working papers in phonetics 4

university of california, los angeles

july 1966

A New Photo-Electric Glottograph

John Ohala

This is a report of our work on a device which is being developed for use in research on laryngeal activity during speech. The device is conceptually similar to the apparatus used by Sonesson (1959, 1960) in his important work on the vibratory pattern of the vocal folds (Cf. Zerlin, 1959). We shall follow Sonesson and call this device a photo-electric (diaphanoscopic) glottograph. (Cf. the terminology of Fabre (1957, 1958, 1959) who calls his apparatus a glottographe de haute frequence.)

Sonesson's apparatus consisted of a strong DC light source place against the neck just below the larynx, the light from which suffused the subglottal region and transilluminated the glottis. A curved inflexible Lucite "light pipe" was passed through the mouth and into the throat and terminated at about the level of the epiglottis. It was aimed at the glottis and transmitted the light coming through the glottal chink to a light sensor outside the mouth. The light sensor transduced the variations in light intensity into variations in electrical voltage, thus providing a direct measure of the degree of opening and closing of the glottis. These variations in electrical voltage could be recorded in a variety of standard ways and then analyzed. Sonesson limited his investigations to steady-state vowels produced with varying pitch and loudness. Due to the rigid light pipe in the subject's mouth, studies of vocal cord vibrations during connected speech would have been extremely difficult if not wholly impossible.

An improvement on Sonesson's technique which immediately suggested itself, then, was inserting the photo-electric pick-up into the throat *via* the nose so that it would not interfere with the normal speech articulations of the tongue and lips, thus permitting an examination of the vocal folds' activity during connected speech. Thanks to the extreme miniaturization of electronic components these days, this is easily accomplished.

In a flexible transparent plastic catheter (4mm. O.D.), sealed at one end, a small photo-resistive diode was encased so that it was about 25 cm from the sealed end. (The light sensor was a Texas Instruments LS-400; 2mm diameter; 13mm long.) The leads from the light sensor extended through the catheter and emerged out the open end. The catheter was inserted into the throat and esophagus *via* the nasal passage so that the light sensor was about 15-16 cm from the external nares. The extra length of catheter preceding the light sensor and extending into the esophagus not only helps to stabilize the position of the light sensor relative to the glottis, but also adds immeasureably to the comfort of the subject, since a loose tube dangling in the pharynx tends to trigger a gag reflex. With this feature plus the catheter's pliancy and thinness we obtained a device that caused no discomfort to the subjects and after an initial 5 minute period following insertion caused no disruption of the natural speech articulations. Tape recordings of subjects speaking with and without the catheter in their throats were indistinguishable.

An approach similar to Sonesson's was used by members of the Haskins group before 1961 (Cooper 1964). They reversed the position of light source and photo-electric pick-up relative to the glottis, having the light transmitted to the supraglottal laryngeal cavity by a thin fiber optic inserted through the mouth, the light variations being read by a photo-cell placed against the trachea just below the larynx. As early as 1964 they mentioned the desirability of inserting the fiber optic through the nasal passage (Lisker and Abramson, 1964) and achieved this in 1966 (Abramson, 1965; Lisker, et al., 1966). A full description of the Haskins' apparatus has not yet been published. They do report some difficulty in obtaining a comfortable velo-pharyngeal closure.

The circuitry we used is very simple: a 6v power supply, a resistor and the light sensor are wired in series, as shown in Figure A. The manufacturer indicates the light sensor's frequency response as flat ± 2 dB. from DC to 90 kHz., with a 1000 ohm resistor. We found that by increasing the resistance we could enhance the light sensor's sensitivity at the expense of the frequency response. We therefore introduced a switch with which we can choose one of two values of resistance depending on which characteristic of the sensor we are interested in most at the time.

At maximum sensitivity, using a 68K ohm resistor, the frequency response is flat ± 2dB. from DC to 1200 Hz; with a 6.8K ohm resistor, the frequency response is extended to 6000 Hz, with an approximately 10 dB drop in sensitivity relative to that obtained with the larger resistance. The only other aspect of the electrical system worth mentioning is that due to the relatively weak signals obtained (with our apparatus, seldom more than 10 mv.) all external leads must be shielded and the subject grounded.

The light source is a 100 watt incandescent lamp in an ordinary spotlight housing with a 3" plano-convex lens; it is powered by a 120 v DC power supply. The spotlight does not generate an uncomfortable amount of heat upon the skin. Rubbing the throat beforehand with glycerine or Vaseline seems to be adequate for the subject's comfort. In order that the position of the glottis remain in as constant a position as possible relative to the spot on the neck where the light is shining, the subject's head and the light must both be fixed. Even with this precaution, though, the larynx still moves somewhat; changes in pitch, changes in the pressure differential across the glottis, changes in tongue position, etc., all involve movements of the larynx. There is also some movement of the light sensor relative to the glottis. This is due mainly to the action of the velum upon which the catheter rests. We know of no simple way of eliminating this. For the most part it is not a serious problem, merely registering the glottal openings with decreased amplitude.

