INFLUENCE OF TIME-VARYING PITCH ON TIMBRE: "COHERENCE AND INCOHERENCE" BASED ON SPECTRAL CENTROID

S. Arthi and T. V. Sreenivas

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, 560012

ABSTRACT

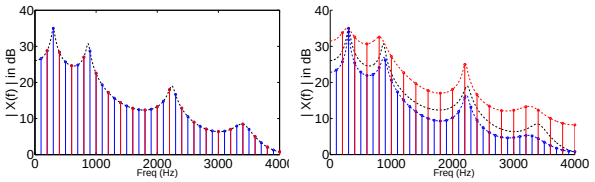
In this paper, the effect of time-varying pitch-glides on timbre perception of vowels is studied experimentally. While earlier studies show that stationary pitch is observed to be a significant dimension of timbre, we explore the effect of non-stationary pitch changes on the perception of vowel timbre using synthesized vowels. In each trial, three listeners rate perceptual timbre change using a MUSHRA-like (MUltiple Stimuli with Hidden Reference and Anchor) methodology. Results show that perceived timbre change is affected by the pitch-shift and nature of pitch-glide. Timbre change is observed to be monotonic with pitch-shift for all types of pitch-glides, implying larger the pitch-shift, higher the perceived timbre change. When the excitation is a pitch-glide, the vowel spectral centroid and the nature of pitch-glide interact. We observe that for vowels of higher spectral centroid, decreasing pitch-glides cause higher timbre change than the timbre change caused by increasing pitch-glides; vice- versa for vowel of lower spectral centroid. Thus, "incoherent" pitch-glides cause higher timbre change than the "coherent" pitch-glides.

1. INTRODUCTION

Conceptually, pitch has been defined as the periodicity in the auditory stimuli (corresponding to harmonicity) and timbre as the spectral envelope of the harmonics. Traditionally, these two dimensions are considered independent. But, a shift in pitch alters the spectral envelope [1], hence affecting the perceived timbre . In fact, perceptual interaction is known to be bi-directional: (i) There is an influence of the multi-component spectra on the perceived pitch. The perceived pitch of a timbre of higher spectral centroid is higher than the perceived pitch of the timbre of a lower spectral centroid, pitch frequency being the same in both cases [2] [3]. (ii) Also, the perceptual brightness of a timbre increases when there is an increase in pitch [4] [5] [6]. This is because an increase in pitch causes a different spectral sampling of stationary timbre, resulting in perceivable changes in the spectral envelope. Fig.1(a) illustrates sampling the spectra at two pitch frequencies 100 Hz and 200 Hz. Fig.1(b) illustrates the estimated spectral envelope (10th order all-pole model estimate) from the discrete spectrum. It can be observed that there is a change in spectral envelope estimate in these two cases because of the pitch.

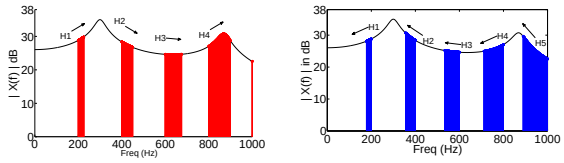
Natural signals like speech, music or bird songs, are often time-varying in nature. They comprise of changing timbre and pitch contours. We do perceive the changing pitch/timbre parameters differently, compared to stationary pitch and timbre [7]. Considering the LTI model of speech production, a vowel with time-varying pitch-glide can be seen as a signal comprising of sum of time-varying harmonics sampling the spectral envelope continuously at different frequencies, (see Fig.2). Fig.2(a) illustrates harmonics of increasing pitch-glide and Fig.2(b) illustrates decreasing pitch-glide. It can be seen that different harmonics sample the spectra differently. In Fig.2(a), while first harmonic (H1) is increasing in strength with increasing pitch-glide, strengths of second and third harmonics (H2 and H3) decrease with decreasing pitch-glide. The effective timbre of the signal is thus dependent, not just on the stationary spectral envelope, but also on the constituent pitch change. Though the spectral envelope is stationary, the resultant timbre of the signal could be non-stationary, caused by the time-varing pitch-glide. Thus, a scalar change in pitch causes a multi- dimensional change in timbre (corresponding to different harmonics components).

In this work, we explore the effect of pitch shift and pitch glides (linearly increasing and decreasing) on stationary timbre using synthetic vowels. Here, pitch-shift refers to a constant shift in pitch, stationary over time; and pitch-glides refer to evolving pitch frequency, non-stationary over the stimulus duration. We also quantify the amount of perceived timbre change introduced by the pitch change using MUSHRA-like listening test.



