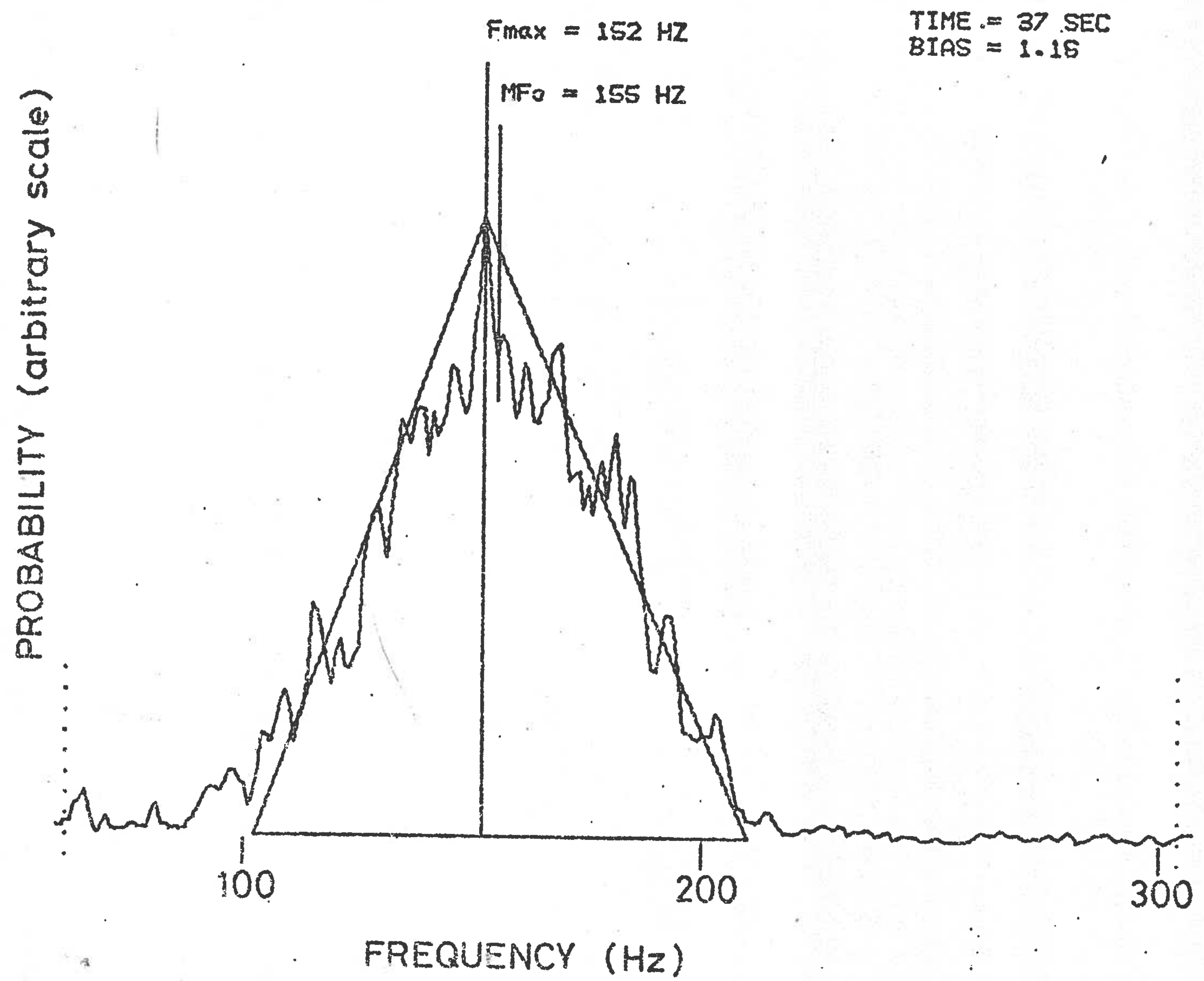

Fig. II-A-4. Fundamental frequency histogram plotted by the computer. The vertical dotted lines are set by the operator and limit the frequency range used in the calculations.

the computer program described above. Thereby, special care was taken so as to base the histograms on exactly the same sample of speech sounds. This was realized by using the amplitude of the contact microphone signal as the gating criterion in both cases, and the lim 1 and lim 2 were set at identical frequencies in both analyses. All histograms were smoothed twice, cf. Eq. (2).

