Voice modification by source filter model-based modified short-time Fourier transform magnitude

Alfred Hofmann1, Ursula Barth1, Ingrid Beyer1, Christine Günther1, Frank Holzwarth1, Anna Kramer1, andErika Siebert-Cole1

1 Springer-Verlag, Computer Science Editorial, Tiergartenstr. 17,  
69121 Heidelberg, Germany

{Alfred.Hofmann, Ingrid.Beyer, Christine.Guenther, Frank.Holzwarth,Anna.Kramer, Erika.Siebert-Cole[}@Springer.com](mailto:%7d@Springer.com)

**Abstract.** The key issue in a specific voice modification is how to change target parameters while keeping other parameters constant. For time-scale modification (TSM), pitch modification (PM) and timbre modification (TM), the target parameters are respectively the tempo, pitch frequency and location and bandwidth of formants. This paper proposes a new method for voice modification by modified short-time Fourier transform magnitude (MSTFTM) based on source filter model (SFM). Different types of voice modification experiments using proposed method are performed. The results show that using SFM-based MSTFTM, a pretty good performance is achieved. Comparisons with another classic MSTFTM method in PM have verified that the method can modify the pitch without affecting the formants.

**Keywords:** voice modification, time scale modification (TSM), pitch modification (PM), timbre modification (TM), signal estimation.

1 Introduction

Voice modification is a technique which can be used to change the characteristics of various sound produced by a person. This field of speech technology can contribute greatly to the Entertainment industry as well as increase diversity of the voice database for multiple speaker Text to Speech (TTS) systems. For example, a language learning system may need to reduce speaking rate so that the pronunciation is much clear, or a TTS system can change the original voice pitch so that a synthetical speech uttered by a man will sound as if spoken by a woman.

Voice modifications are usually referred to as prosodic modifications including four main types: time scale modification (TSM), pitch modification (PM), timbre modification (TM) and intensity modification (IM). The greatest challenge in TSM is to change the audio rate, while preserving other characteristics such as pitch and timbre. The goal of PM is to change the fundamental frequency in order to compress or expand the spacing between the harmonic components in the spectrum while preserving the short time spectral envelope as well as the time evolution. The aim of timbre modification (TM) is to change the locations and bandwidths of formants while keeping the same pitch. IM can be easily achieved by associating an intensity scale factor at each analysis time instant of a signal. Several approaches have been proposed for voice modification. Such approaches include synchronized overlap and add algorithm (SOLA)[1], overlap-add technique based on waveform similarity (WSOLA)[2], phase vocoder method and its refinement [3-4], peak alignment overlap-add algorithm (PAOLA)[5], etc. However, when in PM, above methods change the formants of a voice. PM and TM are mixed together. In the process of changing the pitch of a signal to sharp or flat, either with or without keeping the original audio file length, the sample rate of the audio signal is altered thus changing the fundamental frequency along with all harmonics and spectral envelope. As a result, pitch is changed as well as the locations and bandwidths of formants, which we need to avoid in some applications. Similar cases also happen in TM.

The source filter model (SFM) is a model of voice where the spoken word is comprised of a source component originating from the vocal cords which is then shaped by a filter imitating the effect of the vocal tract. This model of voice production is linear and assumes superposition holds. TSM using short time Fourier Transform (STFT) has been proposed by Portnoff[6]. Griffin and Lim developed an algorithm for signal estimation from modified short-time Fourier transform (MSTFT) and modified short-time Fourier transform magnitude (MSTFTM) [7]. Xinglei, et al improved the real-time performance of Griffin and Lim’s method [8].

The rest of the paper is organized as follows. Section 2 briefly reviews the source filter model. Section 3 introduces an improved MSTFTM. TSM, PM and TM using SFM-based MSTFTM respectively proposed in section 4. Finally, conclusions are drawn in the last section.

2 Source filter model

The printing area is 122 mm × 193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm, if possible.

Use 10-point type for the name(s) of the author(s) and 9-point type for the address(es) and the abstract. For the main text, please use 10-point type and single-line spacing. We recommend the use of Computer Modern Roman or Times. Italic type may be used to emphasize words in running text. Bold type and underlining should be avoided.

Papers not complying with theproceedings’style will be reformatted. This can lead to an increase in the overall number of pages. We would therefore urge you not to squash your paper.

Headings.Headings should be capitalized (i.e., nouns, verbs, and all other words except articles, prepositions, and conjunctions should be set with an initial capital) and should, with the exception of the title, be aligned to the left. Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalized. The font sizes are given in Table 1.

