Robust Query-by-Singing/Humming System against Background Noise Environments

Kichul Kim, Kang Ryoung Park, Sung-Joo Park, Soek-Pil Lee, and Moo Young Kim, *Senior Member*, IEEE

**Abstract — *Under background noise environments, the performance of the Query-by-Singing/Humming (QbSH) system is considerably degraded. Since human pitch information is used as a feature vector for the QbSH system, a noise robust pitch-estimation algorithm is inevitable. Thus, a novel pitch-estimation method is proposed by integrating temporal-autocorrelation and spectral-salience methods. As a pre-processing block, spectral smoothing is applied to enhance the stationarity of the noisy input signal. To calculate the similarity between the MIDI database and input humming signal, the dynamic time warping (DTW) algorithm is used. Jang's corpus and AURORA2 database are selected as humming and background noise signals, respectively. Compared with the standard pitch estimation algorithm in the ITU-T G.729 speech codec, the proposed pitch estimation method improves the average accuracy by 11.7% for the 0 dB signal-to-noise ratio (SNR) noise case. It also improves top-20 ratio and mean reciprocal rank (MRR) of the proposed QbSH system, on average, by 7.4% and 0.13, respectively*[1].

Index Terms — Query-by-Singing/Humming, pitch estimation, background noise, dynamic time warping.

I. INTRODUCTION

In accordance to the development of portable consumer devices and mobile communications systems, an enormous amount of digital music contents is easily accessible. Thus, automatic music-information retrieval (MIR) systems, such as song identification, music genre classification, and music recommendations, have obtained increased attention from both service providers and application developers [1]-[6].

Of the various MIR topics, we propose a novel Query-by-Singing/Humming (QbSH) system to be robust against background noise environments. In a QbSH system, the input humming or signing signal is represented by a melody vector using pitch estimation and voice activity detection (VAD). It is compared with a set of melody sequences extracted from the large size of the music database to identify which song is hummed.

The QbSH system can be classified into note-based and frame-based methods based on the feature vector [7]-[9]. The former method extracts a set of features such as musical interval, duration, and tempo from each note [10]-[15]. To improve performance, the frame-based pitch estimation method was proposed in [16]-[19]. The QbSH system can be also classified into bottom-up and top-down methods based on the pattern matching algorithm [7]-[9]. In the bottom-up method [10]-[17], the local matching between query and target is performed, which can produce the optimal matching result. On the contrary, the top-down method calculates the similarity between the query and target based on the global shape information [7], [18], [19].

The QbSH system based on dynamic time warping (DTW) was proposed by Jang et al., which is a popular frame-based method [17]. The singing data of use are converted into a pitch vector as an input feature. The DTW algorithm is used to calculate the distance between the input pitch vector and a set of pitch vectors extracted from the music database. Another frame-based matching method was proposed by Wu et al., which is called recursive alignment (RA). It can solve the problem of melody alignment in the top-down method [7]. The RA algorithm is similar to the linear scaling (LS) method, but it performs the local matching recursively to solve the alignment problem. Ryynanen et al. proposed a frame-based method that extracts pitch vectors in a time window of fixed length. With these features, a locality sensitive hashing (LSH) method is used as a top-down matcher [19]. Typke et al. proposed an earth mover's distance (EMD) method that calculates the minimum cost between two features of humming and music data with the varying weights [15].

Conventional pitch estimation methods are based on the autocorrelation function (ACF) [4], [20], the average magnitude differential function (AMDF) [21], and the cepstrum-based approach [22]. In [23] and [24], multiple pitch is estimated by using salience, the correlogram model, the auditory model, and mel-scale band analysis. However, these methods yield degraded performance under background noise environments. Since the QbSH system is used in real-life noise environments, noise robustness in feature extraction is a crucial step.

Since only the pitch information is used as an input feature-vector of the proposed QbSH system, we propose a robust pitch-estimation method by integrating time and frequency-

[1] This work was supported by the Ministry of Knowledge Economy grant funded by the Korea government (No. 2010-S-001-01).

