We present a detection-based automatic speech recognition (ASR) paradigm that is capable of integrating both the knowledge sources accumulated in the speech science community and the modeling techniques established in the speech processing community. By exploring this new framework, we expect that researchers in the Interspeech community can collaboratively contribute to developing next generation algorithms that have the potential to surpass current capabilities, and go beyond the limitations of the state-of-the-art ASR technologies.

Chin-Hui Lee is a professor at School of Electrical and Computer Engineering at Georgia Institute of Technology. Dr. Lee received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, in 1973, the M.S. degree in Engineering and Applied Science from Yale University, New Haven, in 1977, and the Ph.D. degree in Electrical Engineering with a minor in Statistics from University of Washington, Seattle, in 1981.

After graduation, Dr. Lee joined Verbex Corporation, Bedford, MA, and was involved in research on connected word recognition. In 1984, he became affiliated with Digital Sound Corporation, Santa Barbara, where he engaged in research and product development in speech coding, speech synthesis, speech recognition and signal processing for the development of the DSC-2000 Voice Server. Between 1986 and 2001, he was with Bell Laboratories, Murray Hill, New Jersey, where he became a Distinguished Member of Technical Staff and Director of the Dialogue Systems Research Department. His research interests include multimedia communication, multimedia signal and information processing, speech and speaker recognition, speech and language modeling, spoken dialogue processing, adaptive and discriminative learning, biometric authentication, information retrieval, and bioinformatics. His research scope is reflected in "Automatic Speech and Speaker Recognition: Advanced Topics", published by the Kluwer Academic Publishers in 1996. From August 2001 to August 2002 he was a visiting professor at School of Computing, The National University of Singapore. In September 2002, he joined the Faculty of School of Electrical and Computer Engineering at Georgia Institute of Technology.

Prof. Lee has participated actively in professional societies. He is a member of the IEEE Signal Processing Society, Communication Society, and the European Speech Communication Association. He is also a lifetime member of the Computational Linguistics Society in Taiwan.

In 1991-1995, he was an associate editor for the IEEE Transactions on Signal Processing and Transactions on Speech and Audio Processing. During the same period, he served as a member of the ARPA Spoken Language Coordination Committee. In 1995-1998 he was a member of the Speech Processing Technical Committee of the IEEE Signal Processing Society (SPS), and later became the chairman of the Speech TC from 1997 to 1998. In 1996, he helped promote the SPS Multimedia Signal Processing (MMSP) Technical Committee in which he is a founding member.

Dr. Lee is a Fellow of the IEEE, and has published more than 250 papers and 25 patents on the subject of automatic speech and speaker recognition. He received the SPS Senior Award in 1994 and the SPS Best Paper Award in 1997 and 1999, respectively. In 1997, he was also awarded the prestigious Bell Labs President's Gold Award for his contributions to the Lucent Speech Processing Solutions product. In 2000, he was named one of the six Distinguished Lecturers by the IEEE Signal Processing Society.
From Knowledge-Ignorant to Knowledge-Rich Modeling: A New Speech Research Paradigm for Next Generation Automatic Speech Recognition

Chin-Hui Lee
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
chl@ece.gatech.edu

Abstract
The field of automatic speech recognition (ASR) has enjoyed a fast technology progress in the last three decades, due to the extensive use of statistical learning algorithms, the availability of a number of large collections of speech and text examples, and fast computing machines. However ASR advances have slowed down quite a bit in recent years. It seems the success of the above knowledge-ignorant modeling approach can be further extended if knowledge sources available in the large body of speech science literature can be properly integrated into the statistical modeling paradigm and objectively evaluated. In this paper we explore a knowledge-rich, data-driven approach to ASR that serves as a candidate paradigm for developing next generation ASR techniques and systems. A few knowledge-supplementary examples will first be illustrated. Some potential collaborative research scenarios will also be discussed.