Here are some examples of headings: "Criteria to Disprove Context-Freeness of Collage Languages", "On Correcting the Intrusion of Tracing Non-deterministic Programs by Software", "A User-Friendly and Extendable Data Distribution System", "Multi-flip Networks: Parallelizing GenSAT", "Self-determinations of Man".

**Table 1.**Font sizes of headings. Table captions should always be positioned *above* the tables.

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
| Title (centered) | **Lecture Notes …** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Headings.** Text follows … | 10 point, bold |
| 4th-level heading | *Remark.* Text follows … | 10 point, italic |

Lemmas, Propositions, and Theorems.The numbers accorded to lemmas, propositions, and theorems, etc. should appear in consecutive order, starting with Lemma 1, and not, for example, with Lemma 11.

2.1 Figures

Please check that the lines in line drawings are not interrupted and have a constant width. Grids and details within the figures must be clearly legible and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). The lettering in figures should have a height of 2 mm (10-point type). Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between the margins and set them in 9-point type (Fig. 1 shows an example). The distance between text and figure should be about 8 mm, the distance between figure and caption about 6 mm.

To ensure that the reproduction of your illustrations is of a reasonable quality, we advise against the use of shading. The contrast should be as pronounced as possible.

If screenshots are necessary, please make sure that you are happy with the print quality before you send the files.

Remark 1.*In the printed volumes, illustrations are generally black and white (halftones). If you send colored figures that are to be printed in black and white, please make sure that they really are legible in black and white. Some colors show up very poorly when printed in black and white.*



**Fig.1.** One kernel at *xs* (*dotted kernel*) or two kernels at *xi* and *xj* (*left and right*) lead to the same summed estimate at *xs*. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption.

2.2 Formulas

Displayed equations or formulas are centered and set on a separate line (with an extra line or halfline space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within each section or within the contribution, with numbers enclosed in parentheses and set on the right margin.

|  |  |
| --- | --- |
| x + y = z . | (**1**) |

Please punctuate a displayed equation in the same way as ordinary text but with a small space before the end punctuation.

2.3 Footnotes

The superscript numeral used to refer to a footnote appears in the text either directly after the word to be discussed or – in relation to a phrase or a sentence – following the punctuation mark (comma, semicolon, or period). Footnotes should appear at the bottom of the normal text area, with a line of about 5cm set immediately above them[[1]](#footnote-2).

2.4 Program Code

Program listings or program commands in the text are normally set in typewriter font, e.g., CMTT10 or Courier.

Example of a Computer Program from Jensen K., Wirth N. (1991) Pascal user manual and report. Springer, New York

program Inflation (Output)  
 {Assuming annual inflation rates of 7%, 8%, and  
 10%,... years};  
constMaxYears = 10;  
var Year: 0..MaxYears;  
 Factor1, Factor2, Factor3: Real;  
 begin  
 Year := 0;  
 Factor1 := 1.0; Factor2 := 1.0; Factor3 := 1.0;  
WriteLn('Year 7% 8% 10%'); WriteLn;  
 repeat  
 Year := Year + 1;  
 Factor1 := Factor1 \* 1.07;  
 Factor2 := Factor2 \* 1.08;  
 Factor3 := Factor3 \* 1.10;  
WriteLn(Year:5,Factor1:7:3,Factor2:7:3,  
 Factor3:7:3)  
 until Year = MaxYears  
end.

2.5 Citations

The list of references is headed “References” and is not assigned a number. The list should be set in small print and placed at the end of your contribution, in front of the appendix, if one exists. Please do not insert a pagebreak before the list of references if the page is not completely filled. An example is given at the end of this information sheet. For citations in the text please use square brackets and consecutive numbers: [1], [2], [3], etc.

2.6 Page Numbering and Running Heads

There is no need to include page numbers. If your paper title is too long to serve as a running head, it will be shortened. Your suggestion as to how to shorten it would be most welcome.

Acknowledgments. The heading should be treated as a 3rd level heading and should not be assigned a number.

References

1. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121

2. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, BerlinHeidelbergNew York (1997) 415–438

3. van Leeuwen, J. (ed.): Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, Vol. 1000. Springer-Verlag, BerlinHeidelbergNew York (1995)

4. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn.Springer-Verlag, BerlinHeidelbergNew York (1996)

Checklist of Items to be Sent to Us

1. A final Word or RTF file
2. A final PDF file
3. A copyright form, signed by one author on behalf of allthe authors of the paper
4. A readme giving the first name(s) and the surname(s) of all the authors of the paper as well as the name and email address of the corresponding author

1. The footnote numeral is set flush left and the text follows with the usual word spacing. [↑](#footnote-ref-2)