

LSP를 이용한 성문 스펙트럼 기울기 추정에 관한 연구

A Study on the Estimation of Glottal Spectrum Slope Using the LSP (Line Spectrum Pairs)

민 소 연* · 장 경 아**
So-Yeon Min · Kyung-A Jang

ABSTRACT

The common form of pre-emphasis filter is $H(z) = 1 - az^{-1}$, where a typically lies between 0.9 and 1.0 in voiced signal. Also, this value reflects the degree of filter and equals $R(1)/R(0)$ in Auto-correlation method. This paper proposes a new flattening algorithm to compensate the weakened high frequency components that occur by vocal cord characteristic. We used interval information of LSP to estimate formant frequency. After obtaining the value of slope and inverse slope using linear interpolation among formant frequency, flattening process is followed. Experimental results show that the proposed algorithm flattened the weakened high frequency components effectively. That is, we could improve the flattened characteristics by using interval information of LSP as flattening factor at the process that compensates weakened high frequency components.

Keywords: Glottal spectrum slope, Line spectrum pairs, Vocal cord characteristics, Vocal tract characteristics, Flattening

1. 서 론

음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 피치검출이나 포만트 검출은 매우 중요하다. 하지만 음성신호에서는 여파기성분과 여기 성분이 상호 작용하기 때문에 피치검출이나 포만트 검출이 매우 어렵다. 특히 음성신호에 잡음이 부가될 경우에는 더욱 어려워진다. 따라서 낮은 SNR 조건에서도 피치정보나 포만트 정보를 유지하는 스펙트럼 신호는 음성처리 분야에서 매우 중요하다고 할 수 있다. 그런데 스펙트럼 신호에는 고조파 성분과 포만트 성분이 함께 나타난다. 따라서 이를 잘 분리하는 것이 피치검출이나 포만트 검출의 관건이라 할 수 있다. 음성신호의 구조에 대한 연구는 음성정보를 추출하거나 강조할 수가 있다. 따라서 음성신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요하다. 본 논문에서는 주파수 영역에서 LSP를 이용하여

* 서일대학 정보기술계열 정보통신전공

** (주)코아로직 소프트웨어실 AT팀

포먼트 주파수를 추출하고 이 과정을 통해 얻은 포락선 기울기를 사용하여 이에 따라 a값을 각 프레임별로 특성에 맞게 보상해줄 것을 제안한다.

2. 성문 특성을 고려한 음성 생성 모델

음성생성에 대한 선형모델은 50년대 후반 Fant에 의해 개발되었는데 그는 음성출력을 음원이 여파기를 통과하여 나오는 신호로 가정하고, 음원과 성도의 각 부분을 독립적인 것으로 간주하는 선형예측모델을 제시하였다. 음원에 대한 모델로 유성음의 음원은 준주기적인 펄스, 무성음의 음원은 백색잡음을 사용하였고, 성대에서 성문이 음원에 미치는 영향은 다음과 같은 성문모델(Glottal Shaping model)로 모델링하였다[1][2][3].

$$G(z) = \frac{1}{(1 - e^{-cT}z^{-1})^2} \quad (1)$$

식(1)에서 T는 준주기이고, cT는 감소인자이며 1보다 충분히 작다.

성문을 지난 신호는 성도를 거치면서 성도의 형태에 따라 몇 개의 공명주파수를 갖게 되는데 이 공명주파수와 대역폭을 2-극(pole) 여파기로 나타내면 성도에 대한 모델은 다음과 같이 구할 수 있다.

$$V(z) = \frac{1}{\prod^k [1 - 2e^{-c_i T} \cos(B_i T) z^{-1} + e^{-2c_i T} z^{-2}]} \quad (2)$$

식(2)에서 k는 포먼트의 개수이고, c_i 와 B_i 는 포먼트 주파수와 대역폭을 결정하는 값으로 대역폭의 좁다는 가정에서 실제 포먼트의 공명주파수와 대역폭은 각각 $B_i/2\pi$, $c_i/2\pi$ 로 결정된다. 성도를 통과한 신호는 마지막으로 입술을 통과하는데 입술에서의 방사특성에 대한 모델은 다음과 같다.

$$L(z) = 1 - z^{-1} \quad (3)$$

음원에 대한 모델을 E(z)로 하면 음성출력 X(z)는 다음과 같다.