Results

Table II-A-I shows the results obtained from these measurements. As regards the average of the distribution there is an almost perfect agreement between the methods. The average difference between the values based on contact microphone and electroglottograph recordings is not greater than 1.0 Hz (SD = .82 Hz). Eventhough this value is very small, most differences are positive indicating that the mean obtained from the contact microphone is slightly higher.

As regards the modes of the histograms greater differences are observed as seen in the same table. This occurs especially in histograms which exhibit several high peaks of almost the same amplitude. Then a small difference in the measurements may easily make another peak the highest one. Even if the average difference is as small as - .7 Hz (SD = 5.96 Hz) the value of this measure can be questioned. Probably a more smoothed histogram would yield a more useful value of the distribution.

The sample time shows very small differences. This is another indication that the differences in the data obtained by the two methods are very small. The BIAS measure, on the other hand, suffers from the discrepancies in the mode values because the BIAS is strongly influenced by a change of the mode. Still, the average discrepancy is as small as .16 (SD = .70). If future research will support the assumption that the skewness of the distribution is a relevant measure, it seems that a more sophisticated measure of the skewness should be used.

Discussion

As mentioned above there are some differences in the signals recorded by a contact microphone and an electroglottograph. As mentioned

Subject	Sex	Mean F_o Hz CM	Mean F_o Hz EKG	Δ	Mode F_o Hz CM	Mode F_o Hz EIG	Δ	Time sec CM	Time sec EIG	Δ	Bias CM	Bias EIG	Δ
GT	fem.	222	221	1	204	206	-2	22	21	1	2.13	1.95	.18
IT	fem.	193	193	0	161	160	1	25	24	1	3.92	3.59	.33
EA	fem.	172	171	1	149	149	0	20	19	1	2.61	2.27	.34
MR	fem.	164	(160)	(4)	120	(119)	(-1)	17	(17)	0	3.71	(3.16)	
SF	fem.	166	-	-	146	-	-	20			1.95		
MK	male	108	106	2	99	100	-1	38	37	1	1.62	1.43	.19
JS	male	96	94	2	77	84	-7	21	21	0	2.33	1.35	.98
BG	male	106	105	1	104	104	0	24	23	1	.86	.86	0
LA	male	95	93	2	84	85	-1	26	25	1	2.19	1.58	.61
EJ	male	104	103	1	98	88	10	30	28	2	1.13	2.45	-1.32
ASL	male	106	106	0	100	112	-12	23	21	2	1.49	0.58	0.91
PK	male	96	96	0	90	85	5	22	22	0	1.11	1.78	-.67
SR	male	155	-	-	152	-	-	37			1.16		
Average				1.0			-.7			1.0			0.16
SD				.82			5.96			.67			0.70

TABLE II-A-I.

the contact microphone senses the vibrations is the wall of the neck. These vibrations reflect variations of the subglottic pressure accomplished by the vibrating vocal folds acting as a sound generator. Such pressure variations will be generated even in cases of incomplete glottal closure, such as in leaky phonation (Kitzing & Löfqvist 1975). On the other hand, the amplitude of the electroglottograph signal seems to be strongly dependent on whether or not the vocal folds make contact, cf. Fig. II-A-5a, b, and c. This dissimilarity will cause differences in fundamental frequency data obtained from the two devices.

Another difference is that the electroglottograph cannot be used on all subjects, as mentioned previously. A thick layer of subcutaneous tissue in the neck may decrease the signal amplitude substantially. This is not true in the case of the contact microphone. Three subjects, which could not be measured with the electroglottograph in our experiment, produced seemingly reliable fundamental frequency data when measured by means of the contact microphone. (It should be stressed, however, that the ratio 10:3 is probably not representative for the percentage of voices that cannot be measured with an electroglottograph. In selecting the subjects we attempted to choose voices which could be expected to cause problems for the devices tested.)

