Evaluation of pitch coding alternatives for vibrotactile stimulation in speech training of the deaf

I.L. Barbacena, A.T. Barros¹, R.C.S. Freire², E.C.A. Vieira¹
¹CEFET/PB, João Pessoa - PB, Brazil, iltonlb@gmail.com
²DEE, UFCG, Campina Grande-PB, Brazil, e-mail: rcsfreire@dee.ufcg.edu.br

Abstract. Use of vibrotactile feedback stimulation as an aid for speech vocalization by the hearing impaired or deaf is reviewed. Architecture of a vibrotactile based speech therapy system is proposed. Different formulations for encoding the fundamental frequency of the vocalized speech into the pulsed stimulation frequency are proposed and investigated. Simulation results are also presented to obtain a comparative evaluation of the effectiveness of the different formulated transformations. Results of the perception sensitivity to the vibrotactile stimulus frequency to verify effectiveness of the above transformations are included.

1. Introduction
Speech feedback is an indispensable stimulus in the speech learning phase of children. In the case of the born deaf there is a need for an alternative form of feedback stimulus of the words spoken by the deaf child [1],[2],[3].

It has been reported in literature [2], [3], [4], [5] that the speech signal of the individuals with hearing impairment or no hearing at all in general contain higher fundamental frequency component than that of the individuals with normal hearing. Their speech also exhibits excessive pitch variability, resulting in an undesirable intonation. This in turn affects the quality and intelligibility of the resulting speech. Boone [8] has concluded that these limitations can be overcome by employing an external alternative feedback of the spoken words and train the individuals with impaired hearing to produce the correct desirable fundamental speech frequency.

At present many software packages are available which help in the training of the deaf speaker in producing a speech vocalization [1]. The visual feedback systems are usually video game type systems, having to hit the right target and an indication of the successful or correct hits, maintaining good motivation and interest of even the children in doing vocalization exercises repeatedly. A tactile stimulus feedback system similar to the visual feedback system can also be designed and employed, especially with the deaf blind children [4],[5]. The tactile feedback may also be combined with the visual feedback to yield a mixed system. There also exists a possibility of using the tactile feedback system separately as a portable device. Most of the tactile feedback stimulus devices employed as an aid in the training of hearing impaired or deaf individuals is the vibrotactile type [6],[7].

In this paper we investigate four different transformations to map the speech fundamental frequency into vibrotactile stimulation frequency. It was conducted a comparative evaluation of these transformations through simulation runs of the fundamental frequency vocalization, thus it was possible to investigate the vibrotactile frequency perception sensitivity of vibrotactile stimulus generated by the different proposed transformations, without influence of the ability of the volunteers.
in the control of his speech fundamental frequency. Results of perception tests were also conducted on a subject with normal hearing to validate the simulation studies. In a second part the frequency of the stimulator output pulse train is being modulated in accordance with the variation in the speech fundamental frequency. The principal objective of the present investigation is to compare the relative effectiveness of the proposed transformations for mapping of speech fundamental frequency $f_o(t)$ into stimulator output pulse train characteristics in the vibrotactile game environment.

The vibrotactile game environment for speech vocalization feedback training/learning is outlined in Figure 1. The learning procedure to control or produces the desired speech fundamental frequency employing the above transformations may be summarized as follows: the user is asked to speak and the fundamental frequency of his speech is detected and a corresponding pulsed stimulation with a certain frequency is applied to the user’s finger tip. As a response to this stimulus, the user modifies the tone of his speech in such a way as to produce the speech with target fundamental frequency $F_a$.

Figure 1. A typical speech vocalization feedback training/learning system

2. Materials and methods

2.1. Proposed Transformations

In a particular system, meant to be used in the speech training or speech therapy, two features are considered very important [1]. The range of variation of the vibrotactile stimulation frequency should be adjustable in accordance with the progress in the speech learning, on the part of the system user, and the therapist should be able to set and choose the system parameters to make the system more suitable to the needs of the individual being trained. As mentioned before, the visual feedback system contains modules to train the user to produce speech having a specified fundamental frequency and also having a feature for indicating the correct hits. There is also a provision for the adjustment of the target fundamental frequency, by the therapist [1].

