Glottal noise during speech

Rothenberg, M.

journal: STL-QPSR
volume: 15
number: 2-3
year: 1974
pages: 001-010

http://www.speech.kth.se/qpsr
I. SPEECH PRODUCTION

A. GLOTTAL NOISE DURING SPEECH

M. Rothenberg*

Abstract

Additive aperiodic components in the voice of one adult male speaker were investigated by using a high pass filter to eliminate most of the periodic components. The glottal volume velocity waveform was recorded simultaneously by inverse-filtering the oral volume velocity. It was found that for flow rates below one liter/second the noise tended to be strongest when the instantaneous glottal volume velocity was about 200 to 300 milliliters/second. When the average glottal air flow was increased, as when the vocal folds were partially abducted for an intervocalic [h], another aperiodic component was sometimes noted at instantaneous flow rates above about 1.2 to 1. liter/second. Visual observations using a fiberoptic probe, spectrograms, and measurements of the pressure just above the glottis indicated that this high-flow-rate noise was actually generated in the vocal tract above the glottis, and not within the glottis. These patterns in the occurrence of aperiodic components were used to add filtered gaussian noise to a three-parameter, "black-box" model of the voice source. Results of informal listening tests are reported.

Introduction

It is generally recognized that during normal, non-whispered speech there can be energy components generated at or near the glottis that are much less periodic than the quasi-periodic pulse trains usually used as a model for the glottal source during voiced speech. Most obvious is the noise-like energy commonly associated with the unvoiced vowel-like sound [h], sometimes referred to phonologically as a glottal fricative (Strevens 1960). However, even during voicing there are often slight aperiodicities that have been modeled by small random variations of glottal period (Lieberman 1961, Dolansky and Tjernlund 1968) or by randomly occurring additive noise components (Fujimura 1968). We report here an attempt to investigate experimentally the mechanisms underlying some of the additive noise components present during voicing and during glottal fricatives.

Additive noise during voicing

There is very little evidence as to the nature of the noise components during voicing, and how they vary as a function of laryngeal adjustment.

* Guest researcher from Syracuse University, Syracuse, N.Y., U.S.A. from September 26, 1973 - March 15, 1974.
To help fill this gap the following experiment was performed. For a single adult male speaker (the author), numerous samples of the glottal air flow waveform were displayed on one trace of a dual tract storage oscilloscope. The glottal air flow was obtained by inverse-filtering oral volume velocity (Rothenberg 1973). On the other trace was displayed the energy in the radiated acoustic pressure waveform above about 5 kHz. A number of sharp-cutoff filters was used at different times to remove the energy below 5 kHz, but the most successful was the "8 kHz" (center frequency) octave band filter of a Brüel & Kjær model 2113 spectrum analyzer. The lower limit of 5 kHz was chosen because for this speaker the energy above 5 kHz appeared to be primarily random in nature, in that the waveform did not repeat from cycle to cycle except for a small component just following the glottal closure. It also seemed likely that the random energy above 5 kHz could be used as a measure of the relative strength of the random energy over a considerably wider bandwidth. (To actually measure the random energy below 5 kHz, where it overlapped with the stronger periodic components, is theoretically possible, but required a much more complex analysis procedure.)

Though there was a considerable variation in the strength and location of the noise energy during the glottal cycle, the most consistent pattern in the occurrence of the noise is shown in Fig. I-A-1. These photos were taken in an almost consecutive sequence while attempting to vocalize with varying degrees of compression of the vocal folds, from "breathy" (vocal folds partially abducted) to "tight" (vocal folds adducted or pressed together more than in normal voicing) (Rothenberg 1973). Though subglottal pressure was not measured, the speaker had considerable previous experience measuring his own subglottal pressure, and could estimate subjectively with some confidence that the pressure stayed within a range of about 7-10 cm H₂O. The Brüel & Kjær octave filter was used. The shutter on the camera was set for a one second exposure. This was long enough to capture about 15 consecutive traces. Afterwards the camera was retriggered with a closed glottis to record zero air flow.

The patterning of the noise relative to the glottal cycle can be seen best in the left hand half of each photo, where the time alignment between the traces was best. Since the inverse-filter and the band pass filter both had a delay of about .7 milliseconds, the upper and lower traces can be considered to be in approximately correct time alignment.
Fig. I-A-1. Simultaneous recordings of glottal air flow and radiated acoustic energy above 5500 Hz for voicing at three levels of vocal fold adduction.
In the "normal" voice, and perhaps also in the "tight" voice, there is a small burst of added energy just after the glottal closure which might have been caused by the periodic energy of the glottal closure. If one neglects this component, the following pattern emerges:

1. The random component was smallest in the "normal" voice, where it tended to be most predominant during the middle of the opening and closing phases, and minimum when the air flow was zero or near its peak value.