$$X(z) = E(z)G(z)V(z)L(z) \quad (4)$$

이를 전극(all-pole)형 합성모델로 다시 정의하면 식(5)와 같이 표현가능하다.

$$X(z) = E(z) \frac{1}{A(z)} \quad (5)$$

여기서,

$$A(z) = \frac{1}{G(z)V(z)L(z)} \quad (6)$$

이 $A(z)$ 는 다음 관계식에 의해 음성으로부터 성도특성을 제거하고 음원을 이끌어 내는 가역여파기를 알 수 있다.

$$E(z) = X(z)A(z) \quad (7)$$

이상에서 살펴본 것과 같이 음성을 음원과 그 음원이 통과하는 성도여파기로 모델링하고 각각을 독립적으로 모델링함으로써 수학적으로 선형방정식의 해를 구할 수 있다[1][2][3].

3. 제안한 LSP 간격정보를 이용한 평탄화 알고리즘

3.1 자기상관법을 이용한 기술기 측정

단구간 자기상관 함수는 식(8)로 표현 가능하다[4][5][6].

$$\Phi_n(i, j) = \sum_{m=0}^{N-1-(i-j)} x_n(m)x_n(m+i-j), \quad 1 \leq i \leq p, 0 \leq j \leq p \quad (8)$$

여기서, 식(8)의 일부를 식(9)와 같이 표현가능하다.

$$R_n(j) = \sum_{m=0}^{N-1-j} x_n(m)x_n(m+j) \quad (9)$$

$$\sum_{j=1}^p a_j \Phi_n(i, j) = \Phi_n(i, 0), \quad \text{for } i = 1, \dots, p \quad (10)$$

자기상관법을 이용하여 식(10)을 풀면 다음과 같이 표현된다.

$$\begin{vmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & \cdot & \cdot & R_n(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ R_n(p-1) & \cdot & \cdot & R_n(0) \end{vmatrix} \begin{vmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{vmatrix} = \begin{vmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{vmatrix} \quad (11)$$

$p=1$ 에 대하여 위의 식을 정리하면 다음과 같은 식으로 표현 가능하다.

$$\alpha_1 = \frac{R_n(1)}{R_n(0)} \quad (12)$$

<그림 1>, <그림 2>는 음성신호와 식(12)를 이용하여 주파수 영역에서 측정된 음성신호의 기울기를 나타낸 것이다. 임의의 음성 프레임에 대한 기울기 측정결과, <그림 1>의 경우에는 대략 -20 dB/decade이다. <그림 2>의 경우에 식(12)를 이용하여 주파수 영역에서 기울기를 측정한 결과, 대략 20 dB/decade로 나타났다.