In the proposed vibrotactile stimulation speech training system, there is a provision for adjustment of: 1) the range of the vibrotactile stimulation frequency determined by the degree of the difficulty in learning process (given by scale factor $S$ as in the Eqs. (1) to (4), and 2), the target fundamental frequency. We investigate linear (Eq. 1), logarithmic (Eq. 2), cubic (Eq. 3) and quadratic (Eq. 4) transformations of the voice fundamental frequency $f_o(t)$, into the stimulator output pulse frequency $f_t(t)$ as defined in Eqs. (1) to (4). These transformations are proposed in the context of a single channel vibrotactile stimulator being employed.

$$f_t(t) = F_c + [f_o(t) - F_a] \cdot S$$  
$$f_t(t) = F_c + \left[ \ln \left( \frac{f_o(t)}{F_a} \right) \right] \cdot S$$  
$$f_t(t) = F_c + [f_o(t) - F_a]^3 \cdot S$$  
$$f_t(t) = F_c + [f_o(t) - F_a]^2 \cdot S$$
where \( f_s(t) \) is the detected fundamental frequency of the speech signal. \( F_a \) is the specified target speech fundamental frequency and \( F_c \) is the value of \( f_s(t) \) when the user gets to produce the speech with the target frequency \( F_a \) and \( S \) is a scale factor indicative of the degree of the difficulty in the learning or training exercise which controls the range of the stimulation frequency.

The variation of the pulsed stimulus frequency \( f_s(t) \) is limited to a range from \( f_{\text{min}} \) to \( f_{\text{max}} \) and the detected fundamental frequency \( f_s(t) \) may vary from \( f_{o1} \) to \( f_{o2} \) which is set by either the user or the therapist. From Eqs. (1) to (4) we observe that \( S = 0 \) represents the highest degree of difficulty in the training/learning process since the feedback is independent of \( f_s(t) \) and \( F_s \). When \( S = S_{\text{max}} \) we obtain the maximum excursion in the pulse frequency \( f_s(t) \), for given values of \( f_s(t) \) and \( F_s \) characterizing the lowest degree of difficulty in the training exercises. Thus for given values of \( F_c \), \( F_a \) and \( f_{o2} \) defined as above, we can obtain expressions for \( S_{\text{max}} \) and \( f_{o1} \) for each of the proposed transformations.

Based on the results from psychophysics experiments, the system designer can choose the values of the stimulation frequency \( f_{\text{min}}, f_{\text{max}} \) and \( F_c \). The speech therapist can choose the upper limit of the voice fundamental frequency \( F_o2 \), and the target frequency \( F_a \) guided by the initial estimate of the fundamental frequency of the speaker's voice signal as well as the target frequency \( F_a \) from the expected evolution of the training/learning session. With these values in hand, and using expressions of the fourth column of Table I one can calculate the value of \( f_{o1} \) and the maximum value of the scale factor \( S \) (i.e. \( S_{\text{max}} = S_{1\text{max}} = S_{2\text{max}} \)) as given in columns 2 and 3 of Table I (\( S_{\text{max}} \) corresponds to the smallest degree of difficulty in the training/learning process).

| Transformation | \( S_{1\text{max}} = \) | \( S_{2\text{max}} = \) | For \( S_{1\text{max}} = S_{2\text{max}} \) |
|---------------|----------------|----------------|-----------------|
| Linear        | \( \frac{f_{\text{max}} - F_c}{F_a} \) | \( \frac{f_{\text{max}} - F_c}{F_a} \) | \( f_A = 2F_c - f_{o2} \) |
| Logarithmic   | \( \frac{f_{\text{max}} - F_c}{\ln \left( \frac{f_{o1}}{F_c} \right)} \) | \( \frac{f_{\text{max}} - F_c}{\ln \left( \frac{f_{o2}}{F_c} \right)} \) | \( f_A = \frac{F_c^2}{f_{o2}} \) |
| Cubic         | \( \frac{f_{\text{max}} - F_c}{(F_{o1} - F_c)^2} \) | \( \frac{f_{\text{max}} - F_c}{(F_{o2} - F_c)^2} \) | \( f_A = 2F_c - f_{o1} \) |
| Quadratic     | \( \frac{f_{\text{max}} - F_c}{(F_{o1} - F_c)^2} \) | \( \frac{f_{\text{max}} - F_c}{(F_{o2} - F_c)^2} \) | \( f_A = 2F_c - f_{o1} \) |