2. The random component during the "tight" voice occurred primarily during the glottal pulse and was rather uniform in amplitude throughout the pulse.

3. The random component during the "breathy" voice occurred primarily when the air flow was minimum, and was least strong (almost at the level of the background and system noise) when the air flow was near its peak.

These observations are consistent with a mechanism for noise production that generates noise most strongly at a glottal opening equivalent to an air flow of about 200 to 300 milliliters/second. This result was surprising in view of the fact that the noise in breathy voicing is commonly thought to be caused by the high air flow. However, an examination of the air flow pattern at the glottal orifice makes these results at least plausible. An increase in noise with increased glottal opening would presumably be due to a transition from laminar to turbulent streaming within the glottis. This would occur if the Reynolds number for the glottal flow,

\[ \text{Re} \approx \frac{\text{Particle velocity in cm/sec} \times \text{glottal width in cm}}{15} \]

exceeded the critical value for an orifice the shape of the glottis (Meyer-Eppler 1953). As the glottal width increases, the particle velocity decreases slightly due to the lowered subglottal pressure, and so the Reynolds number, though it increases, increases more slowly than the glottal opening. The critical Reynolds number for turbulence production is difficult to determine from the literature, since it depends greatly on the details of the shape of the glottis, and of its entrance and exit pathways. Since the discontinuity in area tends to be less as the glottis opens, the critical Reynolds number probably increases somewhat with glottal width. These factors together indicate that there may be no tendency to turbulence at the glottis as it opens. As a simple experiment, one can
blow through the pursed lips separated very slightly to imitate the glottal opening. For intraoral pressures similar in magnitude to the subglottal pressure during speech, there is little audible evidence of turbulence as the separation between the lips is increased.

The small amount of noise that was found to occur at small glottal openings might have been due to a decrease in the critical Reynolds number, as, for example, if a misalignment of the vocal cords caused a strong change in the direction of streaming when the cords were very close. Alternatively, there might have been some other effect more difficult to identify, such as the friction of the air stream along the moist glottal walls.

The amplitudes of the noise components shown in Fig. I-A-1 were about 50 dB below the amplitude of the first formant energy in normal voicing. (The vowel was [æ].) If we estimate that the amplitude of all the random energy in the waveform, including the frequency range excluded by the filter, was about 10 dB higher, we must conclude that the random component in the normal voicing would be only marginally audible, and then only above 3 or 4 kHz, where it is not masked by the periodic component. However, when the voicing was more breathy or more tight, the periodic components became smaller and the random components larger, and the random components could be a minor but perceptible auditory feature.

Additive noise during glottal fricatives

The production mechanism for glottal fricatives such as /h/ in English involves some degree of abduction of the vocal folds. Though the evidence presented above indicates that glottal noise during voicing tends to decrease with glottal area and glottal air flow, listening to [h] type sounds leads one to conclude that there must also be some source of a-periodic energy that tends to get stronger as the glottis opens. To investigate this, the author produced a large number of samples of initial and medial[h], while listening for noise during the [ h ] and observing the glottal air flow waveform on a storage oscilloscope. Pictures of the screen were taken occasionally. It was clear that there was some source of noise in the [ h ] that could occur at high air flow rates in certain vowel contexts. The perceived strength of this noise was very well correlated
with certain features of the glottal air flow waveform. These features are illustrated by the set of oscilloscope photos shown in Fig. I-A-2.

The waveforms in Fig. I-A-2 are of the inverse-filtered glottal air flow in repetitions of the nonsense syllable [æ h æ], low pass filtered above about 800 Hz. They were chosen to represent different degrees of glottal opening during the [h]. The top two waveforms show an opening and closing of the glottis of approximately minimal duration, in which the air flow pattern during the most open state is almost sinusoidal, and offset from zero. This is the type of waveform we might expect when the vocal folds are abducted just enough during the [h] so that they do not come in contact during the vibratory cycle. There was very little perceived noise during these productions. The glottal opening movements in photos C and D of Fig. I-A-2 are stronger and of longer duration. In addition, they show a small flattening of the top of the oscillations whenever the air flow exceeded about 1.3 or 1.4 liter/sec. A significant amount of acoustic noise could be heard during these productions, and one might guess that the reduction in the peak flows occurred because of a transition from a low impedance laminar streaming to a high impedance turbulent streaming somewhere in the vocal tract.