1. Introduction
Speech is considered as the most natural means of communication among human beings. There is also a rich set of human information embedded in speech beyond just word transcription of the spoken utterances. Mining of speech information is therefore of great importance both in theory and in practice. One way to extract such information from spoken languages by machine is to convert speech to text through automatic speech recognition (ASR). With the increasing research and application interests, the field of ASR has enjoyed a fast progress in the last three decades, due to the extensive use of statistical learning techniques, the availability of large collections of speech and text examples, and fast computing machines. However ASR advances have slowed done quite a bit in recent years. It seems the past successes of the prevailing knowledge-ignorant modeling approach can still be further extended if knowledge sources available in the large body of literature in speech and language sciences can be objectively evaluated and properly integrated into statistical modeling. In this paper we explore a knowledge-rich, data-driven modeling paradigm that is capable of going beyond the current limitations of the state-of-the-art ASR technology, and gradually bridging the performance gap between ASR and HSR (human speech recognition). The implied detection-based approach to ASR, taking advantage of both knowledge-based and data-driven modeling paradigms, serves as a candidate paradigm for developing next generation ASR algorithms and systems.

2. Knowledge-Ignorant Modeling
The statistical modeling approach to ASR is motivated by expressing spoken utterances as stochastic patterns. ASR is then accomplished by finding the sequence of words that maximizes the joint probability, \( P(S, W) \), of a given spoken utterance, \( S \), and the corresponding word sequence, \( W \), assuming that \( P(S, W) \) is known. Based on statistical decision theory, it can be shown that the above ASR solution agrees with the optimal Bayes rule that minimizes the total risk of the expected sentence, word, or phone error rates, depending on the underlying problems of interest [1].

Next we consider the Shannon’s channel modeling paradigm that a given sequence of input symbols, \( I \), is passed through a noisy channel, and converted into an output signal, \( O \). We are interested in recovering \( I \) from \( O \), by designing an optimal channel decoder that maximizes the joint probability of \( I \) and \( O \),

\[
P(I, O) = P(O|I)P(I),
\]

where \( P(O|I) \) is the conditional probability of \( O \) given \( I \), and \( P(I) \) is the prior probability of \( I \). Therefore the two ASR perspectives, one based on designing the optimal Bayes decision rule, and the other on optimal channel decoding [2], give the same solution in implementing the maximum a posteriori (MAP) decoding policy that maximizes \( P(W|S) \), or similarly \( P(I|O) \). This powerful tool has been applied to many speech and language processing applications. Since we don’t have an exact knowledge of the joint probability distributions for most practical problems, the forms of the distribution functions of \( P(O|I) \) and \( P(I) \) are often assumed, and their corresponding distribution parameters are then estimated from a large collection of application specific training examples. Therefore, the recognized input sequence is solved by implementing a plug-in MAP decoding policy [1] that plugs the estimated parameters into the assumed distributions in order to evaluate the required probabilities, or likelihood, in MAP decoding. Clearly, the optimal policy in decision and channel decoding no longer holds.

In most speech and language processing problems, the input symbol sequences, such as words, concepts, or part-of-speech tags, can often be approximated by Markov chains, while the output observations are either continuous signals, like speech, or discrete signals, like word sequences. Due to the fact that the input symbols are hidden and not directly observed Markov sources, the output observations can now be considered as discrete or continuous density hidden Markov models (HMMs) [3]. This partly explains the success of using HMMs in many recently reformulated pattern classification applications, even if the observed signals are not necessarily generated by hidden Markov sources. Designing the channel decoder, or computing the acoustic model, \( P(S|W) \) and language model \( P(W) \), demands modeling of \( P(O|I) \) and \( P(I) \), which is often accomplished by collecting a large training set of input-output pairs, and applying statistical learning techniques to estimate all the distribution parameters required to evaluate the two probabilities, \( P(O|I) \) and \( P(I) \). It is clear that the abovementioned approach does not
require any detailed specifications of the input symbols being decoded. With such a knowledge-ignorant modeling approach, it is now quite straightforward to demonstrate ASR capabilities of new tasks for almost any spoken language, without using detailed descriptions about the language. There are now available a vast collection of speech and language corpora, sponsored by many business and government-funded projects in many countries. Advances in hardware, algorithms and data structures have also made implementation of large vocabulary, continuous speech recognition (LVCSR) systems affordable.