(a) Original spectral envelope (b) Estimated spectral envelope

Fig. 1. *(a) Synthesized spectral envelope of /u/ with 5 formant all-pole model sampled at pitch f_0 = 100 Hz and 200 Hz (b) Spectral envelope estimated with linear prediction (normalized with the maximam spectral peak) of /u/ sampled at pitch f_0 = 100 Hz (red) and at pitch f_0 = 200 Hz (blue).*



(a) Increasing pitch-glide (b) Decreasing pitch-glide

Fig. 2. *Spectral envelope of /u/ with 5 formant model sampled (a) with increasing pitch-glide from f_0 = 200 Hz to f_0 = 225 Hz.(red) (b) with decreasing pitch-glide from f_0 = 200 Hz to f_0 = 178 Hz.*

2. LTI MODEL BASED SYNTHESIS

We use an LTI model to synthesize the vowels with time-varying pitch-glides. Let $f_0[n]$ represent the required time-varying pitch frequency function, $0 \leq n \leq N-1$, where $T = \frac{N}{f_s}$, f_s = sampling frequency, T = duration of the signal in seconds. We map the instantaneous pitch frequency value $f_0[n]$ to the voiced excitation signal

Table 1. Pitch-glides presented in this experiment $f_0[n]$ (in Hz). P_c=constant pitch glide; P_{in}=increasing pitch-glide of 3 semitones; P_{de}= decreasing pitch glide of 3 semitones. $P1$-$P11$ represents pitch change (Hz) $f_0[n]$ in the test stimuli presented. R_0 and R_100 are at constant pitch 100 Hz and 200 Hz.

	R_0	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	R_100
$P_c \rightarrow$	100	105	112	118	125	133	141	149	158	168	178	188	200
$P_{in} \nearrow$	100	106-126	112-133	119-141	126-150	133-159	141-168	150-178	159-189	168-200	178-212	189-224	200
$P_{de} \searrow$	100	106-89	112-94	119-100	126-106	133-112	141-118	150-126	159-133	168-141	178-150	189-159	200

e[n] as a sequence of impulses with varying separation between adjacent impulses.

$$e[n] = \sum_{k=0}^{K-1} \delta\left[n - m[k]\right] \quad 0 \le n \le N-1 \tag{1}$$

$$\text{where } m[k] = \left\lceil \frac{f_s}{f_0[m[k-1]]} \right\rfloor + m[k-1];\ m[0] = 0 \tag{2}$$

Here $\lceil . \rfloor$ indicates rounding operation. The number of impulses K for a certain duration of e[n] depends on the pitch-glide $f_0[n]$ and the required signal duration T.

We model the system function, characterizing a vowel, as a 5 formant all-pole system as shown in Fig.3. The average formant frequency values for different vowels, estimated from natural speech utterances, are as shown in the Table.2 [8][9]. Similar to the formant frequencies, bandwidths have also been estimated from natural speech utterances. Using averaged measurements, it has been shown that the formant bandwidths for natural vowels correspond to a piece-wise linear function of the respective formant frequencies F_p [10]. Thus, we have mapped the F_p to the bandwidth σ_p by piece-wise linear approximation [10]. The complex pole pair, parametrized using the formant frequency and its bandwidth is given as :

$$z_p = e^{-\sigma_p \cdot \pi / f_s} \cdot e^{j 2\pi \cdot \frac{F_p}{f_s}} \tag{3}$$

And the transfer function is modeled as 5-formant all-pole system, as shown below:

$$H(\omega) = \frac{\alpha}{\prod_{p=1}^{5}(1 - z_p \cdot e^{-j\omega})(1 - z_p^* \cdot e^{-j\omega})} \tag{4}$$

$$H(z) = \frac{\alpha}{1 + \sum_{k=1}^{10} a_k z^{-k}} \tag{5}$$

The vowel $d[n]$ is modeled as the output of a linear system, with H(z) as transfer function and $e[n]$ as input. Thus,

$$D(z) = E(z) \cdot H(z) \tag{6}$$

$$\text{or } d[n] = -\alpha \sum_{k=1}^{10} a_k \cdot d[n-k] + e[n] \tag{7}$$

Here, α is used to normalize to energy of 23.52 dB across all the stimuli. The stimuli is of duration T = 750 ms. As the pitch glides are non-stationary, a high sampling frequency of $f_s = 48$ kHz is used.