The flattening effect noted in waveforms C and D of Fig. I-A-2 can be seen to be much stronger in waveforms E and F. In E, the compression is initiated at about 1.1 or 1.2 liter/second, and in F the compression seems to start at about 0.9 liter/second. In all productions in which the airflow was similar to that in E and F, the noise in the [h] was highly audible, and there was a feeling of constriction at the back of the throat that seemed to correlate well with the level of the noise. With some practice it was possible to vary this feeling of constriction, and thereby vary the amount of random energy produced, and the amount of "flattening" in the glottal air flow waveform. In waveform F, random oscillations on the non-oscillatory, high air flow part of the waveform make it clear that random energy was being generated during that interval.

To investigate further the mechanism of noise generation at high air flow rates, a number of samples of the glottal air flow waveform during the production of [h] was photographed, with an effort made to use varying degrees of "constriction". Fig. I-A-3 shows a set of 5 waveforms with different degrees of subjective feeling of constriction. The
Fig. I-A-2. Glottal airflow during the [h] of [æ h æ] for different degrees of vocal fold abduction during the [h].
Fig. I-A-3. Glottal air flow near the onset of voicing during productions of /h/ with different degrees of subjective pharyngeal constriction.

- Strong constriction held into the onset of voicing
- Moderate constriction
- Slight constriction
- Little or no constriction during aspiration
amount of noise preceding voicing tended to increase with the degree of subjective feeling of constriction, as can be seen in the increase in the low frequency (under 800 Hz) noise energy in the waveforms for the higher degrees of constriction.

The waveforms suggest to us a production mechanism in which little of the noise at high air flow rates is produced at the glottal orifice, the air flow there remaining laminar to quite wide glottal openings. The noise would be produced instead by turbulent flow at a constriction elsewhere in the vocal tract. According to this interpretation, in the topmost waveform of Fig. 1-A-3 there was no such other constriction in the vocal tract to create turbulence and limit the air flow. This airflow waveform therefore shows most clearly the variation of glottal area as the vocal folds came together from a very open adjustment. On the other hand, on the lower three traces in Fig. 1-A-3 there was an additional constriction somewhere in the vocal tract that created turbulence when the glottal opening was wide. The turbulence at this additional constriction created a high acoustic impedance at that point in the vocal tract, which tended to delay the onset of the glottal oscillations until the glottis closed enough to reduce the air flow below the level required to support the turbulence.

To see where such a constriction could occur in the vocal tract, the glottis of the same speaker was observed by means of a fiberoptic probe inserted through the nose, and lowered to a point in the pharynx just above the epiglottis. Repeated productions of [h] were made in varying contexts. When the [h] preceded a back vowel, such as [a] or [a], there seemed to be a strong correlation between the amount of perceived noise in the h and the degree to which the epiglottis came into view over the open glottis. Conversely, there was little noise when the epiglottis was far forward, unless there was felt to be a constriction elsewhere in the vocal tract, as along the tongue blade with a high [i].

The hypothesis that the high air flow turbulence in the [h] for this speaker was not produced within the glottis was also checked by measuring the pressure above the glottis. The tip of a rubber catheter was suspended to a point just above the glottis, with the other end of the catheter connected to a pressure transducer. The catheter was inserted through the mouth, to a length that must have brought it quite close to the glottis. It was found that the production of a "noisy" [h] in the context [m h w]
at a conversational volume was always accompanied by a pressure of 1 to 2 cm water during the aspirated phase of the [h]. Small adjustments in the position of the catheter did not seem to affect the pressure reading, and so it could be assumed that the reading was not an artifact of the airflow impinging directly on the catheter tip.

It was also observed that in spectrograms made from recordings of [æ h æ] for this speaker the formants tended to rise 10 to 20 percent during an aspirated [h]. This rise could have been caused, at least in part, by the acoustical shortening of the vocal tract that would result from a high acoustical impedance one or two centimeters above the glottis, at the level of the epiglottis. An increase in the formant frequencies could also be partially due to acoustic coupling to the subglottal system through the open glottis (Fant, et al. 1972). This tendency for the formant structure to increase in frequency during the [h] was verified in spectrograms made from the voice of a second adult male speaker. A typical spectrogram for this speaker is shown in Fig. I-A-4, for the phonetic sequence [a ha ? æ h æ ? ε h ε]. A Kay model 6061 spectrum analyzer was used.