A third difference is that the amplitude of the electroglottograph signal reaches its full value sooner than that of the contact microphone in voicing onsets, cf. Fig. II-A-5d. The fundamental frequency meter used is somewhat sensitive to amplitude variations and this can be a cause to some errors in the measurements. By using a more sophisticated fundamental frequency detector these errors may be avoided.

A fourth difference is that the electroglottograph wave-form is almost independent of the articulation. At least it was generally very easy to process with the fundamental frequency detector. The contact microphone wave-form slightly depends on the articulation because of the finite glottal impedance. In other words, the subglottic sound is influenced by the supraglottic sound. The resulting changes in wave-form often cause errors, cf. Fig. II-A-6. Most such errors can be eliminated with the running median smoother, as can be seen in the same figure.

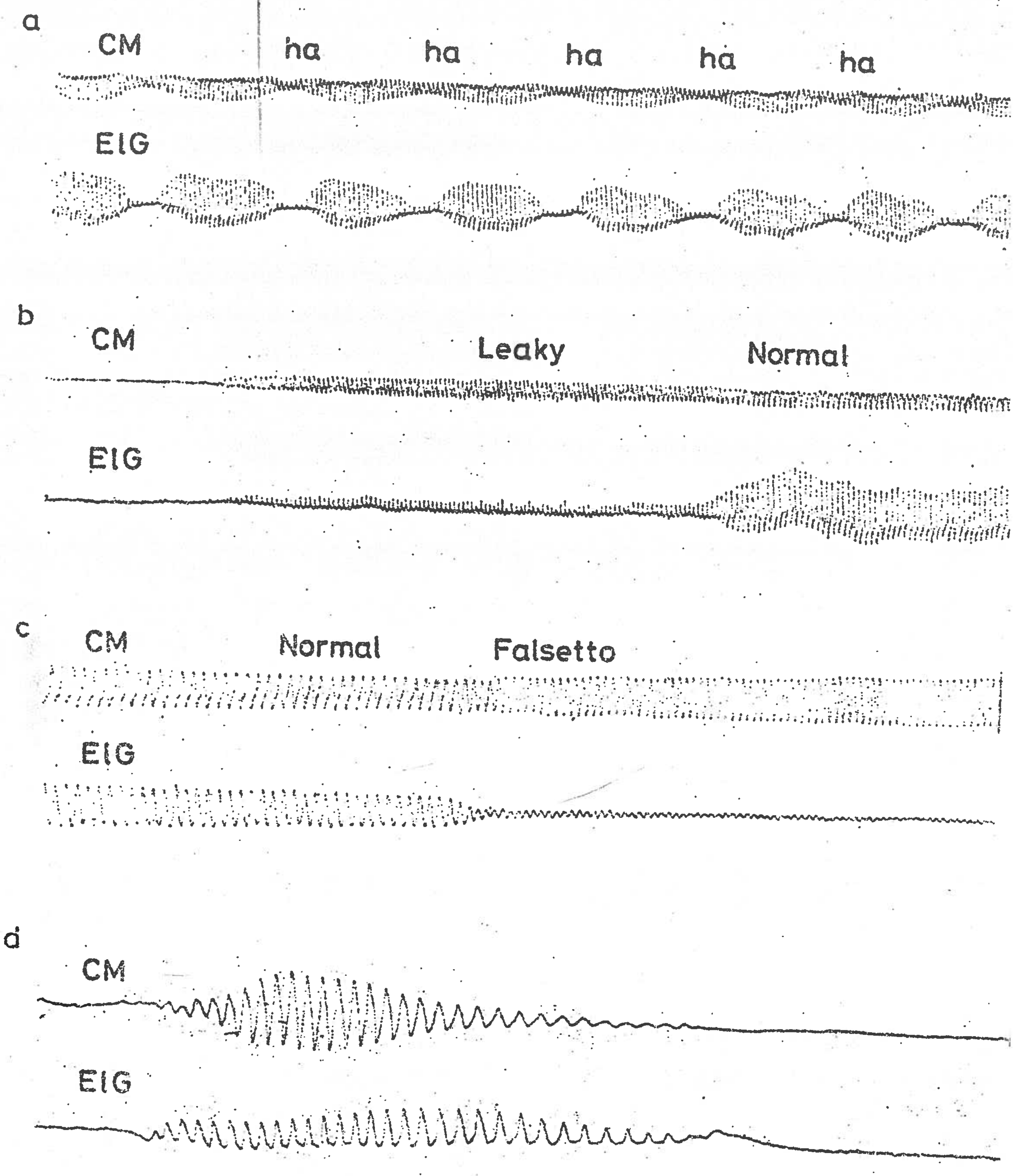

Fig. II-A-5. Comparisons between contact microphone signals (CM) and electroglottograph signals (ElG) for different types of phonation.