The types of records produced by the glottograph are illustrated in Figures 1-7. All were made of one subject, a 24 year old male speaker of the Midwestern dialect of American English. A glottogram is a measure of glottal area as a function of time, evolving from left to right. Thus a positive-sloped line represents an opening movement of the vocal cords, and a negative-sloped line, a closing movement. In these figures there is no absolute calibration of the amplitude of the glottograms, i.e., we have not attempted to quantify any given point on the ordinate with respect to a particular glottal area. However, in any given glottogram, the relative amplitude levels are consistent.

The glottogram on the cover and Figures 1 through 6 were

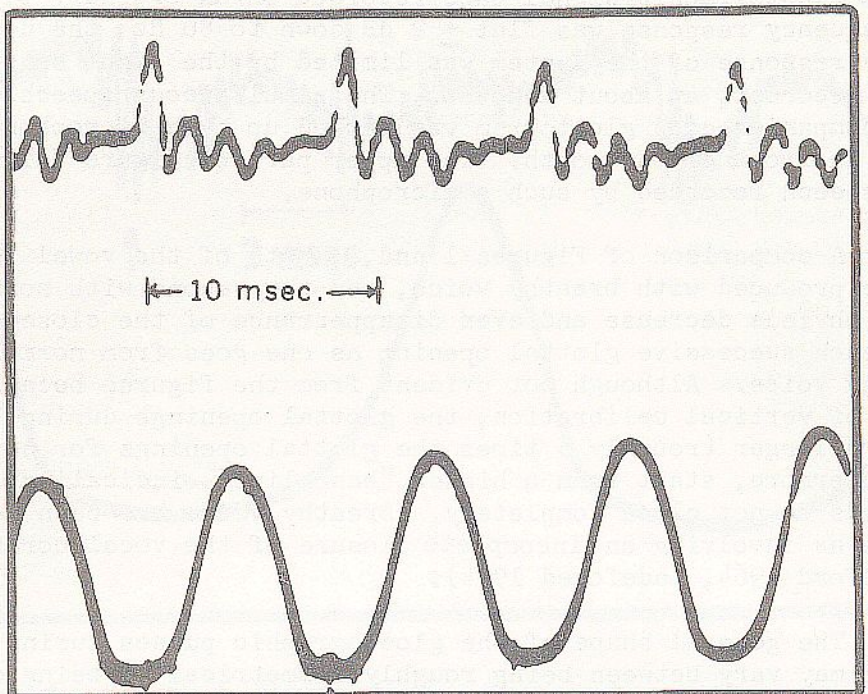


Figure 1. Glottogram (bottom) and simultaneous microphone signal (top) of vowel [ɑ] produced with breathy voice; F_o = 122 Hz.