3. Advances in Hidden Markov Modeling
Knowledge-ignorant modeling is considered as one of the most fruitful areas in characterizing speech and language in recently years. Key advances [1, 4] can be summarized in three broad topics, namely: (1) Detailed Modeling - Software packages [5] are available now in public domains to establish acoustic models with hundred of thousand Gaussian mixture components, and language models with hundred of million of n-gram probabilities. The previous limitation imposed by the curse of dimension, widely known in the pattern recognition community, was alleviated with many advanced modeling techniques that take parameter sharing into account, such as the commonly-used tied-state tree learning strategy in phone modeling; (2) Adaptive Modeling - Adaptive learning of HMM parameters, such as MAP adaptation for parameters of phones and their corresponding structures, is now a standard practice in many systems. Online adaptation of HMM parameters has also been developed to improve learning efficiency and effectiveness with little data; and (3) Discriminative Modeling – Using learning criterion that is consistent with speech recognition and verification objectives, minimum classification error (MCE) and minimum verification error (MVE) learning algorithms for HMM parameters have been shown quite effective in improving model separation, system accuracy, and performance robustness.

4. Limitations with Knowledge-Ignorant Models
Based on the above statistical modeling paradigm we have learned a great deal about how to build practical ASR systems for almost any spoken language without the need of a detailed understanding of the language. However these existing systems are often overly restrictical, requiring that their users have to follow a very strict set of protocols to effectively utilize spoken language applications. Furthermore, the ASR system accuracy often declines dramatically in adverse conditions to an extent that an ASR system becomes unusable, even for cooperative users. When compared with HSR [6], the state-of-the-art ASR systems usually give much larger error rates even for rather simple tasks operating in clean environments. In highly noisy conditions, such as those in moving vehicles, ASR often gives an error rate more than one-two orders of magnitude higher than HSR [7]. Such a performance gap is unacceptable to users and makes the work of application designers extremely difficult.

In addition to the robustness problem with adverse acoustic mismatches, in most real-world environments spoken messages are usually conveyed with spontaneous speech which is often “ill-formed”, with many utterances containing out-of-task, out-of-grammar, and out-of-vocabulary speech segments. Since we don’t have a complete specification to characterize all possible ways of such spoken expressions using finite state network representations of all needed knowledge sources, the commonly-adopted MAP decoding policy for ASR is no longer applicable to this so called “open-set” robustness problem with adverse linguistic mismatches. Therefore the conventional ASR notion, of a complete word transcription of any spoken sentence by any person speaking in any acoustic condition with any language, cannot really be achieved using the current state-of-the-art ASR paradigm. Instead, we need to explore new formulations that can facilitate partial understanding of spoken languages, just like in the case of human speech understanding without the need of recognizing every single sound in a spoken utterance.

However efforts in integrating detailed knowledge, from acoustics, speech, language and their interactions, are hampered by the current ASR formulation as a “blackbox” of models trained to “remember” the training data, because it is not straightforward to integrate all available knowledge sources into the current top-down, knowledge-ignorant modeling framework. This makes it difficult for the ASR community to take advantage of the vast body of literature developed in the speech and language science communities to improve the performance and robustness of ASR systems. Thus, a collaborative paradigm is needed to facilitate innovations and lower entry barriers to ASR research for every single individual or group interested in contributing to ASR advances.

5. Knowledge-Supplemental Modeling
In the last few years, we have witnessed an increasing awareness of attempting to integrate limited knowledge sources into state-of-the-art HMM-based ASR systems to improve recognition accuracy. Three such examples are illustrated as follows, namely: (1) “Sound-Specific” Features [8] - A single voice onset time measurement, or VOT, was shown to be more powerful than 39 spectral features for discriminating a stop pair, such as /d/ vs. /t/. By re-ordering recognition candidates according to VOT, a two-stage alphabet recognizer gave a 50% error rate reduction over state-of-the-art ASR results; (2) Key-Phrase Spotting Mimicking “Foreign Ears” [9] - Human listeners are very good in detecting relevant keywords, buried in utterances of a foreign language. This suggests using a detection approach to mimic keyword spotting by “foreign ears”, or poor ASR systems. For a spontaneous speech application, such a combined keyword detection and utterance verification [10] strategy maintained good accuracy for in-grammar utterances and greatly reduced errors for ill-formed sentences; and (3) “Knowledge-Based” Front-End [11] - An LVCSR system was built based on speech attributes produced by artificial neural network detectors. These “knowledge-based” features were then used to train a set of HMMs. By merging the MFCC baseline system with systems built with 60 attributes and 44 phone features using a ROVER combination [12], we obtained a word error rate of 3.7% for the WSJ 5K test used in Nov92 evaluation, about 20% relative error reduction over the best baseline system.