K. Kim and M. Y. Kim are with the Human Computer Interaction Laboratory, Department of Information and Communication Engineering, Sejong University, Seoul, Korea (e-mail: neraijer@naver.com and mooyoung@sejong.ac.kr).

K. R. Park is with the Division of Electronics and Electrical Engineering, Dongguk University, Seoul, Korea (e-mail: parkgr@dongguk.edu).

S.-J. Park and S.-P. Lee are with the Digital Media Research Center, Korea Electronics Technology Institute, Seoul, Korea (e-mail: bpark@keti.re.kr and lspbio@keti.re.kr).

Contributed Paper
Original manuscript received 02/25/11
Revised manuscript received 04/12/11
Current version published 06/27/11
Electronic version published 06/27/11.

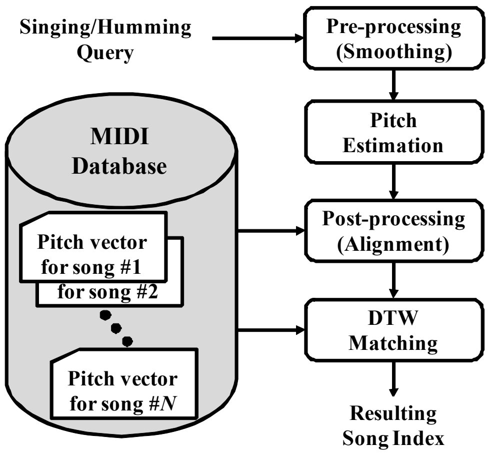

Fig. 1. Block diagram of the proposed QbSH system.

domain methods. Additional pre and post-processing steps improve the stationarity of the corrupted input signal.

The proposed QbSH system, which includes robust pitch estimation and DTW matching, is explained in section II. The experimental results and conclusion are presented in sections III and IV, respectively.

II. PROPOSED QBSH SYSTEM

In a QbSH system, it is important to extract melody from user's signing and humming. The melody includes a note and the duration of the note. We extract the pitch information at a regular interval to represent a note and its duration.

Fig. 1 indicates the overall block diagram of the proposed QbSH system. At the training stage, pitch-vector candidates are extracted from the MIDI database for each song. At a test stage, a pitch vector is estimated from the input singing/humming query by pre-processing, pitch estimation, and post-processing. Since the input query is corrupted by the background noise, a robust pitch-estimation module should be designed. Then, DTW matching is performed by measuring similarities between an input pitch-vector and pitch-vectors in the database.

A. Robust Pitch Estimation

In the time domain, the autocorrelation function (TACF) is calculated by multiplying a signal with a delayed version of itself. TACF is defined by

$$R(\tau)=\frac{\sum_{n=0}^{N-\tau-1} x(n)x(n+\tau)}{\sqrt{\sum_{n=0}^{N-\tau-1} x^2(n) \sum_{n=0}^{N-\tau-1} x^2(n+\tau)}} \tag{1}$$

where $x(n)$, τ , and N are an n-th sample of the input singing/humming signal, the delay of the signal, which corresponds to a pitch candidate, and the frame size, respectively. The pitch value is estimated by τ , which maximizes (1). The TACF has used in various systems such as in the ITU-T G.729 speech codec [4]. Its performance is reported to be reasonably good, but it produces pitch doubling errors.

In the frequency domain, the salience-based method (SAL) is used for extracting multiple F0 [23], [24]. It extracts predominant F0 under musical instruments. First, the spectral energy distribution for each critical band is calculated, and then the whiting process partially flattens the formant structure of the spectrum. Using spectral energy, the standard deviation is calculated by