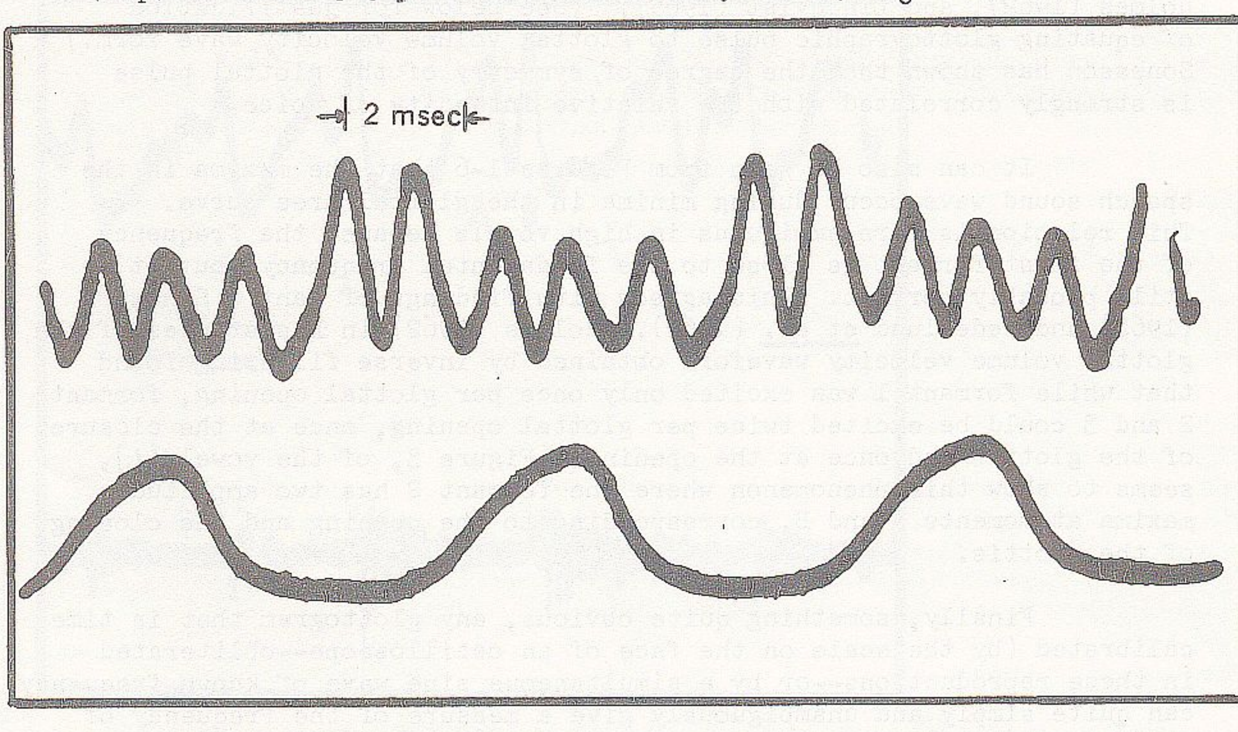


Figure 2. Glottogram (bottom) and microphone signal (top) of vowel [ɑ] produced with normal voice; F_o = 142 Hz.

photographed from the screen of an oscilloscope. Figure 7 is a **spectrogram** made of a glottographic signal tape recorded on an ordinary tape recorder whose frequency response was flat ± 2 dB down to 50 Hz; the upper frequency response of the system was limited by the light sensor, not the tape recorder, at about 1500 Hz. The simultaneous speech sound wave which accompanies each glottogram was picked up by a microphone placed 10-15 cm in front of the mouth. The upper part of Figure 7 is a spectrogram of speech recorded by such a microphone.

A comparison of Figures 1 and 2, both of the vowel [ɑ], the first one produced with breathy voice, and the second with normal voice, shows an obvious decrease and even disappearance of the closed period between each successive glottal opening as one goes from normal voice to breathy voice. Although not evident from the figures because of the lack of vertical calibration, the glottal openings during breathy voice were larger (roughly 5 times the glottal openings for normal voice) and, furthermore, start from a higher "base line"--indicating that the vocal cords do not close completely. Breathy voice has been described elsewhere as involving an incomplete closure of the vocal cords (Moore 1962, Catford 1964, Ladefoged 1964).

The general shape of the glottographic pulses during sustained phonation may vary between being roughly symmetrical to being quite asymmetrical, with the closing phase being shorter and more abrupt than the opening phase. This latter shape of the glottal pulse, quite typical for normal voice, has been reported in or illustrated in Sonesson (1960), Cederlund, et al. (1960), Fant & Sonesson (1962), Miller (1959), Holmes (1962), and Lindqvist (1965). (See below for justification of equating glottographic pulse to glottal volume velocity wave form.) Sonesson has shown that the degree of symmetry of the glottal pulse is strongly correlated with the relative intensity of voice.

It can also be seen from Figures 1-6 that the maxima in the speech sound wave occur during minima in the glottal area curve. This relation is more ambiguous in high vowels because the frequency of the first formant is close to the fundamental frequency, but it is still probably correct. This agrees with findings of Fant & Sonesson (1962) and Cederlund et al. (1960). Holmes (1962) in his studies of the glottal volume velocity waveform obtained by inverse filtering found that while formant 1 was excited only once per glottal opening, formants 2 and 3 could be excited twice per glottal opening, once at the closure of the glottis and once at the opening. Figure 3, of the vowel [i], seems to show this phenomenon where the formant 2 has two amplitude maxima at moments A and B, corresponding to the opening and the closing of the glottis.

Finally, something quite obvious, any glottogram that is time calibrated (by the scale on the face of an oscilloscope--obliterated in these reproductions--or by a simultaneous sine wave of known frequency) can quite simply and unambiguously give a measure of the frequency of voice, not only during vowels, but during nasals and voiced fricatives as well. It should be a trivial point to note that the frequency of the glottal vibrations equals the frequency of the sound emitted from the mouth; however Fabre (1957) shows some confusion about this. Noting that when his subject attempted to match his voice frequency to that of a pure tone, the frequency of his vocal cord vibrations was one