Other notable efforts include two 2003 symposiums and a recent study at the JHU Summer Workshop: (1) Perspectives on Speech Separation, http://www.ebire.org/speechseparation/; (2) Symposium on Next Generation Automatic Speech Recognition, http://www.ece.gatech.edu/~chl/ngasr03; and (3) Landmarks-Based ASR, http://www.clsp.jhu.edu/ws2004/groups/ws04ldmk/.

6. Human-Based Models for Speech Processing
It is interesting to note that human beings perform HSR by integrating multiple knowledge sources from bottom up. It has
long been postulated that a human determines the linguistic identity of a sound based on detected evidences that exist at various levels of the speech knowledge hierarchy, from acoustics to pragmatics. For example, Klatt [13] studied the so-called acoustic landmarks that are assumed invariant to changes in speakers and speaking environments. Stevens [14] and Fant [15] have consistently advocated the approach of detecting and recognizing distinctive features in speech sound from an acoustic-phonetic framework. Indeed, people do not continuously convert a speech signal into words as an ASR system attempts to do. Instead, they detect acoustic and auditory evidences, weigh them and combine them to form cognitive hypotheses, and then validate the hypotheses until consistent decisions are reached. This process has been successfully demonstrated in spectrogram reading by trained experts based on knowledge in acoustic-phonetics [16]. Furthermore, a phonological parsing paradigm [17] for ASR has been proposed by assuming all the distinctive features can be exactly detected. However, these features are not widely used in speech recognition due to the fact that they cannot be reliably detected in continuous speech, especially in adverse acoustic conditions.

In order to bridge the performance gap between ASR and HSR systems, it seems clear that the narrow notion of speech-to-text in ASR has to be expanded to incorporate all related human information “hidden” in speech utterances. This collection of information includes a set of fundamental speech sounds and their linguistic interpretations, the speaking environment that describes the interaction between speech and acoustics, a speaker profile that encompasses gender, accent and other speaker characteristics, such as the emotional state, etc. Collectively, we call this set of speech information, speech attributes. They are not only critical for ASR but also useful for many other applications, including speech coding, speaker recognition, language identification, speech perception, and speech synthesis. Based on this set of speech attributes, ASR can be extended to Automatic Speech Attribute Transcription, or ASAT, a process that goes beyond the current notion of just word transcription.

7. A Detection-Based ASR Paradigm

The above human-based model of speech processing suggests a candidate framework for developing next generation speech technologies that have the potential to go beyond the current limitations. The missing link in utilizing acoustic and linguistic knowledge sources to recognize speech lies in designing a bank of feature detectors that is mathematically rigorous and capable of producing consistent detection results, even in adverse conditions. These robust event detectors are designed using acoustic-phonetic knowledge but are stochastic in nature so that the principles of statistical hypothesis testing and data-driven modeling techniques (as successfully adopted in the top-down, knowledge-ignorant ASR systems) can be extended to such a bottom-up, knowledge-rich modeling approach. These detected “events” can then be combined into higher level knowledge and evidences for phone and word recognition in a probabilistic manner, using hypothesis testing theories and computationally efficient algorithms. Furthermore, there is no need to restrict feature extraction at a fixed frame rate. Analog detectors and other related biologically-motivated processors [18] can also be incorporated into this flexible framework. The methodology distinguished itself from prior acoustic-phonetic approaches [13] practiced in the 1970’s in the consistent use of data-driven designs for speech event detection and in the way the detected cues are fused into higher level evidence in linguistic knowledge integration with hypothesis testing and pattern verification [19].