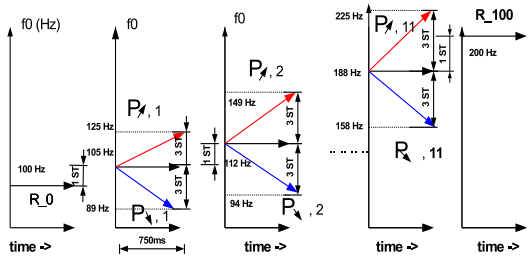
3. PERCEPTION EXPERIMENT

Three male listeners participated in this experiment. They were selected after a set of screaning tests, involving identification of pitch and/or timbre change in a pair of stimuli presented. They were initially exposed to the nature of changes in signal parameters, i.e., pitch and timbre. Then they were presented with randomized test stimuli pairs, with change in pitch or timbre or both. Listeners who could differentiate the pitch change and timbre change were selected. These chosen listeners scored more than 60% in the screening test.

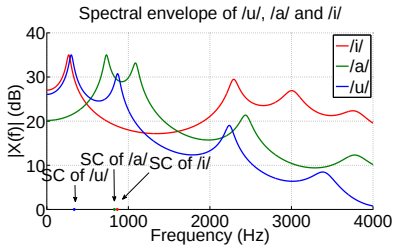
The main experiment comprises of presenting 11 stimuli of one vowel type, each separated by a semi-tone, in a randomized order.

Table 2. Formant frequency details[8][9]

vowel	F1 Hz	F2 Hz	F3 Hz	F4 Hz	F5 Hz
/a/	730	1090	2440	3781	4200
/i/	270	2290	3010	3781	4200
/u/	300	870	2240	3406	4200



