

Biomimetics of sound production, synthesis and recognition

G. Rosenhouse

Swantech - Sound Wave Analysis & Technologies Ltd, Haifa, Israel

Abstract

Biomimesis of sound production, synthesis and recognition follows millenias of years of adaptation as it developed in nature. Technically such communication means were initiated since people began to build speaking machines based on natural speech concepts. The first devices were mechanical and they were in use till the end of the 19th century. Those developments later led to modern speech and music synthesis, initially applying pure mechanics. Since the beginning of the 20th century electronics has taken up the lead to independent electronic achievements in human communication abilities. As shown in the present paper, this development was intentionally made along history in order to satisfy human needs.

Keywords: speaking machines, biomimetics, speech synthesis.

1 Introduction

Automation is an old attempt of people to tame machines for the needs of human beings. It began mainly, as far as we know, since 70 DC (with Heron). In linguistics, the history of speaking machines began in the 2nd century when people tried to build speaking heads. However, actually, the process was initiated in the 18th century, with Wolfgang von Kempelen who invented a speaking mechanism that simulated the speech system of human beings. It opened the way to inventions of devices for producing artificial vowels and consonants. This development was an important step towards modern speech recognition and synthesis, and had to do with one of the fundamental principles of the behaviour of living creatures.

While normally people use five senses for perception, **the only natural way** for human performance of intentional activities is by production of sound and



vibration – people can only generate with their bodies only externally mechanical movement activated by hidden chemical, electro-chemical and thermodynamic processes. This observation was presented in Rosenhouse [1], including the statement that active acoustics is based on **deliberate use of sound and vibration sources in order to influence and cause anything else the person wants**. Thus, modern acoustics became a crossroad between most areas of sciences, biology, technology psychology and arts, which is a view point that inspired the structure of volume I of the above book as well as the one to follow – Rosenhouse [1]. As a result, physical acoustics, electro-acoustics, sound synthesis, communication, art, music, speech and others are coupled now in novel ways into the main stream of science and technology, and in some fields become a major topic for research.

As an example of this understanding, the topic called **movement ecology** (e.g. Nathan, et al. [2]) appeared and became a “hot issue” and a reason for an international inter-university collaboration – but it is not different from intentional or deliberate sound and vibration control as we defined it, although its goals are not the same. The term “control”, which is applied by means of biomimesis, includes motivation, physiological power, navigation and environmental influence (with a link to Darwinistic evolution) is included in acoustics. Hence, in the design of speaking machines, the automata response differs from that of living creatures. Aristotle (4th century B.C.) described in “De Motu Animalium”: “Now we see that the living creature is moved by intellect, imagination, purpose, wish and appetite”. Much later Descartes (1596-1650) spoke also about an “external world”: “All these considerations are enough to establish that it is not reliable judgment but merely some blind impulse that has made me believe up to now that there exist things distinct from myself which transmit to me ideas or images of themselves through the sense organs or in some other way”. However, the development of speaking machines was devoted mostly to the technological achievements.

2 The human voice

The human voice is generated by the vocal cords that open and close the glottis and results in multi-harmonics, periodic and a-periodic wave form. This acoustic signal is filtered by the throat and nose resonating passages according to their geometry and volume. The whole complex generates under the controlled shaping different formants and a variety of vowels, consonants, unvoiced and voiced plosive, fricative (etc.) elements. Quantitative acoustic modelling of sound generation by the vocal tract was suggested, for example, by Fant [3] and by Stevens [4]. Such modelling is used also for investigation of breath sounds for (lung) pulmonary clinical diagnostics. An example of such a research for medical applications is the effects of diameter, length and circuit pressure on sound conductance through endotracheal tubes, by Räsänen et al. [5]. In linguistics, researchers are interested in the effects age and sex, language and dialect, health and behavioural differences – all of which are expressed by the human voice.



3 The need for speech synthesis

Sound synthesis today is a process of generating sound by analog, digital and software techniques. Synthesis means making a whole out of parts. It can be applied in sciences, engineering and arts. Synthesis might be only technical, as in active sound vibration control, but also emotional by involving feelings.

Since the only natural way of human beings to express themselves is by sound, and motion (gestures) the history of sound synthesis is as long as the existence of the human civilization (see table 1). Since the 20th century natural sound was translated into visual letters, and electrical signals - by machines, but still they are not more than artificial devices to improve the limited scope of human natural activities and communication abilities.

Table 1: The history of language from speech to speech synthesis.

Era	Technique	Years	Main stream
1	Natural	Up to 1790	Human voice musical instruments
2	Mechanical	1790-1925	Distorted and hidden human voice and speech machines
3	Electronic –	1925-1945 *	Mostly electronic analogues
4	first analogue	1945-1965 *	Speech research and processing, acoustic phonetics, information theory, psychology, physiology, acoustics
5	and then	1965-1985 *	Electronic technology build up, the use of statistics
6	digital	1985-to date	Computerized speech technology, enhanced statistics, huge data bases, linguistics

* Following Fant [3].