A block diagram of the proposed ASAT paradigm for collaborative speech research is shown in Figure 1. It outlines a four-fold proposition that: (1) We can build a bottom-up speech recognition system that combines information from articulatory and acoustic-phonetic features to form phones, words, and word sequences; (2) We can do it in a modular fashion that can facilitate plug-’n’-play interoperability, allowing for close collaboration across a wide range of groups; (3) We can encourage new speech applications beyond the traditional notion of word transcription, such that researchers in both speech science and speech processing communities can contribute to technology advances; and (4) We can continue to practice the objective evaluation methodology [20] commonly adopted in the ASR community to track technology progress by developing similar evaluation strategies for individual modules and overall system to monitor detection performance, evidence combination effectiveness, as well as feature, phone and word accuracies.

Furthermore, the proposed bottom-up detection paradigm implies a new approach to solving the robustness problems by using a “divide and conquer” strategy. It also enables us to take advantage of the vast body of literature developed outside the ASR community. Knowledge in speech production and auditory processing and perception can also be applied to detection-based ASAT and ASR. By providing a plug-‘n’-play platform, we hope to encourage the many researchers that have worked on the rule-based induction, acoustic-phonetic feature detection, and machine learning techniques in ASR and other speech areas to combine their efforts into such a single system. The proposed approach, when applied to auditory processing, attempts to simulate the human auditory process by assuming that speech is first converted to a collection of auditory response patterns (feature detection), each modeling the probabilistic activity level of a particular acoustic-phonetic event of interest. Knowledge sources and computational models in neuroscience [21] can also be extensively utilized. Detection of the next level of events or evidence, such as phones and words, is accomplished by combining relevant features. Each activity function can be modeled by a corresponding neural system. Both the activation levels and firing rates have been used in neural encoding. Artificial neural networks [22] provide a convenient tool to model neuron combinations. Feedforward neural networks have been used to encode and decode temporal information. Recurrent neural networks have also been used to provide feedback loops to simulate neural processing. Simulating perception of temporal speech events is of particular interest.

Figure 1  Bottom-up ASAT based on speech attribute detection, event merging and evidence verification
8. Summary

In summary, the ASR community is now at a crossroad searching for new directions. We are exploring a new knowledge-rich, data-driven modeling approach to next generation automatic speech recognition under a recently awarded NSF ITR grant: Automatic Speech Attribute Transcription (ASAT): A Collaborative Speech Research Paradigm and Cyberinfrastructure with Applications to Automatic Speech Recognition. It is clear that we have a long way to go before we can develop a complete ASR system that is competitive with the state-of-the-art performances. However, we do believe such a detection-based paradigm is flexible and rich in features, and provides an excellent vehicle for collaborative speech research. To facilitate such a community effort we intend to develop an open and sharable platform, and make all the system modules and tools available to the broader speech research community.

It is believed that by incorporating knowledge sources into speech modeling and processing, the set of recognized attribute sequences, event lattices, and evidences for decision in Figure 1 provides an instructive collection of diagnostic information, potentially beneficial for improving our understanding of speech, as well as enhancing speech recognition accuracy. From some of our preliminary results, we found that the recognized errors produced in knowledge-based, data-driven modeling systems often corresponded to more meaningful confusion of sounds in the same broad phonetic class than those errors obtained with knowledge-ignorant modeling systems, although the features are not necessarily more discriminative in classifying these speech sounds. Based on this set of information, we believe a better set of attribute detectors can be designed and they will contribute to improving both modular and overall system performances.

It is also noted that the performance in the proposed system is “additive”. For example a better module for a feature will not produce poorer performance for the individual module and other modules related to this attribute, including the overall system. To facilitate a community effort to monitor research progress we will design a collection of evaluation sets for each speech attribute. Corresponding performance history will also be documented. Everyone is welcome to participate in this effort. We hope to eventually obtain a collection of “best” modules collectively provided by the community for all the needed features, so that they can be collaboratively incorporated into the “best” overall ASR system of the next generation.

Acknowledgement

The ASAT team includes M. Clements, K. Johnson, E. Fosl-Lussier, L. Rabiner, and C.-H. Lee. The author benefits greatly from many interactions with B.-H. Juang, and appreciates his continuing collaboration. The author also owes his sincere gratitude to his past colleagues at Bell Labs, Murray Hill, for endless and stimulating discussions. Such precious opportunities will be forever missed. Finally, the author thanks his Georgia Tech students, Yu Tsao and Jinyu Li, for generating some of the figures and results used in his oral presentation. Part of their efforts was supported under the NSF SGER grant, IIS-03-96848.