(a) Pitch shifts and pitch glides



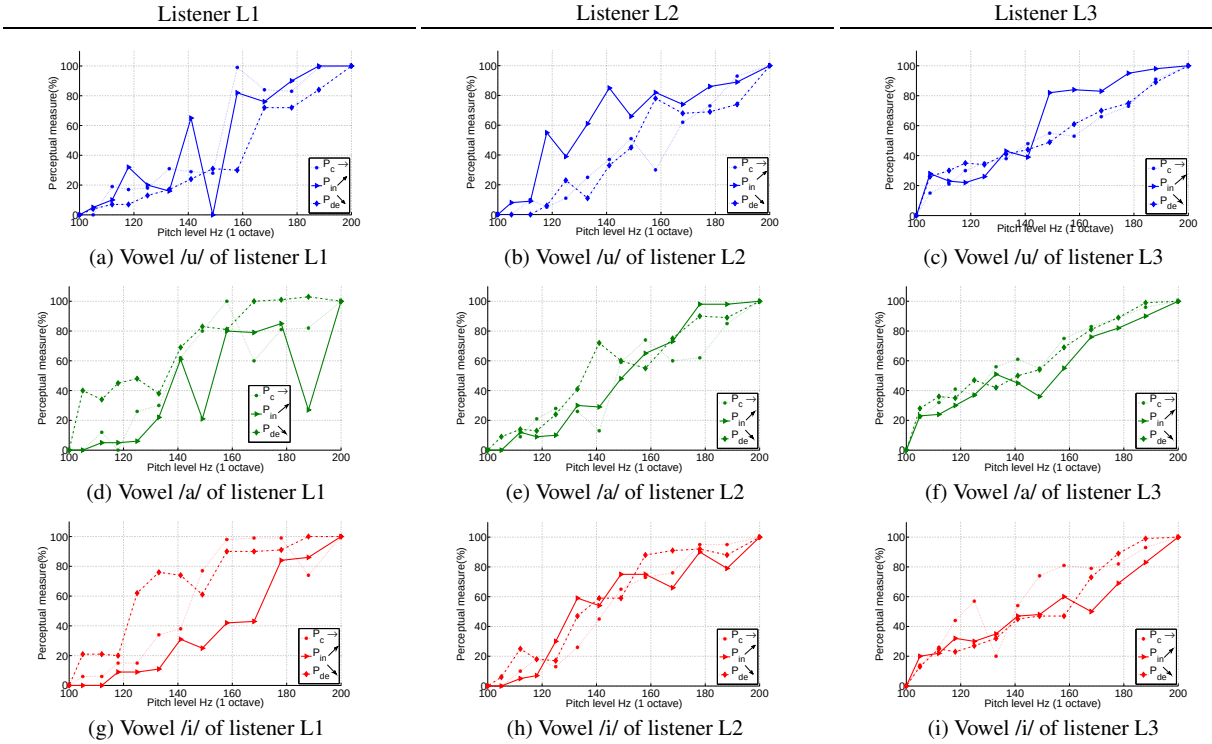
(b) Vowel spectra

Fig. 3. *(a) This figure describes the three types of pitch glides used in this experiment R_0 and R_100 are at constant pitch. The increasing pitch glides are shown in red; decreasing pitch glides are shown in blue; constant pitch are shown in black. Each pitch glide is of duration 750 ms. (b)5-formant spectral envelope of vowels /u/,/a/ and /i/. The spectral centroid (SC) is calculated at pitch 150 Hz. $SC_{/u/}$ = 337 Hz; $SC_{/a/}$ = 829 Hz;$SC_{/i/}$ = 865 Hz. On brightness scale: /u/ << /a/ < /i/*

Each experimental trial comprises of presenting stimuli of one row of Table.3. Totally, there are nine such experimental trials. We choose three cases of vowel timbre, /u/, /a/ and /i/. These vowels are chosen since they form the vertices of the vowel triangle. Among these vowels, /u/ has the least spectral centroid (dull vowel) and /i/ has the highest spectral centroid (bright vowel) and /a/ has a spectral centroid in between. Their spectral envelope is shown in Fig.3(b). Thus, we are studying the timbres over a wide range of spectral centroid. Reference R_0 and R_100 are stimuli with constant pitch at 100 Hz and 200 Hz respectively. These stimuli are made explicit to the listeners. These references are chosen so that all the three types of pitch-glides could have the same references. The listeners were asked to rate the perceptual timbre change with respect to the references on a scale of 0-120. The scale was chosen above 100 because the listener might perceive a timbre change more than R_100. Thus, we can compare the perceptual timbre change across different pitch-glides for the same vowel. We use three types of pitch as show in Table.1. Fig.3(a) represents the pitch-glides in the stimuli. P_c is of constant pitch over the entire 750 ms; P_{in} is a three semi-tone increasing pitch glide; P_{de} is a three semi-tone decreasing pitch glide. We study the nature of timbre change caused by these three types of pitch glides.

3.1. Perceptual scale of timbre change using "MUSHRA"

We have developed a MUSHRA-like (MUltiple Stimuli with Hidden Reference and Anchor) perceptual evaluation tool, in order to



(a) Vowel /u/ of listener L1 (b) Vowel /u/ of listener L2 (c) Vowel /u/ of listener L3

(d) Vowel /a/ of listener L1 (e) Vowel /a/ of listener L2 (f) Vowel /a/ of listener L3

(g) Vowel /i/ of listener L1 (h) Vowel /i/ of listener L2 (i) Vowel /i/ of listener L3

Fig. 4. *Timbre change measure for vowels /u/, /a/ and /i/ for three listeners L1, L2 and L3*

determine the perceptual measure of timbre change. MUSHRA is a well-established experimental procedure used in perceptual rating of audio and speech coders [11]. We take a similar approach in rating the timbre change. We provide fixed anchors as reference stimuli, i.e., 0% (R_0) and 100% (R_100). We ask the listeners to rate the multiple test stimuli presented in the positions in the range of 0-120%. A hidden reference could also be included to validate the listener's base performance. The listeners could play the stimuli as many times as they want to adjust the individual perceptual ratings relative to the other stimuli, until they are fully confident of the rating. This multiple stimuli rating through repeated listening, results in more consistent listener response than the forced binary decisions of the R-AB test (double-blind triple-stimulus with hidden reference test). MUSHRA approach relies less on the user-memory, although comparatively longer listening time is used by the listener. To break the monotony of listening to the timbre samples, the listener can choose to listen to either a noise signal or flute music, in between the tests.