4 The steps from speech to speech synthesis

Long before the era of modern electronic signal processing became established people created unusual means of verbal communication, with the methods available to them. Oracles were found in most of the ancient civilizations as a kind of worship involved with signs from Gods and prophesy, searching for omens and then augury. The oracle had to be a person of knowledge and wisdom highly infallible and authoritative. One famous oracle is the Phytia of the temple of Delphi in Greece (c. 440 BC). For example, see Broad [6]. From the name oracle the Latin word “öräre”, which means “to speak”, was derived (English: “oral”).

These spiritual and religious events that involved verbal communication with the oracle, in a manner similar to human conversation, stimulated later the invention of “speaking heads”, generally built of metal mechanisms. It is a sort of deliberate acoustic information transfer to the believers from the God through the mediator. Hence, vocal prayers definitely constitute acoustic communication (with God).

The obvious result was that ingenious people tried to build machines that could generate human speech or “speaking heads” and a few were successful. Among those inventors was Gerbert of Aurillac or Pope Silvester II (945-1003), who was murdered in 1003. He possessed a speech automaton designed as a huge golden head, given to him by a magician that programmed the talk. This



“speaking head” served as an oracle by giving a “yes-no” answer to any question posed to it. The head was not found, but references still can be found in the Vatican library with some more details. Albertus Magnus (1198-1280) built a “speaking head” as part of a mechanical servant, which is thought to be of the same kind of that of Gerbert of Aurillac, but the whole mechanical servant was later destroyed. Also Roger Bacon (1214-1294), a renowned medieval English scientist and a philosopher, constructed a brazen speaking head. The speaking mechanism behind those machines is not clear even now

Later, the “speaking heads” served as mechanical toys instead of prophecy devices, and are studies within the framework of robotics. In 1779, Christian Krantzenstein of St. Petersburg developed models of the human vocal tract for the five vowel sounds, a, e, i, o, u. This investigation was followed by the bellows-operated “acoustic mechanical speech machine” fabricated by Wolfgang von Kempelen of Vienna in Austria.

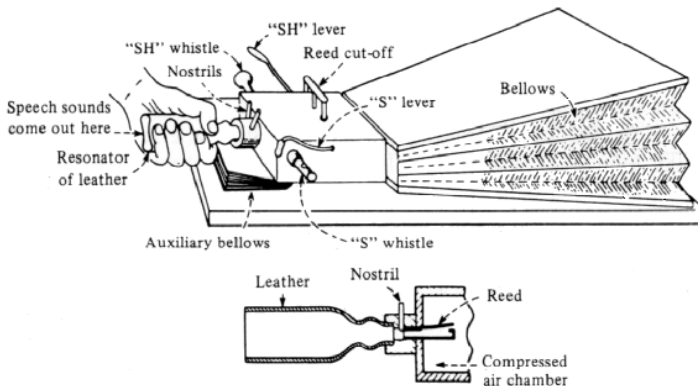


Figure 1: Reconstruction of von Kempelen's speaking machine by Wheatstone (Flanagan [7]).

Von Kempelen described this machine in his book, published in 1791 in German and French, after two decades of developing this topic (see Dudley and Tarnoczy [8] and Pompino-Marschall [9], for instance). This machine included also models of tongue and lips, which enabled the production of consonants and vowels. Charles Wheatstone constructed in 1837 a “speaking machine” based on von Kempelen's design (Fig.1). Faber built in 1857 the Euphonia, and Paget revived Wheatstone's design in 1923.

The neurophysiology of the vocal tract and speech was a parallel research line conducted by Ponizza (1785-1867), Wheatstone (1802-1875), Donders (1818-1889), Broca (1824-1880) and the Nobel Laureate von Békésy (1899-1972). Among the many studies by Donders the most relevant to our study were about the physiology of speech. Following Donders, Helmholtz (1821-1894) built a “Fourier frequency analyzer” as the first synthesizer of sound (Helmholtz [10]), based on “Helmholtz resonators” and electrically activated tuning forks as built

by Koenig (1832-1901) and presented during 1876 and was used in order to study the articulatory generation of speech and music. Helmholtz showed that the timbre of musical notes and of speech vowels were combined of many different overtones (integer multiples of the fundamental), in addition to the fundamental tone. This combination results in the pitch of fundamental frequencies of speech and instrumental music. This way, Helmholtz simulated and understood the nature of human vowel sounds of speech and singing and linked it to the resonances of the vocal tract (each vowel was defined by a few resonant frequencies or formants). In order to create a certain vowel, the muscles of the throat and mouth must change the shape of the vocal tract, producing a new, different set of resonances. Shapes of the vocal tract were built by Chiba and Kajima [11] for generating vowels. Recently Arai [12] constructed a replica of it. The basic theory of those achievements in acoustical phonetics was summarized in a book and a lecture by Stevens [4].