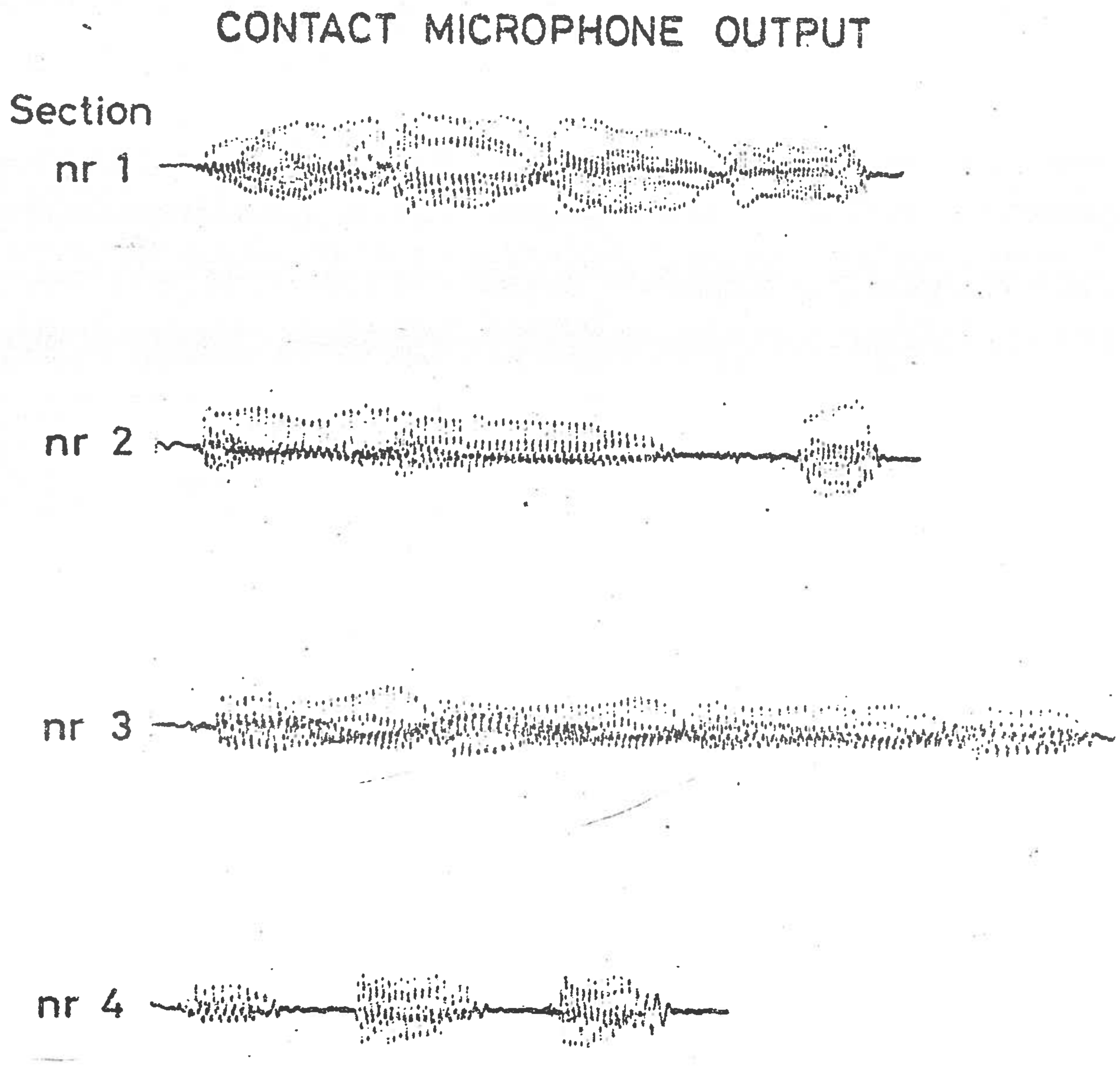

Fig. II-A-6a. The signal from the contact microphone corresponding to the curves in Fig. II-A-6b.