2.2. Vibrotactile Frequency Perception Sensitivity Experiment

Some experimental tests were conducted to quantify the perception sensitivity to vibrotactile stimulation frequency, with a view to evaluate the four transformations described above; in terms of ease of correct target hits, and comparison of performance of the learner for different transformations and for different values of scale factors in the transformations training/learning sessions. In order to study the comparative effectiveness of the different transformations and separate it from inherent ability of the individual in generating the speech with the desired fundamental frequency, we used a procedure which is schematically illustrated in figure 2.

The experimental procedure was done in two steps: the first consist in training the subject to learn to identify the vibrotactile stimulus of frequency \( F_c \) corresponding to the fundamental frequency target \( F_o \) having been achieved. Besides, he/she learns if the stimulus frequency \( f_s(t) \) is greater or smaller than \( F_c \). In the second step the trainee is tested for the capacity acquired during step one, to identify correctly the stimulation frequency \( F_c \). In this step we also store the identified target speech fundamental frequencies, \( f_s(n) \), lying between frequencies \( f_{o1} \) and \( f_{o2} \). This identification is done
through perception of the stimulation frequency $f_t(n) = F_c$. We also store the number of iterations, $n$, required to arrive at the correct identification of the target frequency $F_a$.

The variation of the fundamental frequency $f_o(t)$ is done in discrete steps and each step be defined as $\Delta f_o = (f_{o2} - f_{o1})/N$, where $N$ is the number of discrete frequencies used in the present procedure (figure 2). We used $N = 41$ in the first step and $N = 21$ in the second step. This was done to possibly eliminate the possibility of the subject under training to memorize the number of iterations required to achieve the correct target hit. The value of $N$ can also be chosen randomly between 21 and 41, but we did not use this in our tests.

![Figure 2. Experimental Procedure](image)

The procedure adopted in the second step is to generate a random fundamental frequency $f_o(n)$ in the range $f_{o1}$ to $f_{o2}$. The value of the size of the discrete increase or decrease step is calculated in the light of the number of discrete frequencies arbitrarily chosen for the test. The subject receives a constant amplitude pulsed vibrotactile stimulus $s(t)$ with the frequency $f_t(n)$ during 500 ms in which $n$ refers to the $n^{th}$ iteration. Depending on the subject’s perception of a particular stimulation frequency he/she can choose $k = 1$ (option 1, increase in $f_o(n)$) or $k = -1$ (option 2, decrease in $f_o(n)$). The trainee can also indicate that he/she has hit the target fundamental frequency $F_a$ (option 3). The value of the fundamental frequency identified as the pre-defined correct target frequency and the number of iterations ($n$) required to reach the correct target frequency are the two figures of merit that can be used to make a comparison of the performance of each transformations.
At present we are doing perception tests of fundamental frequency control without simulation of the control of intonation. The learning procedure to control or produces the desired speech fundamental frequency employing the above transformations may be summarized as follows: the user is asked to speak and the fundamental frequency of his speech is detected and a corresponding pulsed stimulation with a certain frequency is applied to the user’s finger tip. As a response to this stimulus, the user modifies the tone of his speech in such a way as to produce the speech with target fundamental frequency $F_a$. The system was developed using MATLAB with the basic characteristics showed in Fig. 3. Some experiments of perception are been made with this software to analyse the use of the proposed transformations studied in this work.

3. TEST RESULTS, DISCUSSION AND CONCLUSIONS

In the actual tests, during one test session, corresponding to each fundamental frequency $f_o$, a stimulation frequency is generated in accordance with each of the four transformations and $F_c$ is chosen to be 10 Hz for the quadratic transformation and 55 Hz to the other transformations. These tests where done only for $S_{1\max} = S_{2\max}$, and two different values of scale factors were employed giving a total of eight experiments during each session. The tests were repeated over 100 sessions. These tests were conducted on a volunteer, with large experience in vibrotactile stimulation. We chose: $f_{o2} = 400$ Hz, $f_{o1} = 0$ Hz and $F_a = 200$ Hz, $f_{\text{min}} = 10$ Hz and $f_{\text{max}} = 100$ Hz.