Up to now, this paper concludes roughly the era of the boost of mechanical speaking heads and the electronic speech synthesis emerged, linked to the development of electronic instruments, digital talking heads, synthesized music and relevant computer programming.

Table 2: Milestones in mechanical artificial speech since 945 to about 1890.

Year (s)	Milestone
945-1003	Talking head (Gerebert d' Aurillac)
1193-1280	Talking head (Albertus Magnus)
1214-1294	Talking head (Roger Bacon)
1779	Resonators for vowel production (Christian Krantzenstein)
1791 (1769)	Mechanical talking machine, also "the Turk" (W. Von Kempelen)
1830	Producing a version of von Kempeln machine (C. Wheatstone)
1857	Euphonia (J. Faber), (1830-1840s)
1862	Mechanical Fourier Analyzer (Helmholtz). Donders, Koenig (1850-1890s)
1941, 2001	Mechanical models for vowel production (T. Chiba, and M. Kajima, 1941); Replica of Chiba and Kajima models of human vocal cavity (Arai, 2001)

5 The influence of the electronic era

Sound synthesizers can be classified into two groups: (1) Sound synthesizers which are used to generate and process sounds, including music, and (2) speech synthesizers that are applied in computers and telecommunication and word synthesizers. Deliberate sound control (DSC) is generally based on sound synthesis, where the control unit collects information and forms the signal that affects the result of the process. Then, the output part takes the values of the signals and by its "built in interpretation unit" synthesizes and produces the necessary output. It can be said that when active sound is used, it necessarily involves synthesis. Synthetic sound can be a purely artificial sound, as frequent in electronic music. It can also be an imitation of regular familiar sounds and acoustic instruments. It is common to produce mixed synthetic and imitative sounds as well, including also various kinds of sound emulations.



The first step in sound synthesis is analyzing its components, where the ability to do so depends on the availability of acoustic measurement devices. At present there are many commercial devices that enable this, including software. An early example is the pitch analyzer by Grützmacher and Lottermoser [13]. The breakthrough into electronics began as early as 1922 – see tables 3, 4. Stewart [14] was the pioneer of a fully electrical synthesis device. Using a two-resonance circuit, this device could simulate the vocal tract for specific single vowel sounds, using the two first formants – fig. 2. It could not produce connected utterances or consonants. His machine was followed by the same synthesizer that was produced by Wagner [15] who developed a vowel synthesizer with a four-resonance circuit, preceding the VODER and VOCODR by Dudley et al. [16]. Following the development of electronics, the invention of the telephone and better understanding of linguistics, Bell labs introduced in the 1930s the vocoder: a keyboard-operated electronic speech analyzer and synthesizer, being clearly intelligible. In 1939 Homer Dudley introduced a refined model of the vocoder, named VODER, for “Voice Operating Demonstrator,” which was the electrical, mechanically controlled equivalent of the speech machine built by Wheatstone (Dudley et al. [16].

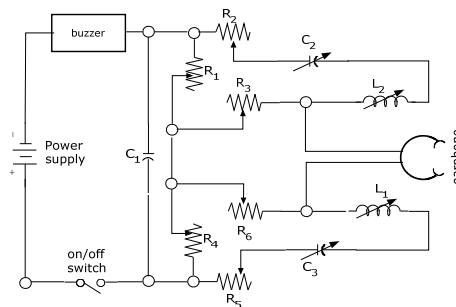


Figure 2: A scheme of the electrical synthesizer for simple speech sounds designed by Stewart [14].

There is a large amount of literature that is useful for the study of sound synthesis. Some of these works are cited here. To start with, Fletcher and Rossing [17] discuss most of the traditional musical instruments that are currently in use in Western music. It gives the necessary basics of physics and mathematics needed to explain the mechanisms by which such instruments produce music and suggestions about the quality and other features of such instruments. A reference book for study, design and realization of electro-acoustic devices, including chapters about transducers in general, and specifically – loudspeakers, microphones and sound recording was written by Rossi [18]. The first part of the book describes sound propagation in acoustic and mechanical systems. Eargle [19] dedicates his handbook to sound engineering and especially, sound recording. It includes chapters on digital recording, analogue digital recording and code–decode noise reduction systems and

production techniques in the studio. The book also reviews multi-channel surround methods, mix-down procedures and trans-aural techniques. Even more devoted to sound engineering is Russ [20], concentrating on sound synthesis analogues, hybrid and digital techniques, sound sampling principles, control of synthesizers and practical synthesis techniques. With all this background, specific chapters about sound reinforcement as in Ahnert and Steffen [21] can be added for the acoustic design of halls (room acoustics, sound reinforcement components, analysis and systems layout and “virtual halls” design).