Table 3. *References and 11 test stimuli are parametrically represented in this table. Along each row, adjacent stimuli are separated by one semi-tone. For each vowel, there are 3 types of pitch-glides ($f_0[n]$). $P_{\rightarrow}$=constant pitch; $P_{\nearrow}$=increasing pitch-glide;$P_{\searrow}$= decreasing pitch-glide; R_0 and R_100 represent reference stimuli at pitch 100 Hz and 200 Hz respectively.*

	$f_0[n]$	Pitch-shift level in semitone					
		0(R_0)	1	2	...	11	12(R_100)
/u/	$P_c\rightarrow$	$P_{\rightarrow,0}T_u$	$P_{\rightarrow,1}T_u$	$P_{\rightarrow,2}T_u$	...	$P_{\rightarrow,11}T_u$	$P_{\rightarrow,12}T_u$
/u/	$P_{in}\nearrow$	$P_{\rightarrow,0}T_u$	$P_{\nearrow,1}T_u$	$P_{\nearrow,2}T_u$	...	$P_{\nearrow,11}T_u$	$P_{\rightarrow,12}T_u$
/u/	$P_{de}\searrow$	$P_{\rightarrow,0}T_u$	$P_{\searrow,1}T_u$	$P_{\searrow,2}T_u$	...	$P_{\searrow,11}T_u$	$P_{\rightarrow,12}T_u$
/a/	$P_c\rightarrow$	$P_{\rightarrow,0}T_a$	$P_{\rightarrow,1}T_a$	$P_{\rightarrow,2}T_a$	...	$P_{\rightarrow,11}T_a$	$P_{\rightarrow,12}T_a$
/a/	$P_{in}\nearrow$	$P_{\rightarrow,0}T_a$	$P_{\nearrow,1}T_a$	$P_{\nearrow,2}T_a$	...	$P_{\nearrow,11}T_a$	$P_{\rightarrow,12}T_a$
/a/	$P_{de}\searrow$	$P_{\rightarrow,0}T_a$	$P_{\searrow,1}T_a$	$P_{\searrow,2}T_a$	...	$P_{\searrow,11}T_a$	$P_{\rightarrow,12}T_a$
/i/	$P_c\rightarrow$	$P_{\rightarrow,0}T_i$	$P_{\rightarrow,1}T_i$	$P_{\rightarrow,2}T_i$	...	$P_{\rightarrow,11}T_i$	$P_{\rightarrow,12}T_i$
/i/	$P_{in}\nearrow$	$P_{\rightarrow,0}T_i$	$P_{\nearrow,1}T_i$	$P_{\nearrow,2}T_i$	...	$P_{\nearrow,11}T_i$	$P_{\rightarrow,12}T_i$
/i/	$P_{de}\searrow$	$P_{\rightarrow,0}T_i$	$P_{\searrow,1}T_i$	$P_{\searrow,2}T_i$	...	$P_{\searrow,11}T_i$	$P_{\rightarrow,12}T_i$

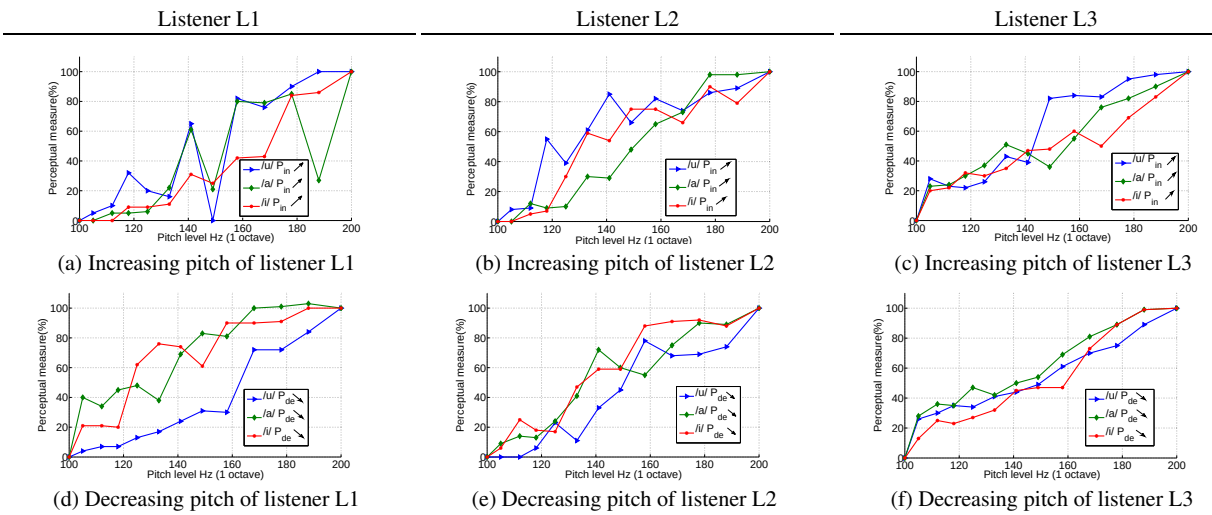
4. RESULTS AND OBSERVATIONS

We present all the listening test results without averaging across listeners or vowels. This will help us examine listener to listener variations across different vowels, since listener percept of each vowel could be different. We refer listeners as L1, L2 and L3 in the following discussion.