$$\sigma_b=\left(\frac{1}{K}\sum_k H_b(k)\,|\,X(k)\,|^2\right)^{1/2} \tag{2}$$

where $|\,X(k)\,|^2$, $H_b(k)$, and K are input power spectrum for each frequency bin k, the mel-scale filter in each subband b, and the length of the Fourier transform, respectively. Next, bandwise compression coefficients are calculated by $\gamma_b=\sigma_b^{\nu}$ where $\nu=-0.67$ is a parameter to determine the amount of spectral whitening. For each frequency bin k, the compression coefficient $\gamma(k)$ is obtained by applying linearly interpolation to γ_b . Then, a whitened spectrum $Y(k)$ is calculated by

$$Y(k)=\gamma(k)X(k)\,. \tag{3}$$

The salience is calculated by

$$s(\tau)=\frac{1}{M_\tau}\sum_{m=1}^{M_\tau}\max_{k'\in k_{\tau,m}\pm\vartriangle k}\left|\frac{Y(k')}{f(k')+\alpha}\right| \tag{4}$$

where $k_{\tau,m}$, $\vartriangle k$, $f(k')$, M_τ , and α are the frequency bin of an m-th harmonic partial, allowable pitch error, frequency value of index k' , the number of an harmonic partial for an time lag τ , and an adjusting parameter, respectively. The first harmonic partial corresponds to F0 and is emphasized in (4).

In [20], we proposed IACF where TACF and the frequency-domain autocorrelation function (FACF) are combined. SAL gives superior performance to FACT. However, it still produces pitch halving errors. A time-domain method such as TACF generates a pitch doubling error. On the other hand, a frequency-domain pitch estimation method such as SAL has the disadvantage of generating a pitch halving error. To reduce these errors, the time-domain method, TACF, is integrated with the frequency-domain method, SAL. It can reduce both pitch doubling and having errors and be robust against background noise environments. Our proposed method is called an integrated SAL (ISAL) that is calculated as a weighted summation between TACF and SAL by

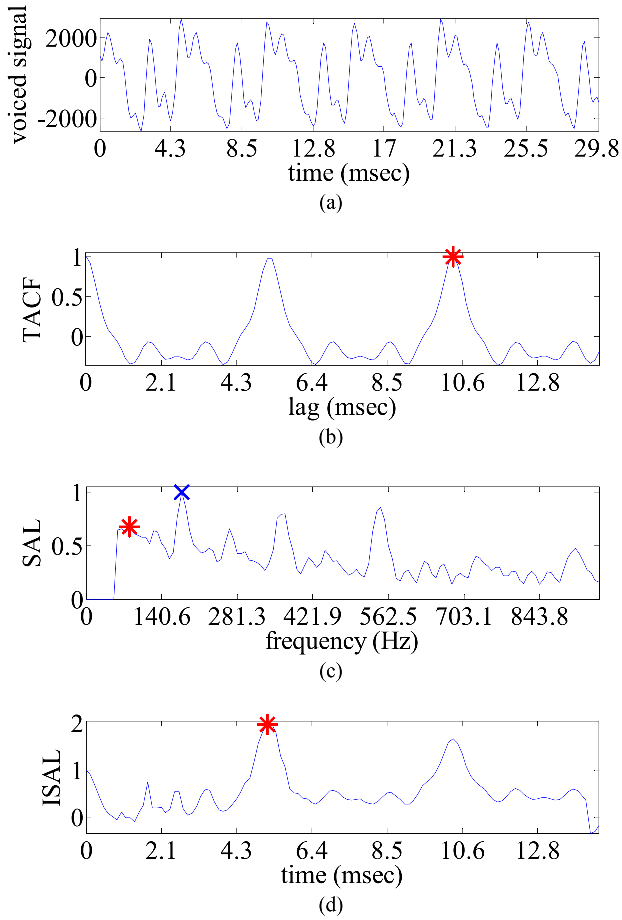

Fig. 2. Performance of ISAL: (a) a voiced input signal, (b) its time-domain autocorrelation function (TACF), (c) its salience in a frequency-domain (SAL), (d) integrated version between (b) and (c) (ISAL).

$$\tilde{R}(\tau) = \beta R(\tau) + (1-\beta)s(\tau) \quad (5)$$

where β is a weighting factor. The salience $s(\tau)$ is normalized with its maximum value to have the same dynamic range as (1). The final pitch value is estimated by finding a maximum ISAL value as given by

$$\tau^* = \arg \max_{\tau \in [a,b]} \{\tilde{R}(\tau)\} \quad (6)$$

where a, b, and τ^* are the minimum and maximum values of the pitch periods, and the estimated pitch value, respectively.