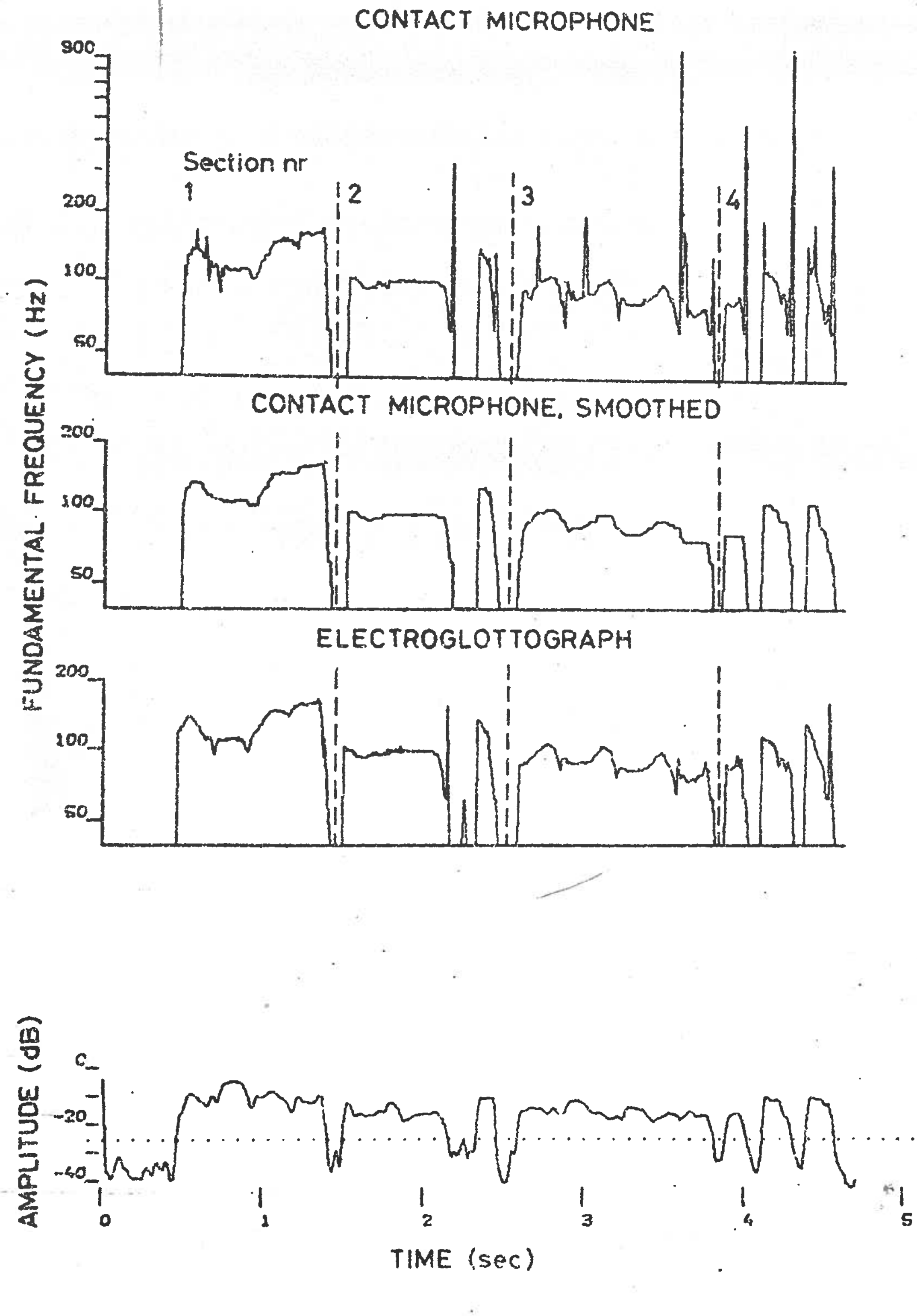

Fig. II-A-6b. Fundamental frequency curves derived from the contact microphone signal and the electroglottograph together with the contact microphone signal amplitude.