Reference

Plenary Session 3  
09:00 - 10:00, Thursday, October 7, 2004  
In Search of a Universal Phonetic Alphabet - Theory and Application of an Organic Visible Speech -
Hyun-Bok Lee, Seoul National University

Abstracts

Phonetic symbols have an important role to play in phonetics, linguistics, language teaching, speech pathology and speech sciences in general, and linguists and phoneticians have tried to devise appropriate phonetic alphabets. Notable among them are Sweet, Bell, Jespersen, Pike, etc. The most successful and popular phonetic alphabet today is no doubt the International Phonetic Alphabet. The International Korean Phonetic Alphabet (IKPA for short) is a system of phonetic symbols that has been devised by the author on the basis of the articulatory phonetic(organic) principles exploited by the Korean King Sejong in creating the Korean alphabet of 28 letters in 1443. The Korean alphabet is not merely a phonetic alphabet of arbitrary nature but a highly sophisticated system consisting of sets of interrelated organic phonetic symbols, each set representing either the shape of the organs of speech, i.e. lips, tooth and velar etc. or their articulatory action. The Korean alphabet is, in a true sense of the word, a set of phonetic symbols designed to represent the organic visible speech of the human being. The author has applied the organic phonetic principles much more extensively and systematically in devising IKPA than the King had done. Consequently the IKPA symbols are just as systematic, scientific, easy to learn and memorize as the Korean alphabet, quite unlike the IPA counterparts which, having been derived mainly from Roman and Greek letters, are mostly unsystematic and arbitrary. The IKPA symbols visualize or mirror the actual speech organs or their action and thus tell us exactly what sort of an articulatory action is involved in producing sounds. It is in this sense that the IKPA deserves to be called a "Universal Visible Speech", which is to be shared by all.

Biographical Sketch

Education
Dept. of Linguistics, Seoul National Univ. (1955-59)  
Dept. of Linguistics, Graduate School, Seoul National Univ. (1959-61)  
Dept. of Phonetics, University College London (1962-64)  
Summer School at Universite Catholique de Paris (1965)  

Career
Librarian at the National Central Library of Korea (1959-62)  
British Council Scholar (1962-64)  
Lecturer (Phonetics/Korean) at Uppsala and Stockholm Univ. (1965-66)  
Fellow at SOAS, London University (1969-70)  
Professor of Phonetics & Linguistics, Seoul National Univ. (1970-2002)  
Director of Language Research Institute, Seoul National Univ. (1986-88)  
Adviser to Korean Broadcasting Station, Ministry of Education, and Ministry of Culture and Sports (1976- )  
Founder and President of the Phonetic Society of Korea (1976-2002)  
President of the Orioreal Society of Korea (1994-2000)  
President of the Linguistic Society of Korea (1997-99)  
Vice-President of the Korean Language Society (1980-2004)  
Visiting Professor at Tokyo Univ. (1977-78), London Univ. (1985-86)  
Warsaw Univ. (1990-92), Tokyo Univ. (2000.5-8), Member of PC-ICSLP (2000- )  
Honorary Adviser to ICSLP Jeju Organizing Committee (2002- )  
Director of the Korean-English Speech Academy (2002- ).
In Search of a Universal Phonetic Alphabet
- Theory and Application of an Organic Visible Speech -

Hyun Bok Lee, CBE, Ph.D.
Professor Emeritus of Phonetics and Linguistics
Seoul National University
hanslee@snu.ac.kr

--- Contents ---
1. Introduction
2. Attempts to Devise Universal Phonetic Alphabets
3. Characteristics and Advantages of the Korean Alphabet
4. HanGeul: An Organic Alphabet of Distinctive Features
5. Principles of the International Korean Phonetic Alphabet
6. Conclusion and Illustration

Abstract

Phonetic symbols have an indispensable role to play in phonetics, linguistics, language teaching, speech pathology and speech sciences in general. And linguists and phoneticians in the east and west have made attempts to devise appropriate phonetic alphabets at one time or another in human history. The most successful and popular phonetic alphabet today is no doubt the International Phonetic Alphabet, which being based mainly on Latin and Greek letters, consists of unsystematic mass of arbitrary symbols.