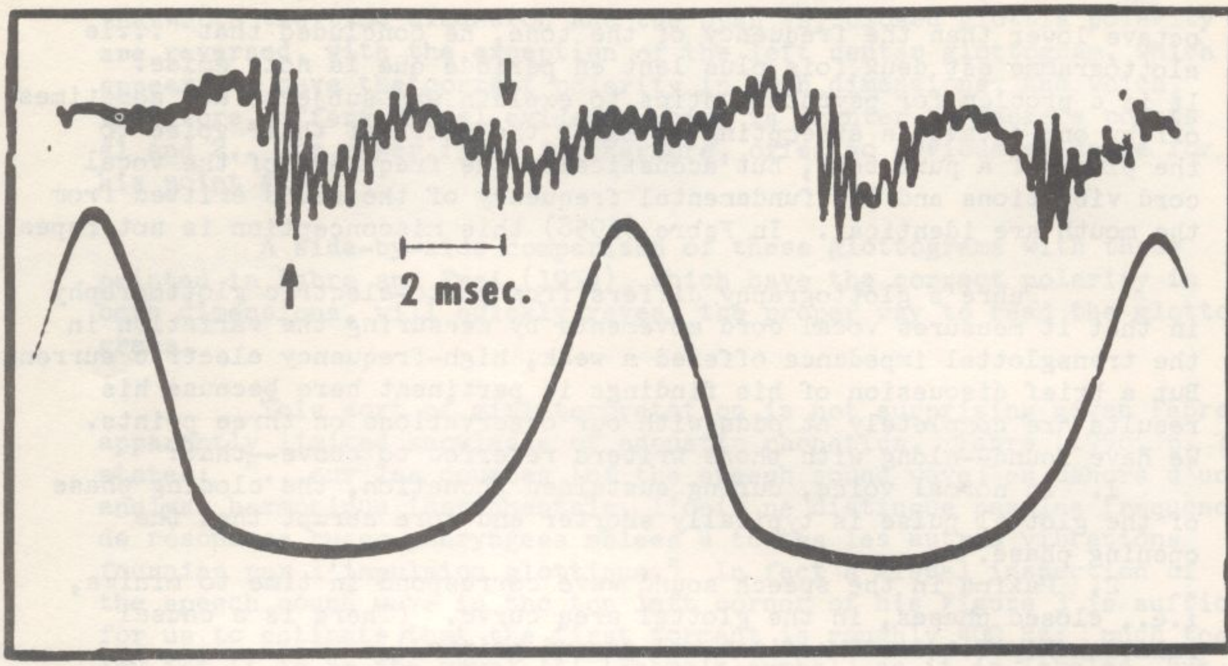


Figure 3. (Retouched) Glottogram (bottom) and microphone signal of vowel [i]; normal voice; $F_o = 94$ Hz.

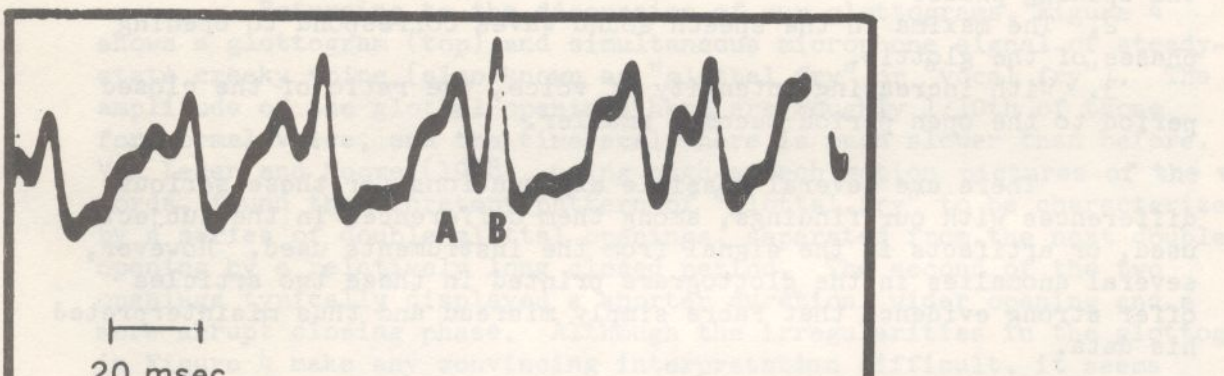


Figure 4. Glottogram (top) and microphone signal of steady-state "creaky voice" (also known as "glottal fry" or "vocal fry").

octave lower than the frequency of the tone, he concluded that "...le glottogramme est deux fois plus lent en période que la note émise." It is a problem for psychoacoustics to explain why subjects are sometimes off by one octave in attempting to match the pitch of their voice to the pitch of a pure tone, but acoustically the frequency of the vocal cord vibrations and the fundamental frequency of the sound emitted from the mouth are identical. In Fabre (1958) this misconception is not repeated.

Fabre's glottography differs from photo-electric glottography in that it measures vocal cord movements by measuring the variation in the transglottal impedance offered a weak, high-frequency electric current. But a brief discussion of his findings is pertinent here because his results are completely at odds with our observations on three points. We have found--along with those writers referred to above--that:

1. In normal voice, during sustained phonation, the closing phase of the glottal pulse is typically shorter and more abrupt than the opening phase.

2. Maxima in the speech sound wave correspond in time to minima, i.e., closed phases, in the glottal area curve. (There is a causal relationship between the first and second points.)

3. With increasing intensity of voice, the ratio of closed period to open period becomes larger. (cf. Sonesson, 1960, etc.)

Fabre (1957, 1958) finds the complete opposite:

1. The closing phase is typically longer and more gradual than the opening phase.

2. The maxima in the speech sound waves correspond to opening phases of the glottis.

3. With increasing intensity of voice, the ratio of the closed period to the open period becomes smaller.

There are several possible explanations for these serious differences with our findings, among them differences in the subjects used, or artifacts in the signal from the instruments used. However, several anomalies in the glottograms printed in these two articles offer strong evidence that Fabre simply misread and thus misinterpreted his data.

A casual inspection of the illustrations published in Fabre (1958) will show conclusively that all the oscillographic displays (glottograms plus simultaneous speech sound wave) in his Figure 1 were printed with the time dimension reversed (the sound waves decay from right to left as they are printed). With the resulting re-interpretation of these six glottograms, point #1 of Fabre (closing phase slower than opening phase) is thus reconciled to our findings (opening phase slower than closing phase). These glottograms do not offer the evidence Fabre claims they do for his point #2, and there is no variation in intensity represented so point #3 is not involved.