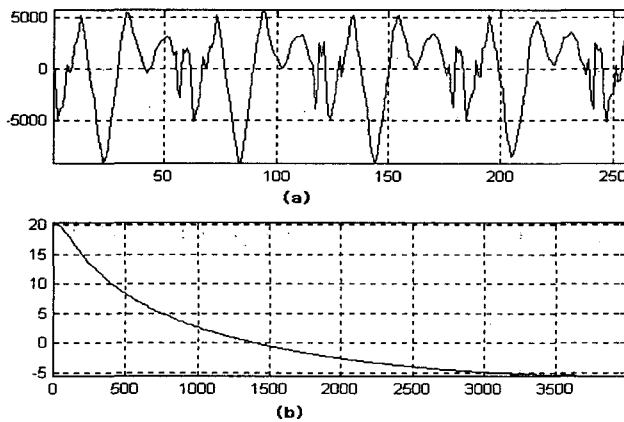


그림 1. 유성음 프레임에 대한 실험결과
(a)음성신호 (b)측정된 기울기

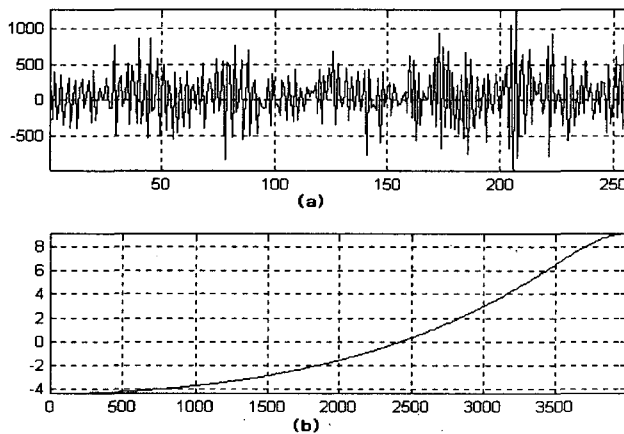


그림 2. 무성음 프레임에 대한 실험결과
(a)음성신호 (b)측정된 기울기

<그림 3>, <그림 4>는 주파수 영역에서 프레임 단위로 기울기를 측정하여 프리-엠퍼시스 필터의 a 값 조정을 통하여 나타난 주파수 영역의 스펙트럼이다.

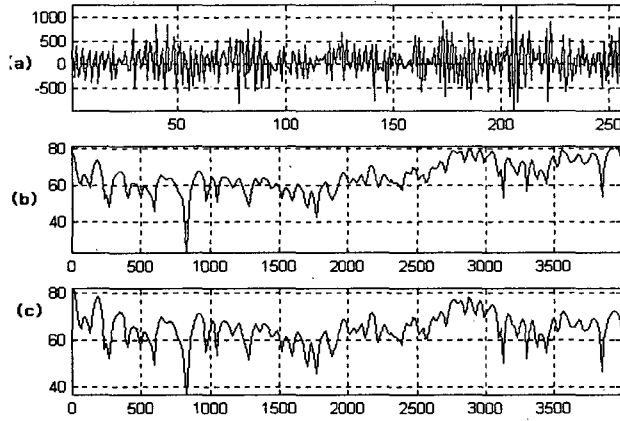


그림 3. 측정된 기울기를 이용하여 a값을 조절한 결과(1)
(a)음성신호 (b) 스펙트럼분석 (c)조절된 a값을 이용한 스펙트럼분석

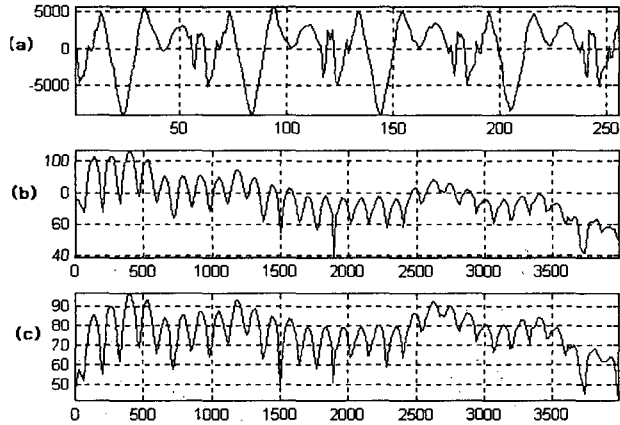


그림 4. 측정된 기울기를 이용하여 a값을 조절한 결과(2)
(a)음성신호 (b) 스펙트럼분석 (c)조절된 a값을 이용한 스펙트럼분석

3.2 제안한 기울기 추정 알고리즘

<그림 5>를 통해 제안한 알고리즘의 과정을 살펴보면 다음과 같다. LSP를 구하는 과정을 보면 음성신호를 입력받고 각 프레임마다 단구간 분석을 한다. 해밍 윈도우를 취한 데이터를 이용해 선형 예측 계수를 찾아내고, 그 다음 과정에서 LSP 파라미터를 얻어낼 수 있다. 10 차의 LSP 파라미터는 각각 주파수 위치를 나타내고 있으며 두 파라미터의 위치간격으로 포먼트가 위치한 주파수 위치를 알아낼 수 있다. 즉, 제안한 방법에서는 3 개의 LSP 파라미터를 하나의 단위로 하여 첫 번째와 두 번째의 간격정보, 두 번째와 세 번째의 간격정보의 값을 추출한다. 이것은 아래 식(14),(15)에 해당한다. 또한 식(13)은 첫 번째 파라미터와 세 번째 파라미터의 간격정보를 나타낸다[8].