TACF has the drawback of generating a pitch doubling error. Conversely, SAL has the drawback of generating a pitch halving error. ISAL can reduce both types of errors by integrating two different methods as shown in Fig. 2. Fig. 2 (a) depicts the input speech signal. In Fig.2 (b), the pitch doubling error corresponding to the maximum peak location denoted by a symbol '*' is produced by TACF. However, the pitch doubling error does not occur in SAL. In Fig. 2 (c), the denoted symbols ‘x’ and '*' indicate the local maximum peaks in SAL and TACF, respectively. In ISAL, by combining TACF and SAL, we can find the correct pitch period as shown in Fig. 1 (d).

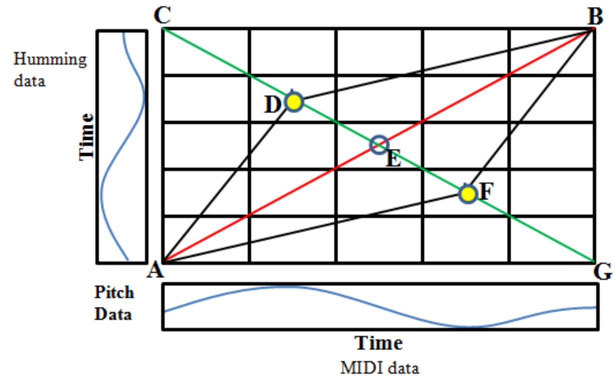

Fig. 3. Searching area of the DTW matching [8], [9].

Pitch estimation methods, such as TACF and FACF, are performed under the assumption that the signal itself is quasi-stationary. However, the singing/humming signal is not stationary when it is corrupted by background noise. Especially, time-varying noise with a low signal-to-noise ratio (SNR) causes significant spectral distortion. In order to reduce the non-stationarity, we apply the first-order recursive averaging method as a pre-processing module as given by

$$\tilde{Y}(l,k) = (1-\alpha)Y(l,k) + \alpha Y(l-1,k) \quad (7)$$

where α , l, and k are the smoothing factor, frame index, and frequency bin, respectively. Because the input speech spectrum fluctuates over time especially under the non-stationary noise condition, this module is mainly used for speech enhancement to improve the stationarity of the corrupted signal.

The post-processing is performed to the output of the pitch estimation module to minimize the inconsistency between input singing/humming and MIDI database. Because the note values in MIDI files are represented by a semitone format, the estimated F0 values from the input query are transformed into a semitone format as given by

$$semitone = 12 * \log_2\left(\frac{F0}{440}\right) + 69\,. \quad (8)$$

Although the corresponding song is sung or hummed, the pitch vector of the singing/humming query can differ from that of a song in MIDI database. To reduce this mismatch, both mean values of the pitch vectors of input query and the MIDI data are set to be zero. In addition, we apply median filter and moving averaging filter in order to further reduce the errors caused by the user vibration and inaccurate pitch estimation.

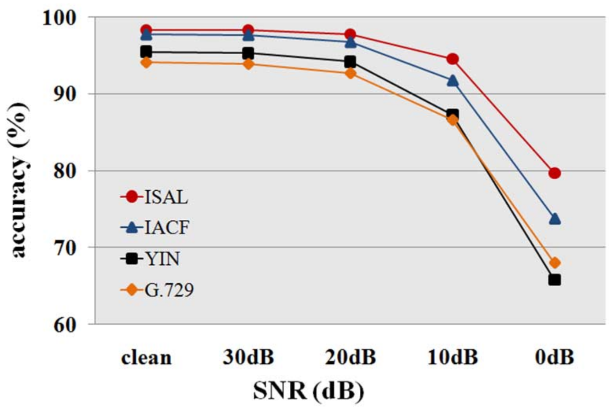

Fig. 4. Performance (GER-10%) of ISAL, IACF, YIN, and G.729 for clean, 30dB, 20dB, 10dB, and 0dB SNR cases.