**Simulation for speech synthesis**

The results of the above experiments were used in adding filtered gaussian noise to an electrical "black box" model of the voice source (Rothenberg, et al. 1974). The source was an ad hoc electrical network designed to have an output similar to an actual glottal air flow waveform, and was controlled by three voltages representing hypothetical neurological command parameters that would act to vary the Frequency, Loudness and Tightness (affect of vocal fold adduction) of the voice for an unconstricted supraglottal vocal tract. Figs. I-A-5 and I-A-6, respectively, illustrate how the output waveform varied over the full range of the Loudness (L) and Tightness (T) parameters. A "normal voicing" could be approximated by the waveform third from the bottom in Fig. I-A-5 and fourth from the bottom in Fig. I-A-6. The voice source was connected to the OVE III serial formant synthesizer at the Dept. of Speech Communication using the control scheme sketched by Rothenberg et al., and subjected to a number of informal listening test using nonsense syllables and connected speech.
Fig. I-A-4. Wide band spectrogram of the phonetic sequence
\[ \text{[a h a ? m h a ? e h e \text{.}]} \].
Fig. I-A-5. Simulated glottal air flow waveform with $L$ varied in equal steps.

Fig. I-A-6. Simulated glottal air flow waveform with $T$ varied in equal steps.
To simulate the noise component found to occur at low flow rates during voicing, a circuit was included in the voice source to insert noise when the simulated air flow was in the appropriate range, about 100 to 400 milliliters/second. The noise amplitude was also made slightly dependent on the waveform derivative, in order to simulate the affect of flow inertance at small glottal openings. For a given glottal area, this inertance would tend to decrease the flow, and therefore the noise, during the glottal opening phase, and increase these parameters during the glottal closing phase. The noise level was made proportional to the loudness parameter. The noise spectrum was chosen by informal listening tests to be falling at about 3 dB per octave, but still might be considerably different than the spectrum of actual glottal noise. The noise level was set about 6 dB above the level at which the insertion of the random component would just be perceived with simulated normal voicing.

Accurate acoustical modeling of the high-flow-rate noise source can be quite complex. In the electrical model of the voice source, we have so far only introduced an additive noise component as the simulated air flow waveform rises above the equivalent of about 1.2 liter/second. The noise amplitude increases roughly in proportion to simulated flow level above the threshold and to the loudness parameter L. (This second dependency may have been unrealistic.) The noise source is the same as used for the low-flow-rate noise. The constant determining the noise level was set by informal listening tests. The formant structure is extrapolated directly from neighboring vowels, with no compensation for vocal tract excitation being above the glottis.

The insertion of the small amount of low-air-flow noise was generally considered to make the synthesized speech more natural sounding, though it is likely that such naturalness could also be obtained by much simpler ad hoc methods, as the scheme suggested by Fujimura (1968). However, the more significant question is whether a more accurate simulation of this noise improves the perception of linguistically significant transitions in laryngeal adjustment, as changes in the voice parameters termed here Loudness and Tightness. Since the level of the noise is small relative to the periodic component, we should expect that the variation of the frequency, amplitude, and spectral shape of the periodic component should be the more dominant factors in the perception of most transitions of laryngeal adjustment, and this was generally found to be the case in listening to transitions
in the model parameters with and without the low-air-flow noise present. However, the informal listening tests used were not adequate to test for smaller, second-order effects.

In the synthesis procedure used, the high-air-flow noise might be generated by the voice source model during any "high-air-flow" sound, such as an unvoiced fricative or stop burst, however, we discuss here only the affect on the perception of [h]. The voice source model was used to generate an [h] sound by appropriately varying the Tightness parameter, i.e., increasing it to a "normal voicing" level after an open glottal state, or increasing then decreasing the parameter for an intervocalic [h]. The high-flow-rate noise, though not entirely satisfactory in spectral quality, seemed to improve the naturalness of those types of [h] sounds in which it might be expected to occur naturally, as after an open glottal state, or before a stressed vowel, when the glottal opening movement might be long (> 150 msec) and result in a larger than average adduction of the vocal folds. A simulation of a shorter intervocalic glottal opening transition, such as might occur in an intervocalic English /h/ in an unstressed position, tended to be heard as a natural [h] whether or not the noise was present. (In the final synthesis procedure, the time constant used for the Tightness parameter tended to prevent the simulated air flow from reaching the threshold for noise generation when the [h] was shorter than about 125 msec.) In fact, fairly natural sounding [h] sounds could be produced in any context with no random components at all added, provided the variation in the tightness parameter was made with a natural time constant. With no noise present, the [h] was apparently signaled by transitions in the spectral distribution of the quasi-periodic components.

Acknowledgments

The author greatly appreciates the numerous conversations concerning the work reported here that he enjoyed with members of the staff of the Dept. of Speech Communication, especially P. Branderud, R. Carlson, B. Granström, and Dr. J. Lindqvist-Gauffin. The speech synthesis experiments were made in cooperation with Mr. Carlson and Mr. Granström, and the fiberoptic probe observations in cooperation with Dr. Lindqvist-Gauffin.

This work was partially supported by the National Institutes of Health, USA.
References