$$LSP(j) = LSP(j+2) - LSP(j) \quad (13)$$

$$LSP_2(j) = LSP(j+2) - LSP(j+1) \tag{14}$$

$$LSP_1(j) = LSP(j+1) - LSP(j), \quad 1 < j < 8 \tag{15}$$

$$\alpha_2(j) = \frac{LSP_2(j)}{LSP(j)} \tag{16}$$

$$\alpha_1(j) = \frac{LSP_1(j)}{LSP(j)}, \quad 1 < j < 8 \tag{17}$$

계산에서 얻어진 α_1 과 α_2 중에서 문턱 값보다 작은 값을 선택하고, 그 주파수 간격 정보를 알아내어 저장한다. 이를 순차적으로 반복하여 포먼트의 위치정보를 얻어낼 수 있다. 문턱 값보다 작은 대역 폭을 가진 인접한 두 LSP 파라미터를 $LSP_{(det)}$ 라 하면 식(18)은 포먼트 주파수에 해당한다.

$$\alpha(n) = \frac{LSP_{(det)}}{2} \tag{18}$$

위의 과정을 통해 얻어진 포먼트 주파수의 개수가 n개라고 가정하고, 주파수 영역에서의 좌표를 나타내면 (a(1),DB(a(1)),(a(2),DB(a(2))),...(a(n),DB(a(n)))이다.

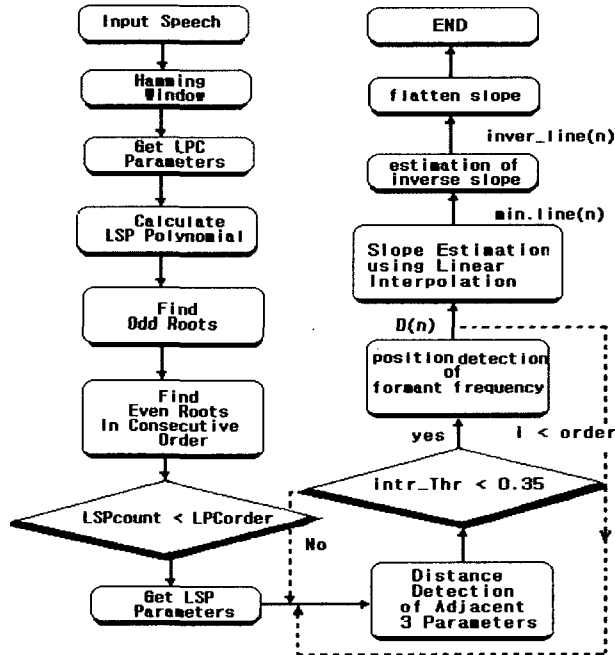


그림 5. LSP 파라미터를 이용한 평탄화 알고리즘

$$I(m) = DB(a(n)) + (m - a(n))f'(a(n)) + \frac{(m - a(n))^2}{2} f''(a(n)) \dots \quad (19)$$

$$f'(a(n)) = \frac{DB(a(n+1)) - DB(a(n))}{a(n+1) - a(n)} \quad (20)$$

식(19), (20)을 이용한 선형 인터폴레이션 과정을 거쳐 주파수영역에서 정확한 선형적 기울기를 추정할 수 있다[7]. 구해진 기울기를 이용하여 스펙트럼 기울기를 평탄화 시키기 위해서는 기울기의 역수인 식(22)를 이용한다. 또한, 식(23)은 주파수 영역에서 평탄화된 스펙트럼의 결과를 나타낸다.

$$\min_{data} = \min(DB(a(n))) \quad (21, 22, 23)$$

$$inver_I(m) = I(m) - 2(I(m) - \min_{data})$$

$$flatten(m) = DB(m) + inver_I(m), \quad 0 < m < frame$$

4. 실험결과

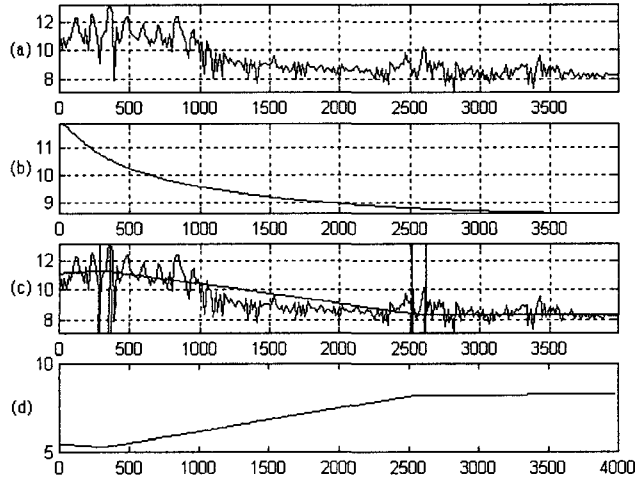


그림 6. 유성음 프레임 구간의 기울기 추정 결과

- (a) 음성신호의 스펙트럼분석
- (b) 자기 상관법을 이용한 기울기추정
- (c) 제안한 방법을 이용한 기울기 추정
- (d) 제안한 방법의 역기울기 추정