4.1. Effect of pitch-shift on timbre

In each experimental trial, all the test samples presented have the same timbre, differing only in the pitch frequency. Each line in the graphs of Fig.4 represents the timbre change rating for the randomized 11 test stimuli of a particular row in the stimuli Table.3. There is a monotonic increase in perceived timbre change with respect to increase in pitch-shift. This is in agreement with the results shown in [5], 16 instruments timbre at different pitch frequencies (stationary) had been used to rate timbre change. In their work, it has been shown that timbre perception is affected by the pitch shift. We extend the work and through our experiments, we have quantified the effect of pitch on timbre through the listening test. From Fig.4, it is clear that there is a non-zero timbre change perceived because of the pitch-shift and it is monotonic in nature, implying a larger pitch change causes a larger timbre change. This monotonic nature is preserved across all three types of pitch-glides.

An interesting observation is that - for some pitch-glides $P_{\nearrow,11}, P_{\nearrow,12}$ (pitch gliding from 178 Hz to 211 Hz or 188 Hz to 225 Hz), even though the component pitch-glide exceeds the reference pitch (R_100) at 200 Hz, the perceived timbre change is less than the reference. In the MUSHRA rating, we have given a provision to rate the timbre change to be more than 100% also. But none of the listeners rated these stimuli above 100%. This indicates that listeners extract an effective or average pitch and the perceived timbre change is influenced by the average or effective pitch.



(a) Increasing pitch of listener L1 (b) Increasing pitch of listener L2 (c) Increasing pitch of listener L3

(d) Decreasing pitch of listener L1 (e) Decreasing pitch of listener L2 (f) Decreasing pitch of listener L3

Fig. 5. *Timbre change measure for vowels /u/, /a/ and /i/ for three listeners L1, L2 and L3 for two types of pitch-glides*

4.2. Effect of pitch-glide on timbre: Coherence in pitch and timbre interaction

Since pitch and timbre are individual percepts and the vowels are ascribed with a qualitative attribute of brightness/dullness, we can examine if the pitch-glide is interactive with the vowel attribute. Since these vowels represent different shades of brightness and the increasing or decreasing pitch glides are inherently "bright" or "dull" respectively, we examine the interaction of these qualitative parameters towards timbre change.

(i) In Fig.4(d-i), considering constant pitch and linear pitch-glides, for /a/ and /i/ (bright vowels), $P_{de}(\searrow)$ causes maximum timbre change for several pitch-shifts. This may be because the dull pitch-glides cause larger timbre change to vowels of high spectral centroid. Similarly, for /a/ and /i/, $P_{in}(\nearrow)$ causes the least timbre change in several cases. The increasing pitch-glides may cause lesser timbre change to the vowels with higher spectral centroid. For /a/, such behavior is seen in L1 clearly in 4(d); L2 showed such behavior with small changes of pitch changes in Fig.4(e); For /i/ such behavior is clearly seen by L1 and L2 in Fig.4(g)(h); L3 shows such behavior for larger pitch differences in Fig.4(i).

(ii) In the case of /u/ (a dull vowel), $P_{in}(\nearrow)$ causes maximum timbre change for several pitch-shifts in Fig.4(a,b,c). On the other hand, for the same /u/, $P_{de}(\searrow)$ causes less timbre change in several cases. This is probably because with respect to the vowel of lower spectral centroid, the increasing pitch-glides contrasts more, causing a more perceived timbre change and decreasing pitch-glides cause less timbre change. For L1, decreasing pitch glide shows such timbre interaction but ratings has large variance, as seen in Fig.4(a). L2 clearly follows this trend in Fig.4(b). L3 shows such a trend only for larger pitch shifts, as seen in Fig.4(c).

