A biologically inspired silicon vocal tract

Keng Hoong Wee, Lorenzo Turicchia, and Rahul Sarpeshkar

The first integrated-circuit vocal tract, combined with a bionic ear processor in a feedback speech-locked loop, provides a new way to process speech.

Electrical circuit models of biological systems provide an intuitive mechanism for engineers’ understanding and are increasingly used to improve the performance of related technology. For example, visual processing performed by the retina can be modeled by a resistive network of interconnected photodetectors and analog processing elements. Complex bio-mechanical systems such as the heart, cochlea, and vocal tract can be modeled using electrical circuits by mapping pressure to voltage, volume velocity to current, and mechanical impedances to electrical impedances, and by representing valves with diodes. Silicon models of the retina have been used in machine vision systems and circuit models of the heart have been used to shed light on cardiac and circulatory malfunction in medicine. Silicon cochlea models have led to improved speech recognition in noise and low-power cochlear-implant processors for the deaf.

In this vein, we have developed the first integrated-circuit vocal tract that uses a physiological model of the human vocal tract to synthesize speech. The system employs articulatory parameters that are intrinsically compact, robust, and linearly interpolatable: it could therefore be well suited to noise-robust automatic speech recognition, speech compression, audio noise cancellation, and could form the basis for future bionic speech prostheses.

Figure 1 shows an analysis-by-synthesis block diagram that creates what we term a speech locked loop (SLL) in analogy with the phase locked loops (PLL) commonly used in communication systems. The auditory processor and controller are analogous to a phase detector and loop filter in a PLL. The vocal tract is analogous to a voltage-controlled-oscillator (VCO). The speech produced by the vocal tract is analyzed and compared to that of the input, and a measure of error is computed. Different sounds are generated until one is found that produces the least error at which time the SLL locks to the input sound with an optimal vocal tract profile produced by the controller. Analysis-by-synthesis methods were previously implemented using computationally expensive digital techniques. Instead, our strategy employs a silicon vocal tract to drastically reduce power consumption, and thus is ideal for portable speech processing systems such as cell phones and personal digital assistants. Power consumption can be reduced further using our previously developed analog bionic ear processor as the auditory processors for the SLL.

Our circuit model of the vocal tract, shown in Figure 2, represents the human vocal tract (composed of nasal, pharyngeal, and oral tracts) as acoustic tubes using a transmission line model. Each transmission line comprises a cascade of tunable two-port elements, corresponding to a concatenation of short cylindrical acoustic tubes (illustrated in Figure 3) of length $\ell$ with varying
Figure 2. Schematic of transmission line vocal tract.

cross-sections. Each two-port element is an electrical equivalent of an LC π-circuit element where the series inductance L and the shunt capacitance C may be controlled by physiological parameters corresponding to articulatory movement (e.g., movement of the tongue, jaw, lips). Speech is produced by controlled variations of the cross-sectional areas along the tube in conjunction with the application of one or two sources of excitation, namely a periodic source at the glottis and/or a turbulent noise source $P_{turb}$ at some point along the tube.

In Figure 2, the glottal source is represented by a voltage source $P_{alv}$ with variable source impedance $Z_{GC}$, which is modulated by a glottal oscillator. We use a circuit model of the glottis that comprises linear and nonlinear resistances connected in series to represent losses occurring at the glottis due to laminar and turbulent flow, respectively. The turbulent source $P_{turb}$ is connected in series with an impedance $Z_{SGC}$. The location of $P_{turb}$ and $Z_{SGC}$ is not fixed in the oral tract, but varies depending on the constriction location. During the production of nasal sounds, the nasal tract becomes coupled to the oral tract via the velar impedance $Z_{V}$; otherwise, $Z_{V}$ is an open circuit. At the lips and at the nose, the transmission lines are terminated by radiation impedances $Z_{rad}$ and $Z'_{rad}$, and the radiated sound pressures $P_{rad}$ and $P'_{rad}$ are proportional to the derivative of the currents flowing in the respective radiating impedances. For simplicity, the electronic vocal tract in its first instantiation only implements the oral tract rather than the oral and nasal tracts.

Our silicon vocal tract is able to generate all speech sounds, given the vocal tract profile and the excitation sources. In order to extensively test and prove the efficacy of our SLL, we introduce a speech-coding scheme based on an anthropomorphic articulatory model that describes the vocal tract profile using seven components, each corresponding to an elementary articulator. Physiologically realistic vocal tract profiles may be represented with reasonable accuracy using these seven articulatory parameters, namely: jaw height, which controls the vertical position of the jaw; tongue body position, which moves the tongue dorsum from the front to the back of the oral cavity; tongue body shape, which indicates whether the tongue dorsum is rounded or unrounded; tongue tip position, which controls the position of the tongue apex; lip height, which varies the mouth opening; lip protrusion, which controls the mouth protrusion; and larynx height, which raises or lowers the position of the larynx. The trajectories of these seven components in time are reconstructed by our SLL, creating what we call an articulogram, which can be used to supplement the spectrogram to enhance the robustness of automatic speech recognition systems.

Figure 4(a) shows the spectrogram of a recording of the word ‘Massachusetts’ lowpass filtered at 5.5kHz. Figure 4(b) shows what we term the vocalogram, a 3D plot of the vocal tract profile as a function of time, extracted from analyzing the recording using the speech locked loop illustrated in Figure 1. Figure 4(c) shows the spectrogram of the same word re-synthesized by our SLL. In Figure 4(c), it is evident that high frequency speech components that were absent in Figure 4(a) have been introduced by the SLL, which inherently synthesizes only speech signals.

Continued on next page
because it is based on a physiological model. Such signal restorative properties are particularly important when dealing with noisy speech.

**Conclusion**

Our integrated-circuit vocal tract based on a physiological model of the human vocal tract can be used with auditory processors (e.g., analog bionic ear processors) in a feedback speech-locked loop to implement speech recognition that is robust in noise. It also has potential for future low-power, real-time speech production, speech compression, audio noise reduction, and bionic speech-prosthesis systems.

**Author Information**

**Keng Hoong Wee**  
Department of Electrical and Computer Engineering  
National University of Singapore  
Singapore, Singapore

Keng Hoong Wee is an adjunct assistant professor. He received his PhD in electrical engineering and computer science from the Massachusetts Institute of Technology. His research interests lie in biologically inspired circuits and systems, biomedical systems, and speech processing.

**Lorenzo Turicchia**  
Research Laboratory of Electronics  
Massachusetts Institute of Technology  
Cambridge, MA

http://www.rle.mit.edu/avbs/whoweare_lorenzo.html

Lorenzo Turicchia is a research scientist whose main research interests are in nonlinear signal processing, especially for audio and biomedical applications, and bioelectronics. His work has included research on cochlear implants, visual prostheses, speech prostheses, speech recognition, and wearable medical devices.

**Rahul Sarpeshkar**  
Research Laboratory of Electronics  
Massachusetts Institute of Technology  
Cambridge, MA

http://www.rle.mit.edu/avbs

Rahul Sarpeshkar is currently an associate professor and heads a research group in Analog VLSI and Biological Systems. His research on bioelectronics has won several awards including the Packard award given to outstanding faculty.

**References**