시뮬레이션을 위해 사용한 음성시료는 뉴스에서 발췌한 남/녀 아나운서 음성으로 8 kHz로 표본화 하고 16 비트로 양자화된 신호이다. 한 프레임의 길이는 256 표본이며 계수는 10 차를 써서 특징벡터를 추출하였다. <그림 6>에서 (a)는 주어진 음성시료 중 유성음에 대한 주파수 영역의 특성을 나타내고, (b)는 기존의 자기상관법을 이용하여 측정한 기울기이다. (c)는 제안한 방법의 실험결과로써, LSP를 추출하여 포먼트 위치를 추정하고 선형 인터폴레이션에 의해 기울기 추정법을 이용하여 구한 기

울기와 주파수 영역의 특성을 나타내고, (d)는 스펙트럼 평탄화를 위한 역 기울기를 구한 그래프이다. <그림 7>은 유성음에 대해 기존의 방법과 제안한 방법 간의 평탄화된 스펙트럼을 비교 분석한 결과이다. <그림 8>은 무성음에 해당하는 프레임에 제안한 방법에 의해 측정된 기울기를 나타낸 그림이고, <그림 9>는 제안한 방법에 의해 주파수 영역에서 보상된 결과를 나타내고 있다. <그림 6>~<그림 9>를 통해 나타난 실험결과에서 제안한 방법이 기존의 자기 상관법보다 평탄화 측면에서 우수함을 알 수가 있다. 예를 들어 <그림 7> (c)의 스펙트럼과 (d)의 제안한 방법을 이용한 스펙트럼을 비교 분석했을 때, 고주파 영역에서 약화된 신호를 평탄화하는 측면에서 우수함을 알 수가 있다.

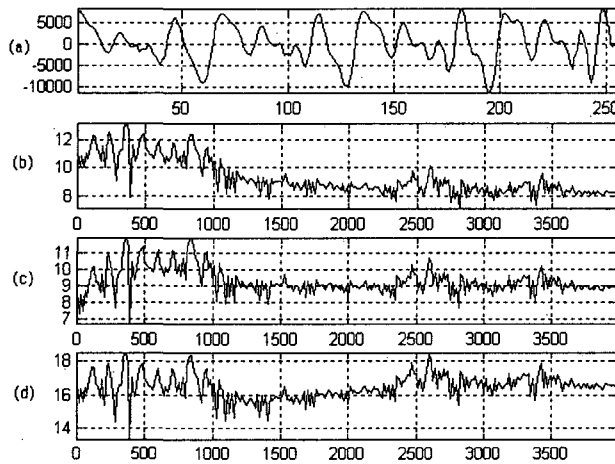


그림 7. 유성음 프레임의 스펙트럼 분석 결과

(a) 음성신호 (b) 원음성의 스펙트럼
(c) 자기 상관법을 이용한 스펙트럼 (d) 제안한 방법을 이용한 스펙트럼

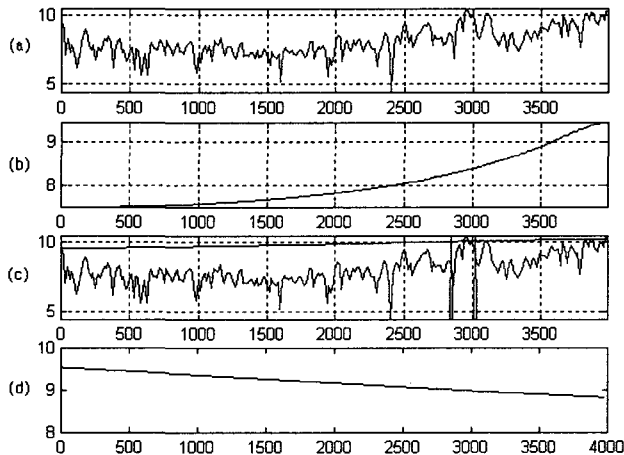


그림 8. 무성음 프레임 구간의 기울기 측정 결과

(a) 음성신호의 스펙트럼분석 (b) 자기 상관법을 이용한 기울기 측정
(c) 제안한 방법을 이용한 기울기 측정 (d) 제안한 방법의 역기울기 측정