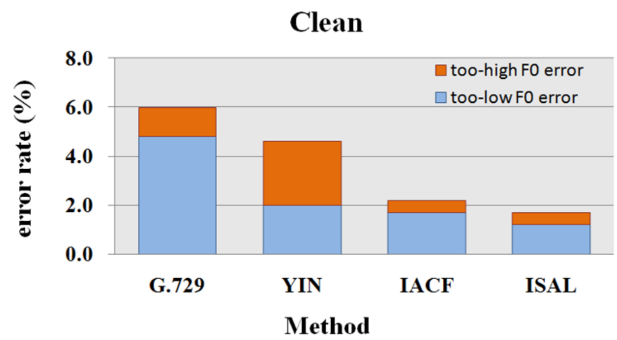

Fig. 5. Evaluation of too-high and too-low errors of G.729, YIN, IACF, and ISAL for clean case.

B. DTW Matching Algorithm

In general, there exists a difference in the lengths of the hummed phrase and the original music when a user hums a song. To reconcile this issue, we use a dynamic time warping (DTW) matching algorithm [8], [9]. DTW matching is a method that can calculate the similarity between the two patterns that have different lengths using insertion and deletion. Fig. 3 shows the searching area of the DTW matching. The horizontal and vertical axes correspond to the original MIDI data and the input humming data, respectively. The DTW matching is not limited to searching the entire searching area based on the assumption that a person hums a song similarly to the original MIDI data, neither too slower nor too faster. Consequently, the input humming data can be successfully matched with the genuine MIDI file just by searching the part of the entire searching area [8], [9]. If the size of the searching area gets smaller, the processing time is reduced, but matching errors increase. By changing the positions of D and F in Fig. 3, we can determine the size of the searching area (the rhombus shape defined by ADBF) of the DTW method [8], [9]. The optimal size can be empirically determined in terms of matching accuracy. In [8], [9], the experimental results show that the accuracy is the best when the length of line (DE) to that of line (CE) is 1/4.

A user can hum or sing an arbitrary part of the original music. So, the proposed DTW method calculates the distance between the humming and MIDI features by sliding the window of the humming feature [8], [9], [25]. The dissimilarity is calculated between the humming and MIDI features based on the following equation:

$$d_{ps}(r_i, q_j) = \frac{\sqrt{\sum_{m=0}^{M-1}[r_i(m) - q_j(m-ps)]^2}}{\sqrt{\sum_{m=0}^{M-1} r_i^2(m)}\sqrt{\sum_{m=0}^{M-1} q_j^2(m-ps)}}, \tag{9}$$

$$0 \le m, m-ps \le M-1$$

where M, $r_i(m)$ and $q_j(m-ps)$ are the feature dimensionality, the i-th MIDI file feature and the j-th humming feature shifted by ps, respectively.

III. Experimental Results

Experiments on the proposed QbSH system were performed against various background noise conditions. For the experiments, Jang's corpus and AURORA2 database were selected. Jang's corpus consists of 48 MIDI files and 4431 singing/humming queries in various channel environments via telephone, microphone, and so on [26]. The query files were recorded by 195 persons at the sampling rate of 8 kHz and the resolution of 16 bits. The ground-truth pitch and voice activity detection (VAD) information were manually labeled every 32 msec. The AURORA2 database was used as real-life background noise. From several noise types, airport, babble, car, exhibition, restaurant, street, subway, and train noises were selected with SNRs of 0, 10, 20, and 30 dB.