Table 3: Some milestones in artificial speech since the early electronic development, to 1960.

Years	Milestone
1876(1849)	Invention of the telephone (Antonio, E. Gray, Alexander Graham Bell)
1922	Electronic speech synthesizer (J.Q.Stewart)
1924	Speech analysis by formants
1930s	Vocoder (Bell Labs)
1932	Third formant in vowels (Obata & Tashim,a)
1939	VODER – voice operating demonstrator (H.W. Dudley, Bell Labs)
1946	Invention of the spectrogram device (W. Koenig, H.K. Dunn, L.Y.Lacey)
1950	Static articulatory model (H.K.Dunn)
1951	Pattern playback synthesizer (F. Copper et al)
1952	Delta modulation (F. de Jaeger)
1953	OVEI (orator verbis ellectris), (G. Fant) – cascade formant synthesizer (OVE II – copying and composing sentences); PAT (parametric artificial talker), Formant synthesis OVE (W. Lawrence); Vocal tract static electrical analogue (K.N. Stevens, N. Kasowski, G. Fant)
1957	Speech synthesis by formants
1958	Articulatory synthesis –DAVO, dynamic vocal tract (G. Rosen, MIT); Concatenative synthesis (Joining two character strings = diphones)
1959	Time assignment of speech interpolation; Synthesis by rule
1960	Acoustical theory of speech production – static analogue (G. Fant)

There are many techniques for sound synthesis, where raw sound source produces the raw signal, which is then modified by filters and envelopes, to create the final form. This kind of modification in the frequency domain is:

$$Y(\omega) = H(\omega)X(\omega) \quad (1)$$

With $X(\omega)$ – the input, ω - radian frequency, $H(\omega)$ – the transfer function, $Y(\omega)$ – the output. Equation (1) stands for either the relation between functions for single input – single output relations, or the relation between the vectors X and Y through the coefficients matrix H in case of many inputs and outputs. Hence, the system as a whole is based on “source and modifier”. In music, the input can be the harmonics of a plucked string, which serves as a wave guide. The transfer function can be the enclosed volume of a guitar that serves a resonator and the

Table 4: Some milestones in artificial speech; 1960-2000.

Year(s)	Milestone
1960s,1970s	First full TTS; Vocal tract models - Bell Labs (P. Mermelstein, C. Coker, etc.) ASY computational models of speech production based on vocal tract models (P.E. Rubin, T. Baer., P. Mermelstein, 1970s – Haskins Labs)
1961	1 st computer- based phonetic - synthesis – by – rule program (J. Kelly. L. Gerstman)., In English in 1964 (J. Holmes, I. Mattingly, J. Shreamer)
1962	Computer simulations (J.L. Kelly, C. Lochbaum); also later: Nakata Kazuo, T. Mitsuoka, , 1965; E. Matsui, 1968; T. Honda, Inoue, S., Ogawa, Y., 1968; P. Mermelstein., 1969, 1973
1962, 3,4	Computer speech synthesis (B. Dennis, E.C. Whitman, R.S. Tomlinton); Cepstral analysis (1964)
1967-1970	LPC (linear predictive coding of speech, 1967); TTS – text to speech (Norico Umeda et al), Prosody (A.W.F. Huggins). Also, first prosodic synthesis by rule (I. Mattingly) 1968; Windbag vocal tract analogue speech synthesizer (B. Baxter, W. J. Strong, 1969); Dynamic time warping for speech recognition; Adaptive Delta modulation, “speak and spell” using formant analysis (Texas instruments) - 1970
1973	Adaptive differential delta modulation
1975	Markov method for speech recognition
1976	Sentences produced by an articulatory model (J. Flanagan, K. Ishizaka; Sub-band coding); Optical scanner (Kurzweil)
1977	Adaptive transform coding; Concatenation of linear prediction diphones (J.Olive)
1978	Large scale integrated circuit speech synthesizer; Votrax – Votrax based type-n-talk (R. Gagnon) - 1979 (1978)
1980	Vector quantization for speech coding; TI speak 'n' spell; Texas instruments (R. Wiggins)
1981	Klattalk. Mitalk - S. Allen, S. Hunnicut and D. Klatt, 1987)
1982	Multipath-excited linear-prediction
1984	Sinusoidal modes
1985	ANN – Artificial neural networks; PSOLA – pitch synchronous overlap add (H. Traunmüller)
1986	Comparison of synthesis and natural sentence (D. Klatt); Speech synthesis based on sinusoidal representation (R. J. McAulay & R. J. Quatieri)
1990-2000	Articulatory synthesizer considering tongue shape (S. Maeda, 1990); Reading machine system for the blind (Kurzweil, Bhahena, Baum, 2000)

output of the instrument is the sound radiated into air. The general shape of the output envelope of such a signal is given in fig. 3a, and it consists of an attack time - A, which is the rise of the signal from zero to the full size, Decay time – D, which is descent from the full size to the designated one, Sustain level – S, which occurs when the key is held down and the signal is steady and finally, the release time – R, when the key is released and the signal decays from the steady state to zero. Altogether this is the ADSR decay curve. For a guitar string the envelope is more like the curve in figure 3b, with rise and decay times. When

Table 5: Some basic sound synthesis methods.