A fifth difference between the two devices is their behavior in cases of vocal fry. Here, the vocal folds do not close regularly, so that the electroglottograph sometimes gives lower frequency readings than the contact microphone, cf. Fig. II-A-7. However, for very low-frequency vocal fry phonation the electroglottograph may not always register correctly the instances of glottal closure. This is because the electroglottograph is also sensitive to larynx movements and changes in the constriction of the larynx. These movements superimpose low-frequency signals on the glottogram and even if a HP-filter is used errors may occur in triggering the circuit delivering the square-wave output.

Summarizing one may say that there are several reasons for expecting slightly differing results in fundamental frequency data obtained from an electroglottograph and from a contact microphone: the electroglottograph normally depicts the frequency of glottal closures, while a contact microphone also mirrors the sound generated by the vocal fold vibrations. Moreover, the variability in the contact microphone amplitude and wave-form cause spurious errors in the fundamental frequency measurement. Therefore, for the purpose of the examining fundamental frequency events in speech, an electroglottograph should be tried in the first place.

Conclusions

Both equipments examined in the present investigation have strong and weak points. As for the electroglottograph it rarely fails when it functions at all: those voices that can be measured are generally measured accurately. It records the glottal closures rather than the sound generated by the vibrating vocal folds. The contact microphone, on the other hand, can probably be used on all subjects, even in cases when the electroglottograph fails. The output signal should be LP-filtered in order to provide a signal suited for fundamental frequency measurements. The signal mirrors the sound generated by the vocal fold vibrations. There is no exact agreement between the fundamental frequency data obtained from the two devices. This is a consequence mainly of the fact that variations in the contact microphone signal