In his Figure 2 in the same article, which is identical to Figure 1 in the 1957 article, the time dimension is correct, but Fabre misread the open vs. closed glottis polarity, i.e., what he labeled "open glottis" is really closed glottis and vice versa. With the resulting re-interpretation of these glottograms they agree with our points #1, 2, and 3.

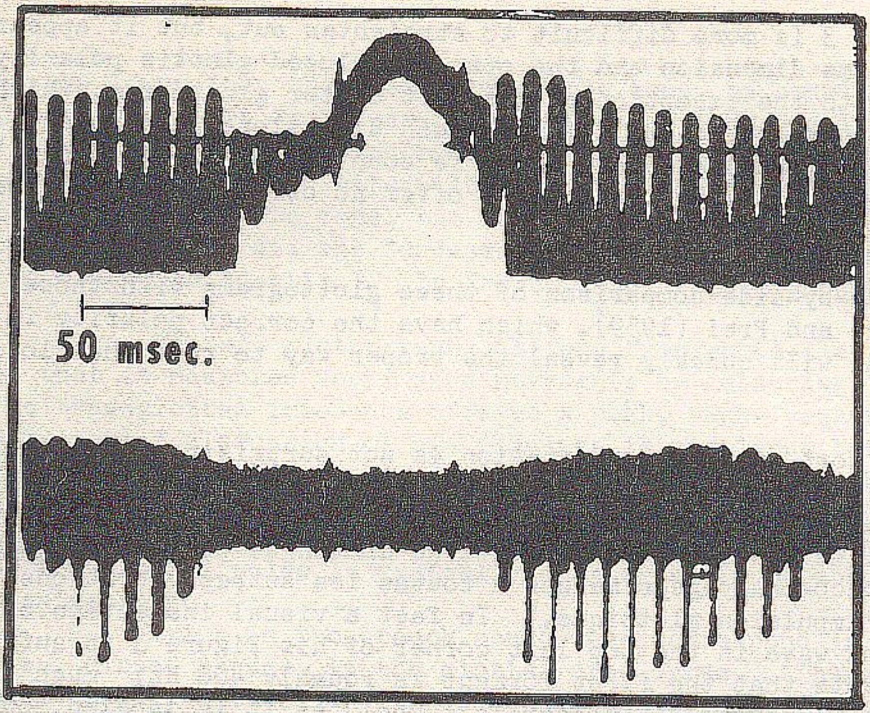
Figure 3 is more difficult to straighten out, but it appears that both the time dimension and the open vs. closed glottis polarity are reversed, with the exception of the left center glottogram, which appears to have the correct polarity in both dimensions, and which, therefore, offers visual evidence which is counter to Fabre's points #1 and 2. The other five, furthermore, offer no obvious evidence for his point #1.

A side-by-side comparison of these glottograms with those printed in Fabre and Frei (1959), which have the correct polarity in both dimensions, will quickly reveal the proper way to read the glottograms.

This sort of misinterpretation is not surprising given Fabre's apparently limited knowledge of acoustic phonetics. Fabre (1958, p. 773) states: "...sur les courbes [of the speech sound wave] en dehors d'une analyse harmonique instrumentale, l'oeil ne distingue pas les fréquences de résonance bucco-pharyngees mêlées a toutes les autres vibrations fournies par l'impulsion glottique." In fact a visual inspection of the speech sound wave in the top left corner of his Figure 3 is sufficient for us to estimate that the first formant is roughly 400 Hz: much too low for it to be the vowel 'A' [Fabre's symbol] as it is labeled, but is more appropriate for the vowel 'E'. The top four speech sound waves in his Figure 3 are thus mislabeled as to what vowel they represent. Furthermore, the article by Fabre and Frei (1959) abounds with misinformation on acoustic phonetics.