Group	Technique	Definitions and comments
Analogue synthesis for audio signal that can be processed by filters and amplifiers	Additive	Superposition of tones of different amplitudes
	Subtractive	Filtering complex sound to shape harmonic spectra
	FOM/VOSIM	The techniques have few parameters to extracts harmonic peaks.
	Wave table synthesis Generally suit MIDI and cheap sound cards	Looping short, complete wave shapes to get sustained one or several sound segments producing composite sample.
	Phase distortion	Varying during the playback the speed of the stored waveforms stored in wave tables
	Wave shaping	Deliberate signal distortion to modify the result
	Mixed/Hybrid synthesis	Combining recorded and artificial sounds into a new signal.
Digital	Direct digital synthesis	Software modification of the generated waveforms
	FM	Sound signals are emulated by several sine/cosine functions. 2-6 oscillators for modulation are common.
	Frequency modulation	
	Vector synthesis	Fading between any numbers of sound sources.
	LPC	Linear predictive coding for speech/audio synthesis.
	Re-synthesis	Modification of sampled digital for the playback use
Sample based synthesis	Sampling	Continuous audio signals converted into recorded discrete digital samples by pulse code modulation (PCM), using transistors since the 1950s. Applying computer musical instrument (CMT) since 1979. After the 1990s CD-Rom became a popular replacement.
	Granular synthesis	Combining many short sound pieces, to long sound recording, imitating advanced compositions/classical musical instruments/non-physical sounds.
Physical modelling	Wave sequencing	Linear combinations of small segment to a new sound.
	Physical modelling	Mathematically modelled sound waveform to simulate a physical sound source, by equations and algorithms
	S & S Sample and synthesis	Technologies of wave table and sample replay, adding filtering and shaping
	Digital wave synthesis	Replaces signals with discrete digital forms to enable computerized processing of synthesizing sounds.
	Karplus – Strong synthesis (1983, 1987) Jaffe and Smith (1983)	Loops of short waveform through a filtered delay to simulate a plucked string, drum timbers, knocking, applause etc.
	Formant synthesis	Rule based synthesis; see tables 3, 4
Other available variants. Some methods belong to more than one category.		

hitting the cymbal the rise time is extremely short. Other similar musical instruments have different waveform shapes. Generating sounds by means of analogue, and/or digital and the dominating software techniques leads by proper design to the desired imitated shapes.

Synthesizers can be categorized into two groups – synthesizers with built - in synthesis modules, called “performance synthesizers” and “modular synthesizers” with modules that can be connected in different forms. There are many techniques of sound synthesis and some are given in table 5. Tables 4, 5 can be arranged following Klatt: 1. Development of speech synthesizers (1939-1982); 2. Segmental analysis by rule (1959-1968); 3. Synthesis by rule of

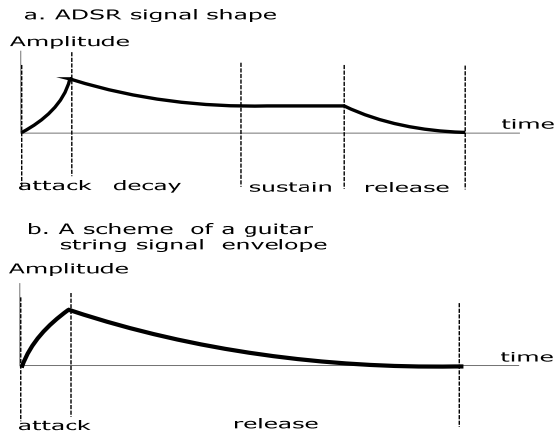


Figure 3: A scheme of ADSR decay curve.

segments and sentence prosody (1968-1980); 4. Fully automatic text-to-speech conversion (1968-1985); and in addition: 5. Computerized speech technology, enhanced statistics, huge data bases, linguistics (1985-to-date).