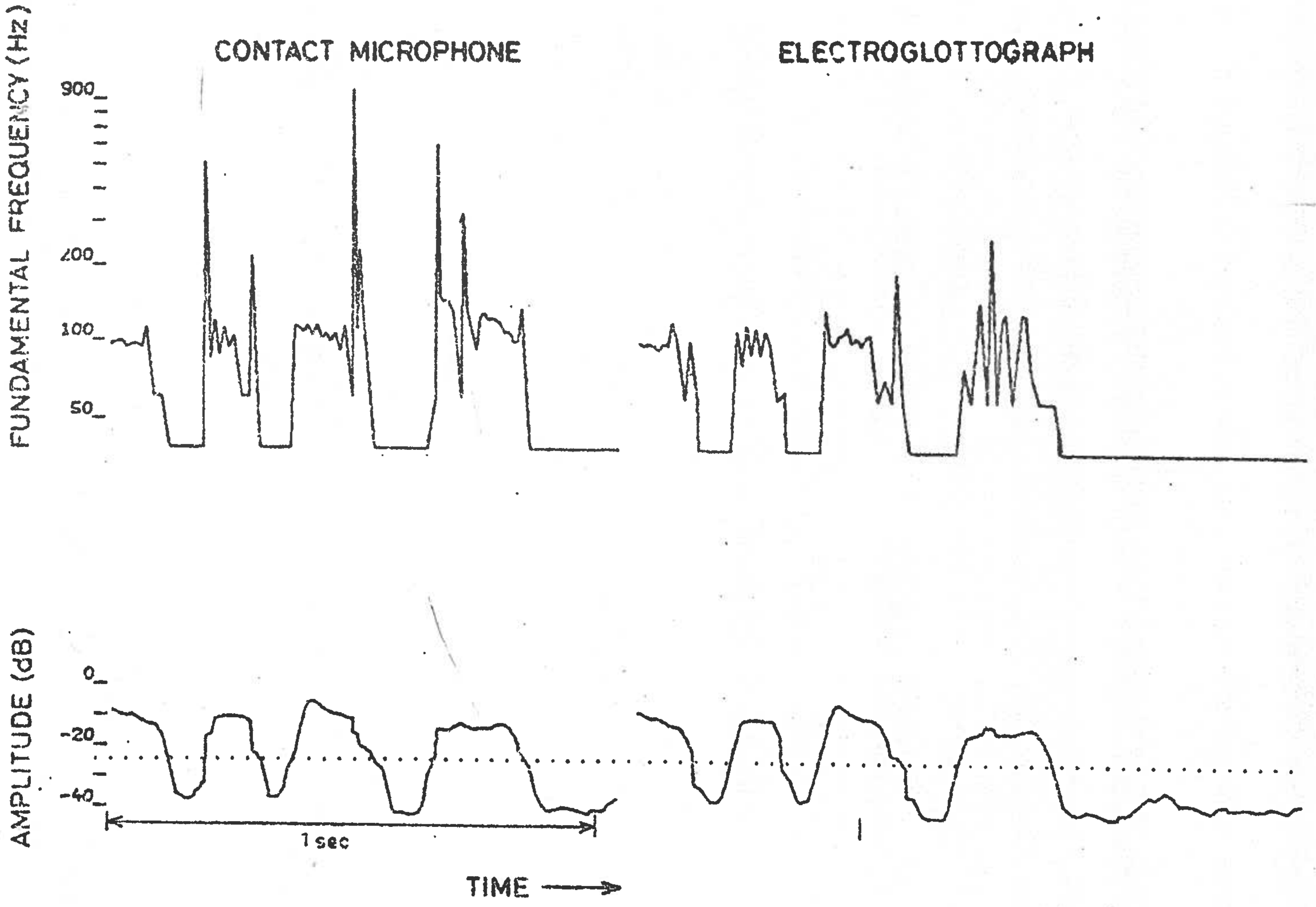

Fig. II-A-7. Comparison between fundamental frequency curves as measured by contact microphone and electroglottograph during vocal fry phonation.

wave-form cause errors in the fundamental frequency measurements. However, such errors can be easily eliminated by a running median smoother. Also, the frequency of the vocal fold closures is not invariably identical with the fundamental frequency of the vibrations in the wall of the neck. It depends on how the fundamental frequency is measured and defined. For example, the fundamental frequency of the sound differs from the glottal closure frequency in certain instances of vocal fry. However, when used for building up fundamental frequency histograms the differences in the data obtained by the two devices are negligible for practical purposes.

For the practical situation in the phoniatric clinic it seems that the electroglottograph should be preferred as long as it works, and replaced by the contact microphone in remaining subjects.

Acknowledgments

This work was supported by the Bank of Sweden Tercentenary Foundation under Contract no. 67/48.

References

Le CLUSE, F. (1977): "Elektroglottografie", Academical Thesis, Utrecht.

FANT, G., ONDRÁČKOVÁ, J., LINDQVIST, J. & SONESSON, B. (1966): "Electrical glottography", STL-QPSR 4/1966, pp. 15-21.

FOURCIN, A.J. (1974): "Laryngographic examination of the vocal fold vibration", Ventilatory and phonatory control mechanisms, Ed. B. Wyke, Oxford University Press.

FRITZELL, B., GAUFFIN, J., HAMMARBERG, B., SUNDBERG, J., & WEDIN, L.(1977): "Clinical applications of acoustic voice analysis", STL-QPSR 2-3/1977, pp. 31-43.

KITZING, P. & LÖFQVIST, A.(1975): "Subglottal and oral airpressures during phonation", Medical and Biological Eng., Sept., pp. 644-648.

KITZING, P., RUNDQVIST, H-E. & TALO, E. (1975): "Fundamental frequency of the voice in continuous speech", Working Papers 12, Phon. Lab., Lund University, pp. 85-98.

LARSSON, B.(1977): "Pitch tracking in music signals", STL-QPSR 4/1977, pp. 1-7.

RABINER, L., SAMBER, M. & SCHMIDT, C. (1975): "Application of nonlinear smoothing algorithm to speech processing", IEEE Transact. Acoust., Speech, Signal Processing, Vol. ASSP-23, Dec., pp. 552-557.