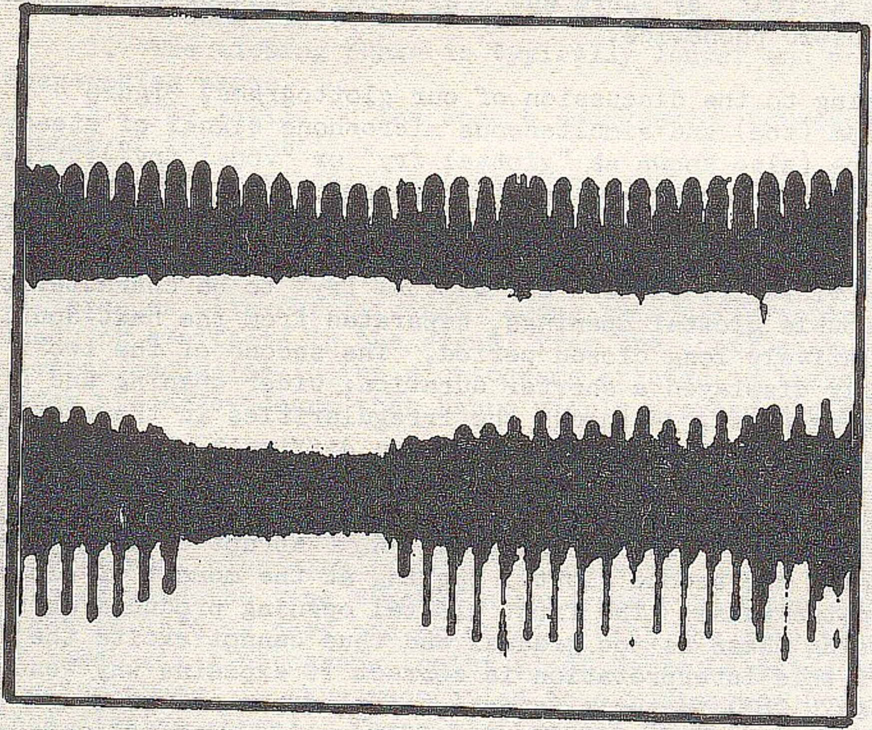
Returning to the discussion of our glottograms, Figure 4 shows a glottogram (top) and simultaneous microphone signal of steady-state creaky voice (also known as "glottal fry" or "vocal fry"). The amplitude of the glottal openings here are roughly 1/10th of those for normal voice, and the time scale here is much slower than before. Von Leden and Moore (1958), using high-speech motion pictures of the vocal cords, found the vibratory pattern of "glottal fry" to be characterized by a series of double glottal openings, separated from the next double opening by a relatively long closed period. The second of the two openings typically displayed a shorter duration, wider opening and a more abrupt closing phase. Although the irregularities in the glottogram in Figure 4 make any convincing interpretation difficult, it seems likely that the glottal openings marked A and B represent just such a double opening as von Leden and Moore reported. The glottal opening B has a shorter duration, wider opening and a more abrupt closing phase. Interestingly, only the second glottal opening at the moment of closure produces an acoustic pulse. The first glottal opening produces no sound because its closing phase is too gradual and thus cannot excite the vocal tract. If this interpretation is correct it explains why Wendahl, Moore and Hollien (1963) in their investigation of the acoustic and perceptual correlates of "vocal fry", found no evidence that double pulses were essential characteristics of "vocal fry". The double glottal openings do indeed exist in creaky voice, it seems, but the first of the pair of openings need not have an acoustic correlate. We often observed three glottal openings in quick succession, but again, only the last one at the moment of closure produced a sound pulse.

In Figure 5 is shown a glottogram (with greatly reduced time scale) of the sequence /a'pa/ abstracted from the larger sequence



(Retouched) Glottogram (top) and microphone signal of [ɑ'pɑ].

Figure 5



(Retouched) Glottogram (top) and microphone signal of [ɑ'bɑ]; same time scale as Figure 5.

Figure 6

/'pa'pa'pa'pa.../. For a /p/ in this environment it is clear that during the voiceless period, i.e., the period of non-vibration, the vocal cords move apart such that the maximum glottal area is greater than the maximum glottal area during vibration. The absolute size of this maximum area seems to vary as a function of the tempo of speech, being larger for a slow rate of speaking. In the transition from vibrating state to non-vibration, and vice versa, there are some vibrations of the vocal cords which are not picked up by the microphone. This could be due to the acoustic impedence of the lips which would drastically reduce the amplitude of the sounds arriving at a microphone in front of the mouth, or because these vibrations are too gradual and do not cause a pressure change large enough and rapid enough to excite the vocal tract (cf. Lisker & Abramson 1964).