6 A note about Granular synthesis

Isaac Beeckman (1588-1637) was a physicist, known especially for notes about new modern concepts of inertia, free fall, and in acoustics – the rule that the fundamental frequency of a vibrating string is inversely proportional to its length. He never published his scientific work and de Ward (1939-1953) rediscovered his lost diary and notes from which we know about his work. See also van Berkel [22] and Cohen [23] was first to present the concept of sonic granulation, by assuming that “quanta” of sonic data traveled through air. Gabor [24] introduced the “quantum physics theory of granular sound”. He stated that any sound could be synthesized by matching a set of large number of simple sound grains, which is the basis of granular synthesis. The gain is built by an instrument that generates by equation an envelope of a narrow Gaussian curve that drives a sine oscillator. The grain itself is a unit of audible energy of low sound level, possessing any waveform and duration.

The grain signals are measured along the time axis as a unit, but can be decomposed into temporal information – duration (of a few milliseconds – usually 10-50 ms), envelope and waveform shapes and location along the time axis, also spectral information in the frequency domain, which is the frequency of the waveform within the grain. Grain frequency means the rate of repetition of the waveform in the grain envelope. The granular synthesis is the assembling of the short wave signals of the individual grains into the desired large scale acoustic signals (macro sound: music, etc.). Synchronous granular synthesis applies approximately uniform grain density (spacing) that can generate formants

of unchanging pitch. The grains are statistically scattered in “clouds” of certain band widths, carrying the grain parameters. The grain composition is its shape or envelope – amplitude and duration (functional, constant, or random). A typical grain shape is shown in fig. 4. The duration and the band width are inversely proportional and interpolation over the duration of the cloud can change the grain wave form for each grain. Some grain properties are given in table 6.

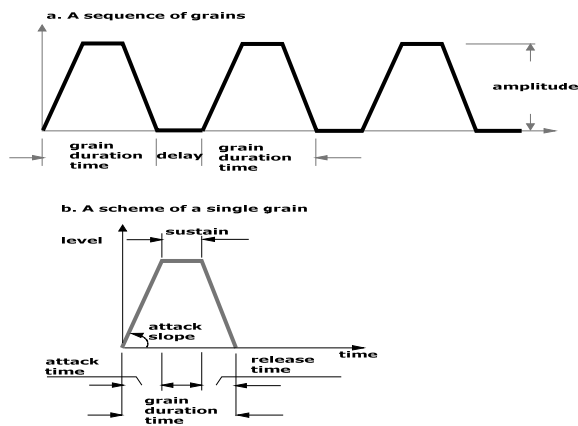


Figure 4: A typical grain shape.

Table 6: The effect of grain properties on the acoustic impression.

Grain property	Frequency band width	Acoustic result
Short duration		“Explosive” sound
Long duration		Smooth perception
Sparse grain density		Points form
High grain density	Narrow	Pitched stream with formants spectra
	Wide	Full sound distribution

Although the shape of the grain envelope changes its spectral content, it is the waveform that influences perception and not the amplitude. Hence, the relations among the attack, sustain, release and duration of the grain are responsible for the listeners’ awareness, while the frequency depends on the duration and delay between consecutive grain occurrences. Practically it means that changing the slope of a regular micro-sound varies the resulting spectral content of the grain and steeper “attack” causes broader band width.

The next step, following Beeckman’s physical form was contributed by Norman Wiener [25]. He used quantum physics to show that sound could be minimized to “fundamental” particles (“atoms”). Following this theoretical approach, researchers developed the concepts behind granular synthesis. Some of them became “milestones” in grain synthesis. The Greek composer and architect Iannis Xenakis (1922-2001) postulated after Gabor that all sound was integration



of grains of elementary sonic particles of some quanta (Xenakis [26] and see also Capanna and Xenakis [27] and Xenakis and Kanach [26]). His way to use this term is in accordance with the definition of the “mechanical” granular synthesis: A method by which electro-magnetic tape sound is cut into small grains that are then organized in a different order to form other sounds, or sonic textures. Roads [28] is considered to be the first to implement an instrument for digital granular synthesis, after he studied Xenakis’ approach and installed it in the computer. In 1985 Roads introduced the theory of granular synthesis based on acoustic quanta as defined in Gabor’s theory and Xenakis’ compositional theory of sound grains. Roads’ 1996 survey of computer music is an important contribution, with an extensive reference to analyses, techniques and different applications of granular synthesis.

Influenced by Roads [28], Truax [29] became a pioneer in real time synthesis. Truax dealt with implementation of real time granular synthesis for the DMX-1000 signal processor. This achievement was realized in 1986. He also introduced a system for a time varying parameter synthesis. Truax explained the distinction between delay time, sampled sound, and granular synthesis, and showed various applications of granular synthesis to pitch and time based transformations. The huge amount of grains per second (thousands) that is involved in each composition necessitates computational assistance, as, for example by FOF module in the program CSound.