Thus we can note that for vowels of high spectral centroid, decreasing pitch-glides cause large timbre change. For vowels of low spectral centroid, increasing pitch-glides would cause significant timbre change. In [6], when the pitch is changed incoherently (increase in pitch-shift when the spectral centroid decreases or vice versa), the percentage of identification of timbre change falls. Whereas, when the pitch is changed coherently (increase in pitch-shift when the spectral centroid increases or vice versa), the percentage identification of timbre change increases. The nature of interaction between pitch and timbre can be interpreted in terms of perceptual "congruence" [6]. These experiments are reported for a single formant stationary signal. We are able to show the effect of coherency in time-varying pitch-glides in presence of vowels, giving phonetic context. An increasing pitch-glide is coherent with an increase in spectral centroid and a decreasing pitch-glide is incoherent with increase in spectral centroid.

4.3. Comparison of timbre change across vowels

In order to understand the nature of timbre change across vowels, we re-arrange the observations, grouping across vowels in Fig.5. Though the references R_0 and R_100 are different across the experimental trials, the perceptual ratings are superposed to better understand the perceptual timbre change. Fig.5(a,b,c) represent timbre change ratings for increasing pitch glide and Fig.5(d,e,f) represent the timbre change ratings for decreasing pitch glide.

(i) In Fig.5(a)(b)(c), for increasing pitch-glide, /u/ causes largest timbre change at several pitch levels. L1 shows such tendencies for small and very large pitch-shifts in Fig.5(a). L2 shows for medium pitch-shifts in Fig.5(b) and L3 shows for large pitch-shifts in Fig.5(c). On the other hand, /i/ shows the minimal timbre change for L1 and L3 in Fig.5(a)(c).

(ii) For decreasing pitch-glides, clearly /u/ shows least timbre change for L1 and L2 at several pitch levels in Fig.5(d)(e). L3 shows such behavior for larger pitch-shifts in Fig.5(f). /a/ and /i/ cause more timbre change than /u/ as perceived by L1 and L2 as seen in Fig.5(d)(e).

From the observations, it is seen that for the dull vowel /u/, increasing pitch glide causes more timbre change than the decreasing pitch glide; For bright vowels /a/ and /i/, decreasing pitch glides cause more timbre change than the increasing pitch glide. Thus, coherent nature of interaction is observed, when the ratings are compared not just across pitch glides (Sec. 4.2) but across vowels also.

5. CONCLUSION

The main observation of this work is time-varying pitch glides interact with vowel timbre. An increasing pitch glide causes more timbre change in a vowel with low spectral centroid than in an vowel with high spectral centroid. Similarly, a decreasing pitch glide causes more timbre change in a vowel of high spectral centroid than in a vowel of low spectral centroid. Thus an increasing pitch glide is coherent with high spectral centroid and incoherent with low spectral centroid; and vice-versa for decreasing pitch glide; This perceptual property of coherency can be used in instrument synthesis, music analysis and data sonification applications.

6. REFERENCES

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[2] C. M. Warrier and R. J. Zatorre, "Influence of tonal context and timbral variation on perception of pitch," *Perception & psychophysics*, vol. 64, no. 2, pp. 198–207, 2002.

[3] F. A. Russo and W. F. Thompson, "An interval size illusion: The influence of timbre on the perceived size of melodic intervals," *Perception & psychophysics*, vol. 67, no. 4, pp. 559–568, 2005.

[4] R. D. Melara and L. E. Marks, "Interaction among auditory dimensions: Timbre, pitch, and loudness," *Perception and Psychophysics*, vol. 48, no. 2, pp. 169–178, '90b.

[5] J. Marozeau and A. D. Cheveigné, "The effect of fundamental frequency on the brightness dimension of timbre," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 383–387, 2007.

[6] E. J. Allen and A. J. Oxenham, "Symmetric interactions and interference between pitch and timbre," *J. Acoust. Soc. Am.*, vol. 135, no. 3, pp. 1371–1379, 2014.

[7] C. L. Krumhansl and P. Iverson, "Perceptual interactions between musical pitch and timbre," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 18, no. 3, pp. 739–751, 1992.

[8] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Pearson Education, 1993.

[9] A. De Cheveign, "Formant bandwidth affects the identification of competing vowels," *Proceedings of the 14th International Congress of Phonetic Sciences San Francisco*, vol. 4, pp. 2093–2096, 1999.

[10] H. K. Dunn, "Methods of measuring vowel formant bandwidths," *J. Acoust. Soc. Am.*, vol. 33, no. 12, pp. 1737–1746, 1961.

[11] ITU-R, "Recommendation b.s., 1534-1: Method for the subjective assessment of intermediate quality levels of coding systems," January 2003.