
1:Θφ4

Composer / researcher:

Dimitri Voudouris

Composed:

2007

Composition:

1: $\Theta_{\varphi 4}$

Duration:

12 min 28 sec

For:

Four artificial female voices

Index	Page

Artificial vocalization:

Written text is a transcription of our audible, but also of our inner, cognitive "vocalization." Although we occasionally transcribe speech directly, written text is first of all a transcription of our inner voice rather than our oral expression. Nevertheless, this inner voice is still deeply connected to speech. Written text is an extension of, or a tool for, our voice. It would seem that our voice is not just the part of our expression that we register with our ears and hear. To explore the connection between voice and code,between our physical vocal expression and programming languages, means that we first have to look at the connection between our vocal expression, our expression in written text and other notation systems (such as notation systems for music and song) and the meanings we have ascribed to these expressions. Given that our voice has been "notated" in other ways as well since the advent of recording media, we have to also consider the influence of recorded and artificial voices on how we perceive and express ourselves. Artificial voices and the subliminal languages of machines have changed our perception of voice even further. Our voice has become a complex of oral, written, and machinic expression.

A great deal of research has gone into voice coding, though most of it has focused on efficient coding of speech. The most common implementation in use today is Code- Excited Linear Prediction (CELP), which describes a family of coding techniques based on Linear Predictive Coding (LPC) for the calculation of filter coefficients, roughly corresponding to the resonant (wideband) response of the vocal tract, along with modeling of the excitation vector,roughly corresponding to the glottal excitation function produced by the vocal folds. Although both are produced using the same mechanisms,perceptually important qualities of the singing voice are quite different from that of speech. These differences make speech codecs less applicable for highquality transmission of singing. This research investigates these differences and explores ways of modifying a standard CELP coder to better transmit the singing voice. Specifically, data taken from singing as opposed to speech is used to create codebooks that are more applicable for singing. The resulting codec removes some of the artifacts generated when using CELP encoding for singing, providing toll quality transmission of singing while maintaining an overall low bitrate. In singing, each note that is sung is fairly constant and quantized in pitch, as opposed to speech, in which pitch varies unpredictably and continuously. CELP excitation model is not ideally suited for the singing voice because speech contains a fairly even mixture of voiced and unvoiced sounds, singing is almost entirely voiced. A comparison of the residuals after the determination of pitched and stochastic excitations for a singing voice signal reveals that the stochastic codebook does little to reduce the residual error in singing. The residual error data taken from recorded singing was used to replace the stochastic codebook. Approximately one continuous minute of sung material from an individual singer was used to create the new codebook. A Principle Components Analysis (PCA) of the residual subframes was performed using singular value decomposition, and the eigenvectors corresponding to the highest 128 eigenvalues were selected for the codebook. This is a much different type of codebook than the one used in the standard CELP coder. This codebook consists of basis vectors which are combined to model the individual excitations.

.

Composition:

Four languages were chosen to represent the composition of 1:Θφ4 [for four female voices] namely Greek, Spanish, Portuguese and Italian. I focused on these languages because of the various differences in voiced and unvoiced sounds that are produced whilst singing, plus the expressive richness associated with traditional and ancient patterns of such singing techniques. A combination of 16 vowel to consonant formations and visa versa arrived at approximately 2% African dialect . I centred my research on dialect and selected vowel sounds which form expressions noted in throat sounds and various linguistic patterns of speech such as utterances, turns, intonational and phonological phrases. Cross references between the languages with the location of the word in a larger prosodic domain was noted. Words and phrases were selected and designed from basic artificial language parameters . These selections were necessary in designing words or phrases that had a singing and expressive quality.