Today, specific synthesis programs that run on operating systems such as Windows of Microsoft, Macintosh and UNIX are available. This sophistication is necessary for high quality musical compositions through control of the many events that generate continuous heard sound due to the sequence of grains. Tape recorders did not suffice for that goal since it demanded laborious work for special tasks. This difficulty was eased by the appearance of digital synthesis that became a part of the grain synthesis. For example, quadraphonic amplification by granular synthesis is common due to its remarkable capabilities of spatial positioning. The signal from the grain oscillator is processed through an amplifier for spatial positioning of the grain, which means that each grain can be located arbitrarily in the stereo field. The typical duration of a grain is somewhere between 5 and 100 milliseconds and it is considered as a point of sound. If the duration of the grain is less than 2 milliseconds it will be perceived as a click. Roads [28] recommends 20 ms for grain duration and even 10 ms. Yet, Truax [29] claims that in order to make the grain perceived as a continuous texture, its maximum duration should be about 50 ms. A pulsating rhythm can be heard in the background if the delay time between the grains reaches 60 ms. However, Truax has performed examples of 15 ms duration and 5 ms delay, which means that the ratio of duration to delay time is 3.

The most musically important aspect of an individual grain is its waveform. The variability of waveforms from grain to grain plays a significant role in the flexibility of granular synthesis. Fixed-waveforms (such as a sine wave or saw wave), dynamic-waveforms (such as those generated by FM synthesis), and even waveforms extracted from sampled sounds may be used within each grain. The result of combining a huge amount of grains can generate “clouds” of grains,



which are masses of sound, consisting of micro-sounds that are distributed statistically. A simple granular “cloud” may consist of only a handful of particles, but a sophisticated “cloud” may be comprised of a thousand particles or more. Hence, a vast amount of processing power is required to perform granular synthesis. Also, real-time granular synthesis requires a supply of grain-generating devices.

Several currently available microcomputers are capable of implementing real-time granular synthesis, but the high cost of these machines might be prohibitive.

The flexibility of granular synthesis is very large, and in addition, there are variants of granular synthesis, for example: Wavelet (grainlet) synthesis duration is determined by its pitch, and it starts and ends at zero phase; in glisson synthesis the contents of each grain are modified by glissando, and in each grain of pulsar synthesis a pulsar is generated by an impulse generator. Vibrato can be achieved by amplitude modulation of the main oscillator with a low frequency oscillator. Two oscillators with similar frequencies, acting simultaneously, yield a special musical effect (as known from hearing theories).

7 Conclusion

Biomimesis of sound production recognition and synthesis are consequence of evolutionary development of communication means. Sound has served since the beginning and up to now for religious purposes, including diverse ways of expression that emerged along this development in human communication (for example, the mimesis of natural sounds, use of musical scales, and especially harmonics in classical music, and speech and music synthesis by electronic means). As sheer interest in science and technology increased in the second half of the 19th century, new innovations in this area have yielded applications for spoken and sung communication, as well as in all areas of engineering and medicine. The AV devices are a recent example of these innovations.

References

- [1] Rosenhouse, G., *Active Noise Control – fundamentals for acoustic design*, Vol. I. WIT Press: Southampton and Boston, 2001; Rosenhouse, G. *Deliberate Sound and Vibration Control – Acoustic Sources and Networks in Biology, Biomimesis and Intentional Acoustics*. Vol.II, WIT Press: Southampton and Boston (to appear in 2010).
- [2] Nathan, R., Getz, W.M., Revilla, E. Holyoak, M., Kadmon, R, Saltz, D. & Smouse, P.E., Movement ecology paradigm for unifying organismal movement research. *Proc. Natl Acad. Sci. USA*, 105, pp. 19052-19059, 9.12.2008.
- [3] Fant, G., *Acoustic Theory of Speech Production*, Monton, The Hague: Netherlands, 1960, Fant, G., Speech technology and research retrospect and prospect. *Interspeech 2005*, Lisbon, 6.2005.