We compared the proposed ISAL with the conventional ITU-T G.729 TACF [4], YIN [21], and IACF [20]. The performance of these pitch estimation methods was measured based on the gross error rate (GER). In this paper, we measured the GER-10% where the correct F0 interval is defined between ±10% of the ground-truth F0. If the estimated F0 is above or below 10% of the ground-truth F0, it is defined as too-high and too-low errors, respectively. Too-high and too-low F0 errors correspond to the pitch halving and doubling errors, respectively. The reference VAD was applied to measure the performance of the matching algorithm by using the mean reciprocal rank (MRR) and top-20 recognition rate.

Fig. 4 shows the GER-10% accuracy of the conventional pitch estimation schemes, G.729, YIN, and IACF, and the proposed ISAL. The conventional G.729, YIN, and IACF produce significantly degraded performance in noise environments. For all SNR cases, ISAL gives the best performance in pitch estimation accuracy by combining TACF and SAL. The performance improvement of ISAL is more distinct in low SNR cases such as 10 and 0 dB. In

TABLE I
TOP-20 RECOGNITION RATE (%) AND MRR FOR THE OPTIMAL SYSTEM AND THE PROPOSED ISAL.

performance	top-20	MRR
Ground-truth pitch value	94.4	0.71
Pitch value estimated by ISAL	93.1	0.69

TABLE II
TOP-20 RECOGNITION RATE (%) FOR THE CONVENTIONAL METHODS (G.729, YIN, AND IACF) AND THE PROPOSED ISAL.

Environmental Conditions	G.729	YIN	IACF	ISAL
clean	89.5	87.9	92.4	93.1
30 dB	89.5	87.6	92.3	93.0
20 dB	88.7	86.2	91.3	92.5
10 dB	83.2	78.3	85.6	89.4
0 dB	69.8	58.8	70.4	77.2
average	84.1	79.8	86.4	89.0

TABLE III
MRR FOR THE CONVENTIONAL METHODS (G.729, YIN, AND IACF) AND THE PROPOSED ISAL.

Environmental Conditions	G.729	YIN	IACF	ISAL
clean	0.63	0.60	0.68	0.69
30 dB	0.63	0.60	0.68	0.69
20 dB	0.61	0.58	0.66	0.69
10 dB	0.50	0.46	0.57	0.63
0 dB	0.30	0.23	0.39	0.43
average	0.53	0.50	0.59	0.63

average, G.729, YIN, IACF, and ISAL have the accuracy of 87.1%, 87.6%, 91.6%, and 93.7%, respectively. In 0 dB SNR, ISAL yields better accuracy of 11.7% than G.729.

Fig. 5 depicts the too-low and too-high F0 errors of all the methods in the clean environment. Both G.729 and YIN are time-domain methods that include a weighting function to reduce the pitch doubling error that corresponds to the too-low F0 error. However, G.729 produces too-low F0 errors more frequently than too-high F0 errors. YIN gives more balanced F0 errors and better performance than G.729. By combining time and frequency methods, IACF and ISAL produce less too-low and too-high F0 errors than the time domain methods.

Table I shows the performance of the QbSH systems by using the ground-truth pitch value and the proposed ISAL. In terms of both top-20 ratio and MRR, ISAL gives the almost similar performance to the reference pitch.

Table II and III present the top-20 and MRR scores of the QbSH systems using G.729, YIN, IACF, and the proposed ISAL for every SNR case. Compared with G.729, YIN has better GER-10% in pitch estimation, but worse top-20 and MRR scores in QbSH recognition rate since it cannot estimate the exact pitch contour. For clean, 30dB, 20dB, 10dB, and 0dB SNR cases, the proposed method, ISAL, gives the best performance in pitch estimation and QbSH recognition rate.

IV. CONCLUSION

To reduce performance degradation under background noise environments, we propose the robust pitch estimation method, ISAL, which provides the noise robust feature into the QbSH system. The proposed method integrates time-domain and frequency-domain methods to reduce pitch doubling and halving errors. The spectral smoothing enhances the stationarity of the input signal especially for the non-stationary noise at lower SNR. To design a QbSH system, DTW is used to calculate the similarity between various MIDI file candidates and the input humming signal in a F0 domain.