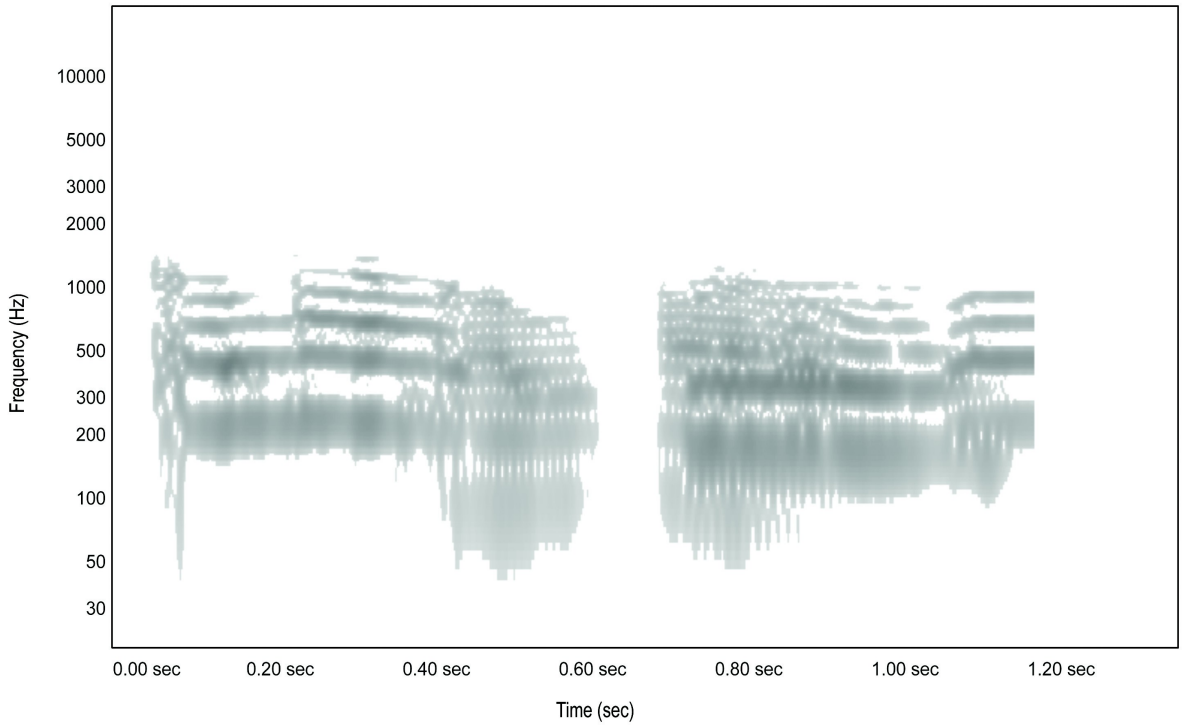

Diagram 1: The spectrogram is of atrificial consonant to vowel transition prior to processing [*formed from four language groups*] composed of two syllables.
The sound has African consanant *clicking*
as noted at 0.00 sec -0.1 sec.

The words created where tested under realistic circumstances they were spoken and sung with precision breaking the silence with prosodic utterance.
Prior to voice processing and synthesis various acoustic patterns and regularities were examined. E.g. The timbral consequences of higher larynx analysed by the use of a listening test synthesized ascending scales. The aim was to assess the perceptual relevance of three acoustic consequences that could be expected to accompany a rise of the larynx.

- 1] a small increase in the formant frequencies
- 2] a decrease in of the vibrato extent
- 3] a decrease in the level of voice source fundamental

Through this analysis I came to the conclusion that the first formant frequencies must be avoided implying that when this is increased that it is not lower than the fundamental. This is the reason of the pitch dependent jaw and lip opening that can be observed in female vocalists when they sing at high pitches.