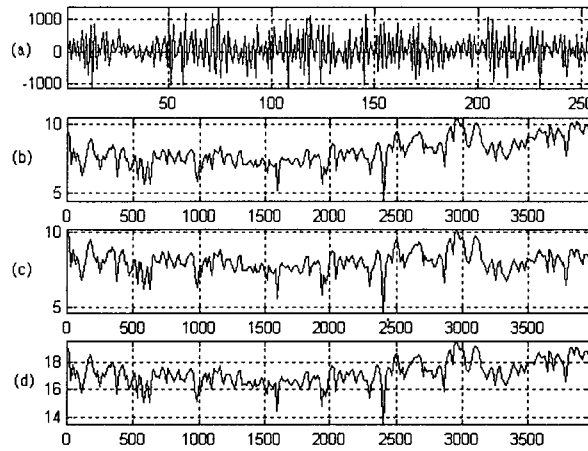


그림 9. 무성음 프레임의 스펙트럼 분석 결과

- (a) 음성신호 (b)원음성의 스펙트럼
 (c) 자기상관법을 이용한 스펙트럼
 (d) 제안한 방법을 이용한 스펙트럼

5. 결 론

본 논문에서는 주파수 영역에서 선형 예측 쌍 파라미터의 간격 정보를 사용하여 포먼트 주파수를 찾아내어 기존의 자기 상관법에 비해 세밀한 선형화된 기울기를 추정하고 이를 이용하여 약화되어진 고주파 특성을 보상하는 새로운 방법을 제안한다. 주파수 영역의 포락선 기울기를 평탄화해줌으로써 신호 대 잡음비를 높이고 음질향상에 유효하다. 임의의 음성 프레임 구간에 대한 실험 결과에서, 제안한 방법을 이용한 경우, 음성 프레임 구간 특성을 적용적으로 반영하여 고주파 영역에서 약화된 특징을 강조함으로써 평탄화된 스펙트럼을 추출할 수가 있었다.

감사의 글

본 논문은 2005년도 서일대학 학술연구비에 의해 연구되었습니다.

참 고 문 헌

- [1] 배명진. 1998. *디지털 음성분석*, 동영출판사, pp. 38-42, 108-118.
 [2] Parsons, T. 1986. *Voice and Speech Processing*, McGraw-Hill, pp. 262-267.
 [3] Rabiner L. R. & Juang, Biing-Hwang. 1993. *Fundamentals of Speech Recognition*, Prentice-

Hall AT&T, pp. 112-113.

- [4] Kondo, A. M. 1994. *Digital Speech*, John Wiley & Sons, pp. 44-46, 79-95.
- [5] Shaughnessy, D. O. 1996. *Speech Communication*, IEEE Press, pp. 188-191.
- [6] Rabiner, L. R. & Schafer, R. W. 1978. *Digital Processing of Speech Signal*, Prentice Hall, pp. 38-48.
- [7] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992. *Numerical Recipes in C*, Cambridge University Press, 2nd Edition, pp. 108-110.
- [8] 민소연, 장경아, 배명진. 2002. "선형스펙트럼쌍을 이용한 성문특성이 제거된 성도특성 추출법에 관한 연구." *한국음향학회지*, 제21권 제7호.

접수일자: 2005. 11. 08

게재결정: 2005. 12. 01

▲ 민소연

서울시 중랑구 면목8동 49-3 서일대학 (우: 131-702)

정보기술계열 정보통신 전공

Tel: +82-2-490-7583

E-mail: symin@seoil.ac.kr

▲ 장경아

서울시 동작구 상도5동 181-364번지 201호 (우: 156-095)

(주)코아로직 소프트웨어실 AT(Advanced Technology)팀 선임연구원

Tel: +82-2-2016-5673

E-mail: kajang74@corelogic.co.kr