In 0 dB SNR, ISAL produces a 11.7% performance improvement in pitch estimation compared with the G.729 open-loop pitch search. The performance improvements in terms of Top-20 and MRR of the proposed QbSH system using ISAL are 7.4% and 0.13 compared with that using G.729, respectively. Although the proposed ISAL gives the enhanced GER, top-20 rate, and MRR, we will further improve the performance at a lower SNR than 0 dB.

ACKNOWLEDGMENT

We would like to thank Myeongseok Lee for his help in performing the experiments.

REFERENCES

[1] W. Son, H.-T. Cho, K. Yoon, and S.-P. Lee, "Sub-fingerprint masking for a robust audio fingerprinting system in a real-noise environment for portable consumer devices," *IEEE Trans. Consum. Electron.,* vol. 56, no. 1, pp. 156-160, 2010.

[2] J.-F. Wang, J.-C. Wang, M.-H. Mo, C.-I Tu, and S.-C. Lin, "The design of a speech interactivity embedded module and its applications for mobile consumer devices," *IEEE Trans. Consum. Electron.,* vol. 54, no. 2, pp. 870-876, 2008.

[3] X. Zhu, Y.-Y. Shi, H.-G. Kim, and K.-W. Eom, "An Integrated Music Recommendation System," *IEEE Trans. Consum. Electron.,* vol. 52, no. 3, pp. 917-925, 2006.

[4] ITU-T Recommendation G.729, Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP).

[5] S.-P. Heo, M. Suzuki, A. Ito, and S. Makino, "An effective music information retrieval method using three-dimensional continuous DP," *IEEE Trans. MULTIMEDIA,* vol. 8, no. 3, pp. 633-639, 2006.

[6] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.,* vol. 10, no. 5, pp. 293–302, 2002.

[7] X. Wu and M. Li, "A top-down approach to melody match in pitch contour for query by humming," in *Proc. International Symposium of Chinese Spoken Language Processing*, 2006.

[8] G. P. Nam, K. R. Park, S.-J. Park, S.-P. Lee, and M. Y. Kim, "A new query by humming system based on the score level fusion of two classifiers," *International Journal of Communication Systems*, accepted for publication, 2011.

[9] G. P. Nam, T. T. T. Luong, H. H. Nam, K. R. Park, and S.-J. Park, "Intelligent query by humming system based on score level fusion of multiple classifiers," *EURASIP J. Advances in Signal Process.*, in submission.

[10] R. J. McNab, L. A. Smith, and I. H. Witten, C. L. Henderson, and S. J. Sunningham, "Toward the digital music library : tune retrieval from acoustic input," in *Proc. the first ACM Conf. Digital Libraries*, pp. 11-18, 1996.

[11] R. J. McNab, L. A. Smith, D. Bainbridge, and I. H. Witten, "The new zealand digital library melody index," *Digital Libraries Mag.*, vol. 3, no. 5, pp. 4-15, 1997.

[12] A. Kornstadt, "Themefinder: a web-based melodic search tool," *Computing in Musicology,* vol. 11, pp. 231-236, 1998.

[13] R. Typke, F. Wiering, and R. C. Veltkamp, "A survey of music information retrieval systems," in *Proc. Int. Conf. Music Information Retrieval,* pp. 153–160, 2005.

[14] S. Blackburn and D. DeRoure, "Tool for content based navigation of music," in *Proc. ACM Multimedia*, pp. 361-368, 1998.

[15] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. V. Oostrum, "Using transportation distances for measuring melodic similarity," in *Proc. Int. Conf. Music Information Retrieval,* pp. 107-114, 2003.

[16] A. Ghias, J. Logan, D. Chamberlain, and B. Smith, "Query by humming – musical information retrieval in an audio database," in *Proc. ACM Multimedia*, pp. 231-236, 1995.