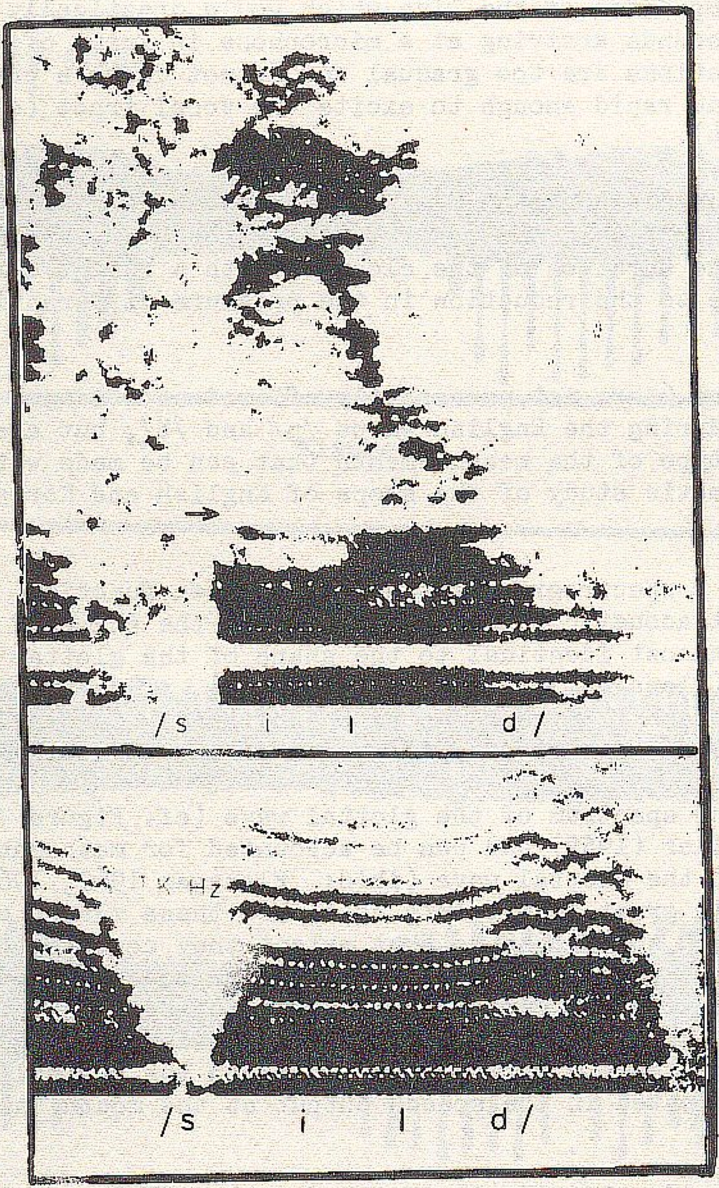
Figure 6 presents a glottogram for the sequence /a'ba/ taken from the larger sequence /'ba'ba'ba.../. In this case the vibrations continue throughout the duration of the closure, with a slight decrease in their amplitude due to the reduction in the pressure differential across the glottis.

Figures 5 and 6 are not necessarily offered as evidence on the state of the glottis during the English stops /p/ and /b/, but are merely illustrations of the type of the measurements that can be made with the glottograph. A systematic study of the stops of English and Korean is now under way.

Figure 7 is a spectrogram of the glottal area function. It is basic to current acoustic phonetic theory that the shape of the glottal area wave is almost identical to the shape of the glottal volume velocity wave (Flanagan 1962, 1965; Fant 1958). Thus we can treat a glottographic signal as if it represented the original sound produced by the glottis, unaffected by the resonances of the supra-glottal cavities.

Zeroes in the spectrum of the glottal wave (cf. Figure 7) have been noted by Miller (1959) and can be accounted for mathematically by Fourier analysis of the glottal wave (ibid; Flanagan 1962, 1965). Although there is still great uncertainty as to how these zeroes affect the perception of speech sounds and to what extent they contribute to identifying the speaker, their acoustic effect on the speech output is quite obvious as is shown by a comparison of the spectrograms of the glottal source and the speech output for the word "sealed" in Figure 7. The zero that occurs at the 7th harmonic (700 Hz) in the glottal source during the vowel is also manifested in the speech output at the mouth (see arrows).

Current acoustic phonetic theory is fairly well agreed that any acoustic coupling between the supraglottal cavities and the glottis is small and negligible. But our investigations did suggest that there is a very important mechanical coupling between the glottis and the supraglottal cavities, manifested every time the tongue or jaw is moved and every time a closure or release of a closure in the supraglottal cavity causes a sudden change in the pressure drop across the glottis. This mechanical coupling is responsible not only for changes in the frequency of vocal cord vibration but also for very real dynamic changes in the shape of the glottal wave and thus in the glottal spectrum. This can be seen in Figure 7 in that part of the spectrogram corresponding



Spectrogram of the word "sealed".

Simultaneous spectrogram of the glottal wave.

Figure 7

to the closure for /d/, in which the zero previously located at 700 Hz now shifts downward to 620 Hz and a new zero becomes apparent at 350 Hz, thus reflecting a change in the shape of the glottal wave.

Perhaps the most useful feature of the photo-electric glottograph is the simplicity of its construction and use. Compared with much of the electronic gadgetry currently in use in speech research, the glottograph is about as straightforward to use as a meter stick. And as the necessity of sampling several physiological parameters of speech simultaneously becomes ever more apparent, researchers in the field will have to pay more attention in the design of their experiments to the ease with which the data may be obtained, recorded and analyzed.

I gratefully acknowledge the encouragement and helpful suggestions of Peter Ladefoged, Hans von Leden and Norris McKinney, and the considerable technical assistance of Stan Hubler and Ralph Vanderslice.