- [4] Stevens, K.N., *Acoustic Phonetics*, The MIT Press: Cambridge MA, London, 1998 (paper back, 2000), 628 pages; Stevens, K.N., Acoustics of speech and hearing, 6.551J/HST714J, lecture 22, 30.11.04.
- [5] Räsänen, J.O., Rosenhouse & G., Gavrieli, N., Effects of diameter, length, and circuit pressure on sound conductance through endotracheal tubes. *IEEE Trans. of Biomed. Engng*, **53**(7), pp. 1255-1264, 2006.
- [6] Broad, W.J., *The Oracle: The Lost Secrets and Hidden Message of Ancient Delphi*, Penguin Press: NY, 2006.
- [7] Flanagan J., *Speech Analysis, Synthesis, and Perception*, Springer-Verlag: Berlin, Heidelberg and New York, 1972.
- [8] Dudley, H. & Tarnoczy, T.H., The speaking machine of Wolfgang von Kempelen. *J. Acoust. Soc. Am.*, **22**(2), pp. 151-166, 1950.
- [9] Pompino – Marschall, B., Von Kempelen's contribution to the theory of acoustic articulation. *Grazer-Linguistische-Studien*, **62**, pp. 137-147, 2004; Pompino – Marschall, B., Von Kempelen, et al. *ZAS papers in Linguistics*, **40**, pp. 145-159, 2005.
- [10] Helmholtz, H.L.F., *The Sensation of Tone*, Dover publications: NY, pp. 121, 122, 1954; The first addition appeared in 1961.
- [11] Chiba, T. & Kajima, M., *The Vowel, Its Nature and Structure*, Tokyo – Kaiseikan Pub. Co.: Tokyo, 1941.
- [12] Arai, T., The replication of Chiba And Kajima's mechanical models of the human vocal cavity, *J. Phonetic Soc. Japan*, **5**(2), pp. 31-38, 2001.
- [13] Grützmacher, M. & Lottermoser, W., Über ein Verfahren zur Trägheitsfreien Aufzeichnung von Melodikurven. *Akustische Zeitschrift*, **2**, pp. 242-248, 1937.
- [14] Stewart, J.Q., An electronic analogue of the vocal organ. *Nature*, **110**, pp. 311-312, 1922 (words: mama, anna, wow-wow, yi-yi)
- [15] Wagner, K.W., Ein neues elektrisches Schprechgerät Nachbildung der menschlichen Vokale, *Preuss, Akad. Wiss. Berlin. Abh.*, **2**, s. 1-44, 1936.
- [16] Dudley, H., Riesz, R.R. & Watkins, S.A., A synthetic Speaker. *J. Franklin Institute*, **227**, pp.739-764, 1939; Dudley, H., The Vocoder. *Bell Labs Record*, **1**, pp.122-1267, 1939a; Dudley, H., Remaking speech. *J. Acoust. Soc. Am.*, **11**(2), pp. 169-177, 1939b.
- [17] Fletcher, N.H. & Rossing, T.D., *the Physics of Musical Instruments*, 2nd ed. Springer Verlag: NY, 1998.
- [18] Rossi, M., *Acoustics and Electro acoustics*. Artech House: Norwood, MA, 1988.
- [19] Eargle, J.M., *Handbook of Recording Engineering*, 3rd ed., Chapman & Hall and ITP: NY, 1996
- [20] Russ, M., *Sound Synthesis and Sampling*, Focal Press: Oxford, UK., 1996
- [21] Ahnert, W. & Steffen, F., *Sound Reinforcement Engineering*, E & FN Spon: London, 1999.
- [22] van Berkel, K., *Isaac Beeckman (1588-1637) en de mechanisering van het wereld beeld*, Amsterdam, 1983 (Isaac Beeckman (1588-1637) on the mechanization of the world picture).



- [23] Cohen, M. (ed.), *Isaac Beeckman*, D. Riedel Publishing Company: Dordrecht, The Netherlands, 1990.
- [24] Gabor, D., Acoustical quanta and the theory of hearing. *Nature*, 15, pp.591-594, 1947.
- [25] Wiener, N., Spatio-Temporal Continuity, Quantum Theory and Music, in *The Concept of Space and Time*. D. Reidel Publishing Company: Dordrecht, pp. 539-546, 1964.
- [26] Xenakis, I., *Formalized Music – Thought and Mathematics in Composition*, Bloomington University Press: Indiana, 1971, 273 pages. Translated into English by C. Butchers, G.H.Hopkins and J., Challifour. New revised edition – compiled, edited and translated by Sharon Kanach, Stuyresant, NY, Pendragon Press, 1992, 387 pages; Xenakis, I.& Kanach, S., *Music and Architecture*: Pendragon Press, 2008, 337 pages.
- [27] Capanna, A., Iannis Xenakis – Architect of light and sound. *Nexus Network J.*: Winter Spring, 3(1), pp. 19-26, 2001.
- [28] Roads, C., Automated Granular synthesis of sound. *Computer Music J.*, 2(2), pp. 61-62), 1978; Roads, C., Granular Synthesis of Sound. Reprinted in *Foundations of Computer Music*, C.Roads & J. Strawn (eds). MIT Press: Cambridge, MA, pp.145-159, 1985; Roads, C., Introduction to Granular Synthesis. *Computer Music J.*, 12(2), pp. 11-13, 1988; Roads, C., *Computer Music Tutorial*, The MIT Press: Cambridge, MA, pp. 168-184, 1996; Roads, C., *Microsound*, The MIT Press: MA, 2001 (NY ed.: 2004).
- [29] Truax, B., Real-time granular synthesis with digital signal processor. *Computer Musical J.*, 12(2), pp.14-25, 1988; Truax, B., Discovering inner complexity: Time shifting and transposition with real-time granulation technique. *CMJ*, 18(2), pp. 38-48, 1994.