[17] J. S. R. Jang and M. Y. Gao, "A query-by-singing system based on dynamic programming," in *Proc. International Workshop on Intelligent Systems Resolution*, pp. 85-89, 2000.

[18] R. Typke and L. Prechelt, "An interface for melody input," *ACM Trans. Computer-Human Interaction*, vol. 8, no. 2, pp. 133-149, 2001.

[19] M. Ryynanen and A. Klapuri, "Query by humming of MIDI and audio using locality sensitive hashing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2249-2252, 2008.

[20] Y. D. Cho, M. Y. Kim, and S. R. Kim, "A spectrally mixed excitation (SMX) vocoder with robust parameter determination," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* vol. 2, pp. 601-604, 1998.

[21] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimation for speech and music," *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1917-1930, 2002.

[22] S. Ahmadi and A. S. Spanias, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm," *IEEE Trans. Speech Audio Process.,* vol. 7, no. 3, pp. 333–338, 1999.

[23] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Process.,* vol. 11, no. 6, pp. 804-816, 2003.

[24] A. Klapuri, "Multiple fundamental frequency estimation by summing harmonic amplitudes," in *Proc. Int. Conf. Music Information Retrieval,* pp. 216-221, 2006.

[25] J. Song, S. Y. Bae, and K. Yoon, "Mid-level music melody representation of polyphonic audio for query-by-humming system," in *Proc. Int. Conf. Music Information Retrieval,* 2002.

[26] C.-C. Wang, J.-S. R. Jang, and W. Wang, "An improved query by singing/humming system using melody and lyrics information," in *Proc. Int. Society for Music Information Retrieval Conf.,* pp. 45–50, 2010.

BIOGRAPHIES

Kichul Kim is a Master student at the Dept. of Information and Communication Engineering, Sejong University, Seoul, Korea. He received a B.Sc. degree in Electrical Engineering from Sejong University, Seoul, Korea, in 2009. His research interests include speaker recognition, speech enhancement, and music information retrieval.

Kang Ryoung Park received BS and MS degrees in Electronic Engineering from Yonsei University, Seoul, South Korea, in 1994 and 1996, respectively. In 2000, he earned a PhD degree in Computer Vision from the Department of Electrical and Computer Engineering also at Yonsei University. He is currently an Associate Professor at the Division of Electronics and Electrical Engineering at Dongguk University. His research interests include computer vision, image processing and biometrics.

Sung-Joo Park received MS degrees in Electronics Engineering from Kyungpook National University, Daegu, South Korea, in 1997. He is currently a PhD candidate at Kwangwoon University and a Senior Research Engineer at Digital Media Research Center of Korea Electronics Technology Institute. His research interests include signal processing and the convergence of broadcast and telecommunication.

Seok-Pil Lee received BS and MS degrees in Electrical Engineering from Yonsei University, Seoul, South Korea, in 1990 and 1992, respectively. In 1997, he earned a PhD degree in Electrical and Electronics Engineering also at Yonsei University. He is currently a head of Digital Media Research Center of Korea Electronics Technology Institute. His research interests include A/V signal processing and the convergence of digital broadcast and telecommunication.

Moo Young Kim (M'96, SM'10) is an Associate Professor at the Dept. of Information and Communication Engineering, Sejong University, Seoul, Korea. He received an M.Sc. degree in electrical engineering from Yonsei University, Seoul, Korea, in 1995, and a Ph.D. degree in electrical engineering from KTH (the Royal Institute of Technology), Stockholm, Sweden, in 2004, respectively. From 1995 to 2000, he worked as a Member of Research Staff of the Human Computer Interaction Laboratory at Samsung Advanced Institute of Technology, Kiheung, Korea. From 2005 to 2006, he was a Senior Research Engineer of the Dept. Multimedia Technologies at Ericsson Research, Stockholm, Sweden. His research interests include speech and audio coding based on information theory, speech and speaker recognition, speech enhancement, joint source and channel coding, and music information retrieval.