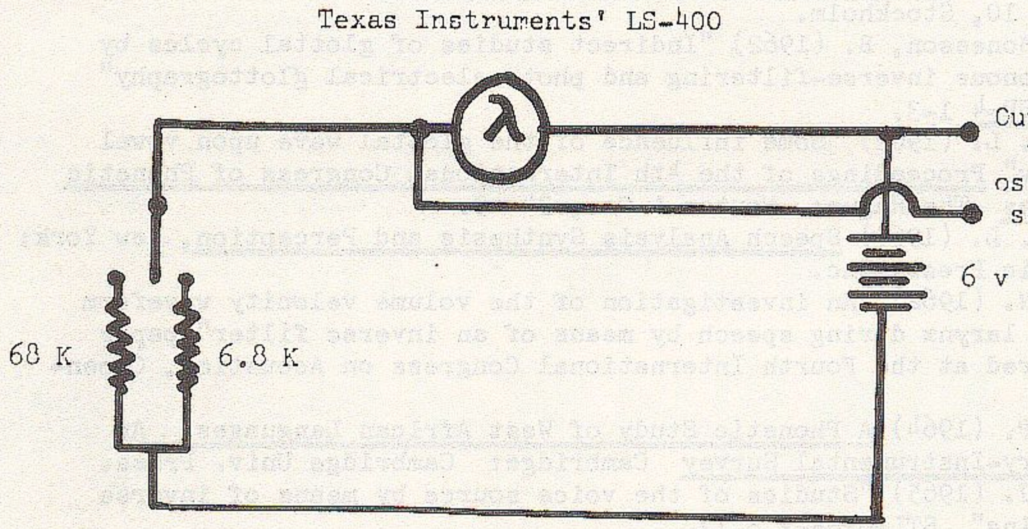


Figure A. Circuit diagram of glottograph.

Bibliography

Abramson, A. S., Lisker, L. & Cooper, F. S. (1965) "Laryngeal activity in stop consonants" Status Report on Speech Research-4 New York: Haskins Laboratories, 6.1-6.13.

Catford, J. C. (1964) "Phonation types: the classification of some laryngeal components of speech production" In Honour of Daniel Jones ed. by Abercrombie, D., et al., London: Longmans, 26-37.

Cederlund, C., Krokstad, A. & Kringlebotn, M. (1960) "Voice source studies" Speech Transmission Laboratory, Quarterly Progress & Status Report-1, 1-2.

Cooper, F. S. (1964) Discussions recorded in International Conference on Research Potentials in Voice Physiology, Syracuse, N.Y. [held in 1961] ed. by Brewer, D., New York: State Univ. of New York.

Fabre, P. (1957) "Un procede electrique percutane d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute frequence" Bull. Acad. Nat. Med. 141, 66-9.

Fabre, P. (1958) "Etude comparée des glottogrammes et des phonogrammes de la voix humaine" Ann. Otolar. (Paris), 75, 767-75.

Fabre, P. (1959) "La glottographie electrique en haute fréquence, particularités de l'appareillage" Comptes Rendue, Sociéte de Biologie 153, 1361-64.

Fabre, P. & Frei, A. (1959) "Analyse harmonique des glottogrammes et des phonogrammes de la voix chantee" Ann. Oto-Laryng. 76, 459-63.

Fant, G. (1958) Acoustic Theory of Speech Production Roy. Inst. Techn. Report 10, Stockholm.

Fant, G. & Sonesson, B. (1962) "Indirect studies of glottal cycles by synchronous inverse-filtering and photo-electrical glottography" STL-QPSR-4 1-3.

Flanagan, J. L. (1962) "Some influence of the glottal wave upon vowel quality" Proceedings of the 4th International Congress of Phonetic Sciences The Hague: Mouton & Co., 34-49.

Flanagan, J. L. (1965) Speech Analysis Synthesis and Perception, New York: Academic Press, Inc.

Holmes, J. N. (1962) "An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter" paper delivered at the Fourth International Congress on Acoustics, Copenhagen.

Ladefoged, P. (1964) A Phonetic Study of West African Languages: An Auditory-Instrumental Survey Cambridge: Cambridge Univ. Press.

Lindqvist, J. (1965) "Studies of the voice source by means of inverse filtering" STL-QPSR-2 8-13.

Lisker, L. & Abramson, A. S. (1964) "A Cross-language study of voicing in initial stops: acoustical measurements" Word 20, 384-422.

Lisker, L. & Abramson, A. S., Cooper, F. S. & Schvey, M. H. (1966) "Transillumination of the larynx in running speech" abstract of paper delivered at 71st meeting of the Acoustical Soc. of Amer., Boston.

Miller, R. L. (1959) "Nature of the vocal cord wave" J. Acous. Soc. Amer. 31. 667-77.

Moore, P. (1962) "Observations on the physiology of hoarseness" Proceedings of the 4th International Congress of Phonetic Sciences The Hague: Mouton & Co., 92-5.

Moore, P. & von Leden, H. (1958) "Dynamic variation of the vibratory pattern in the normal larynx" Folia Phoniatrica 10, 205-38.

Sonesson, B. (1959) "A method for studying the vibratory movements of the vocal cords. A preliminary report" Journal of Laryngology 73, 732-37.

Sonesson, B. (1960) "On the anatomy and vibratory pattern of the human vocal folds -- with special reference to a photo-electric method for studying the vibratory movements" Acta-Oto-Laryngologica Supplement 156, Lund.

Wendahl, R. W., Moore, G. P. & Hollien, H. (1963) "Comments on vocal fry" Folia Phoniatrica 15, 251-55.

Zemlin, W. R. (1959) "A comparison of high speed cinematography and a transillumination-photoconductive method in the study of the glottis during voice production" unpublished M. S. Thesis, University of Minnesota.