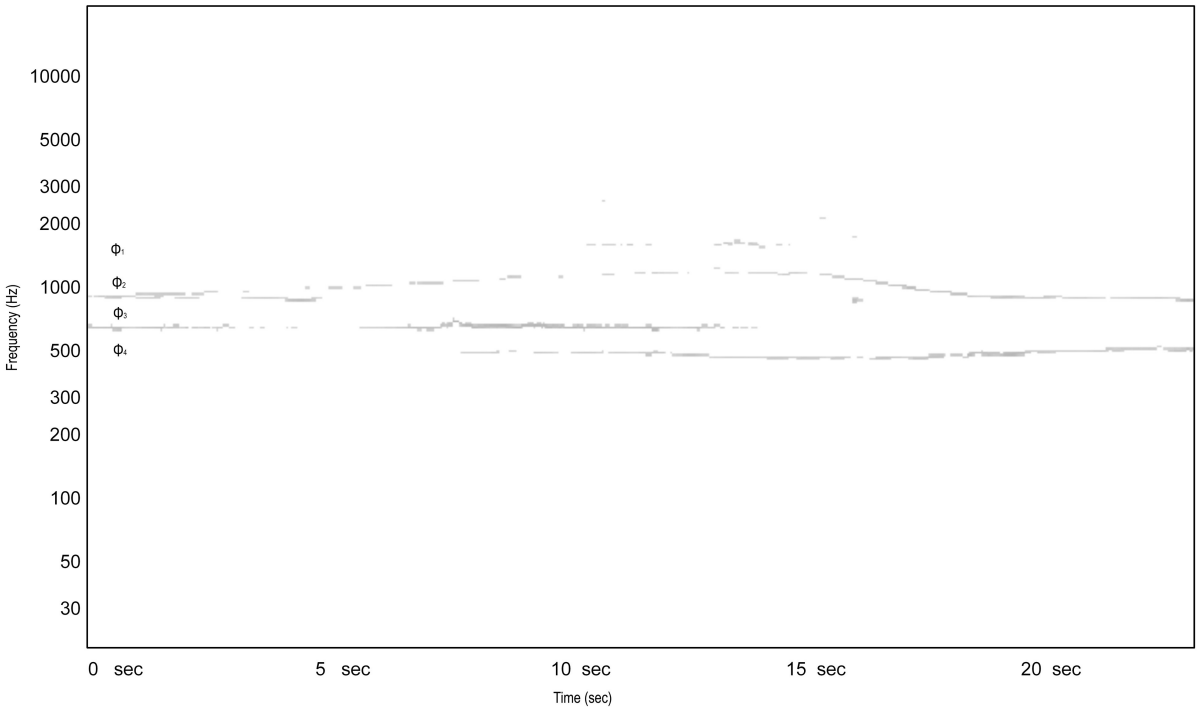

Diagram 2 : The spectrogram showing four female voices (Φ1-Φ4) in the processing of 1:Θ*φ4*

Simple cyclical patterns of singing were adopted and processed resulting in the annotated diagram below: singing language of voice Φ_1 connects at point x on voice Φ_2. Voice Φ_2 connects at point x on voice Φ_1 etc.

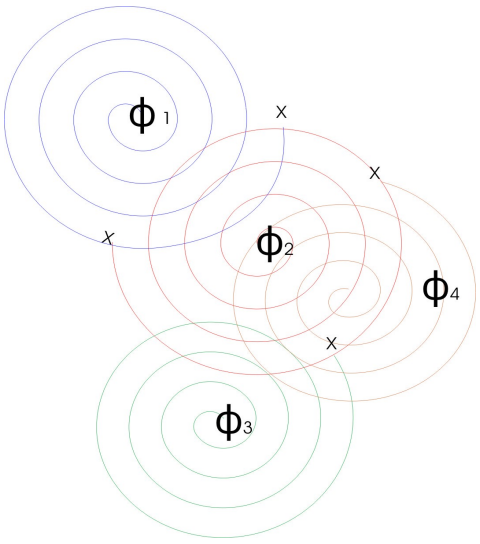

Diagram 3 : One of 16 variations of **c**yclical pathways of four voices

This is one of 16 variations in cyclical patterns were the various voices follow prepared pathways releasing each package of information at points x and go through a variety of transforming alterations.

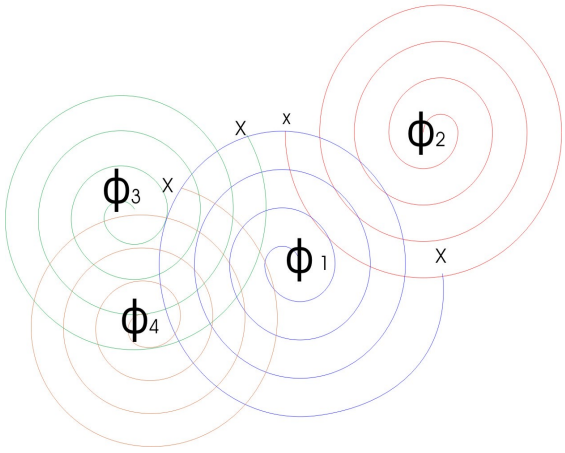

Diagram 4 : Second of 16 variations of cyclical pathways of four voices

The properties of spaces had varying reverberation and echo properties. The vocalists were placed at various audible distances and with alterations done in pitch and amplitude with respect to properties of space. This analysis latter helped me in selecting certain artificial words and phrases that were used in singing synthesis, the results were plotted graphically and through Matlab I was able to analyse, select and eliminate vocal expressions in the synthesis of 1:$\Theta\varphi 4$.

References :

[1] J. Sundberg. *The Science of the Singing Voice*. Dekalb, IL:Northern Illinois University Press, 1987.
[2] N. J. Miller. *Filtering of Singing Voice Signal from Noise by Synthesis* . Unpub. Ph.D. thesis. Univ. of Utah, 1973.
[3] R. McAulay and T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. On Acous., Speech, and Sig. Proc.*, vol. 34, pp. 744-754, 1986.
[4] P. R. Cook. *Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing* . Unpub. Ph.D thesis. Stanford U., 1990.
[5] P. Lansky and K. Steiglitz, "Synthesis of Timbral Families by Warped Linear Prediction," *Computer Music Journal*, vol. 5, no. 3, pp. 45-49, 1981.
[6] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals* . Englewood Cliffs, NJ: Prentice-Hall, 1978.
[7] Hunt, A., Wanderley, M. and Paradiso, M."The importance of parameter mapping in electronic instrument design", *Proceedings of the 2002 Conference on New Interfaces for Musical Expression (NIME-02)*, Dublin, Ireland, 2005.
[8] Tian, Y., Kanade, T. and Cohn, J. "Robust Lip Tracking by Combining Shape, Color and Motion", *Proceedings of the 4th Asian Conference on Computer Vision (ACCV'00)*, 2000.

END