The results of these tests are shown in Table II, where the percentage mean error in $f_o(t)$ and the average of the number of iteration are, respectively, done by:

$$E_f = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{f_o(i) - F_a}{F_a}$$

$$N_i = \frac{1}{N_s} \sum_{i=1}^{N_s} n(i)$$

where $N_s$ is the number of sessions (in this case $N_s = 100$), $n(i)$ and $f_o(i)$ are the number of iterations used and the fundamental frequency chose by the volunteer in the $i^{th}$ session (Fig. 2).
We observe from table I that the smaller percentual errors are obtained from the quadratic transformation and the number of iterations is larger for cubic transformation. For the quadratic transformation, the $f_t(t)$ versus $f_o(t)$ curve has an inflection in the $F_a$ point. The better result of this transformation occurs because the trainee can achieve the target by the perception of the inflection and not only by the perception of $f_t(t) = F_a$.

The cubic transformation has a larger number of iterations because the $f_t(t)$ versus $f_o(t)$ curve of this transformation has a region of uncertainty around $F_a$.

We conclude by experimental results that the quadratic transformation had a better performance than other transformations because this transformation has an inflection in the target fundamental frequency. We also conclude that the transformations with different slopes after and before the target fundamental frequency lead to a fast decision of target achievement. The ability to control the speech fundamental frequency is a factor that can aid a trainee to achieve the target frequency, which can imply in the possibility of development of games with diverse characteristics and degree of difficulty.

| Transformation Type | $E_{f_o}$ | $N_{it}$ | $E_{f_o}$ | $N_{it}$ |
|---------------------|-----------|----------|-----------|----------|
| Linear              | 11        | 7.7      | 22.2      | 11.9     |
| Quadratic           | 3.5       | 6.1      | 8.7       | 12.5     |
| Cubic               | 36.4      | 13.7     | 38.2      | 21.3     |
| Logarithmic         | 12.3      | 8        | 24.9      | 6.7      |

At present we are doing some perception experiments based on a vibrotactile system implemented on MATLAB, using the transformations proposed in this work.

ACKNOWLEDGMENT

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES/SETEC/PIQDTEC for the support in the form of research and study fellowships during the course of these investigations.

REFERENCES

[1] J. Mahshie, “The use of sensory aids for teaching speech to children who are deaf”, In: Plant, G & Spens Karl-erik (org.) Profound deafness and speech communication, (Whurr Publishers, Londres, chapter 21, 461-491, 1995).
[2] I. R. Summers, “Single Processing Strategies for Single Channel Systems”, In: Summers, I.R. (org) Tactile Aids for The Hearing Impaired, (Whurr Publishers, Londres, chapter 5, 110-127, 1992).
[3] I.R. Summers, and D.A. Gratton, “Choice of Speech Features for Tactile Presentation to the Profoundly Deaf”, IEEE Transactions on Rehabilitation Engineering, 3(1), (1995).
[4] L.E Bernstein, S.P Eberhardt, and M.E. Demorest, “Single-channel vibrotactile supplements to visual perception of intonation and stress”, The Journal of Acoustical Society of America 85(1), 397-405 (1989).
[5] M. Rothenberg, & R.D Molitor, “Encoding voice fundamental frequency into vibrotactile frequency”, The Journal of the Acoustical Society of America 66(4), 1029-1038 (1979).
[6] M.H. Goldstein and A. Proctor, “Tactile aids for profoundly deaf children”, Journal of Acoustical Society of America 77(1), 258-265 (1985).

[7] L. E Bernstein, “Toward Future tactile Aids”, In: Plant, G & Spens Karl-erik (org.) Profound deafness and speech communication, (Whurr Publishers, Londres,chapter 8, 147-162, 1995).

[8] D. Boone, A voz e a terapia vocal (The voice and the speech therapy), (Editora Artes Médicas, São Paulo, Brazil, 1994).