Hunnmir Jeongeum, the original version of the Korean alphabet is a highly sophisticated system consisting of sets of interrelated organic phonetic symbols, each set representing either the shape of the organs of speech or their articulatory movements. The Korean alphabet is, in a true sense of the word, a set of phonetic symbols designed to represent the visible speech sounds of human beings.

In an attempt to devise an ideal and universal organic phonetic alphabet the author has applied extensively the organic principle that was exploited by the King Sejong of Korea in 1446 in creating Hunmir Jeongeum. The International Korean Phonetic Alphabet (IKPA for short) is a system of phonetic symbols, which is just as systematic, scientific, easy to learn and memorize as the Korean alphabet. The IKPA symbols visualize or mirror the actual speech organs or their action and thus tell us exactly what sort of an articulatory action is involved in producing speech sounds. It is in this sense that the IKPA is called "A Universal Visible Speech".

1. Introduction

Ladies and gentlemen! I welcome you all to Korea, the land of phonetics. Korea has been known to the outside world as the land of morning calm, or hermit kingdom for centuries. As you can see yourself, there is nothing really calm about Korea nowadays. In fact she has now turned out to be a highly hustling and bustling society armed with IT industry. Believe it or not she is now a leading country in internet and mobile phones. "Korea is engaged", as the weekly magazine Time aptly described on the cover page some time ago to dramatize the popularity of mobile phones in Korea.

More than anything else, however, Korea deserves to be known as "The Land of Phonetics and Spoken Language Processing", for it was here in Korea that the most remarkable phonetic alphabet was invented by the King Sejong in 1443. With the promulgation of Hunmir Jeongeum (Right sounds for teaching the nation), the Korean alphabet of 28 letters by the king in 1446, Korea was turned into the land of phonetics. The Korean alphabet, now known as "HanGeul" (Unique and great alphabet) was devised on the basis of articulatory and auditory phonetic principles as well as the modern phonologial concepts. It is an "Organic Alphabet", consisting of simple and yet versatile letters reflecting the shape of the articulatory organs in action. Therefore HanGeul has rightly earned the reputation of being "the greatest masterpiece of human intellect and a truly universal alphabet". Ladies and gentlemen, you are literally surrounded by organic phonetic symbols and visible speech wherever you are in Korea. Let me tell you what I mean by this.
2. Attempts to Devise Universal Phonetic Alphabets

Serious attempts have been made by phoneticians and linguists in the east and west to devise universal phonetic alphabets. Notable among the pioneers are Bell, Sweet, Jespersen etc. Their attempts were highly rewarding and yet their ideas were unfortunately all short-lived. For instance, the Scottish Educationalist, Alexander Melville Bell, father of Alexander Graham Bell, the inventor of the telephone, invented a system of representing the sounds of speech according to the way in which they are articulated, which he called *Visible Speech*. Henry Sweet a student of Melville Bell had improved upon this and adapted it to become what he called the *Organic Alphabet*, "based on a physiological analysis of the actions of the organs of speech".

It is highly interesting to remind ourselves on this occasion that Bell and Sweet appear to have to a large extent used the similar principles that the King Sejong of Korea applied in creating the Korean alphabet over five centuries ago. Just as the King Sejong used the symbol ㅏ in Hunmin Jeongeum as a component in the letters for the velar sounds [g] and [k], as resembling the shape of the tongue "blocking the throat", and ㅓ as a component in the letters for the lingual-dental sounds [l, d, n] to represent the tongue "touching the upper jaw", and ㅗ, representing the shape of the lips, in the labial sounds, so Sweet used a full circle 〇 in his Organic Alphabet to represent the open throat in breath, and the symbol ㅏ to represent the narrowed throat with closed vocal cords, as in the production of voice. Sweet's symbol for the teeth for dental consonants (such as English th in "thin") is ㅗ based on the shape of teeth. "Open", i.e., fricative, consonants are represented by a broken circle, stopped consonants by the broken circle closed by a bar. A voiced consonant is represented by adding the voiced "bar" to the corresponding voiceless consonant. Notice that the King Sejong used an additional stroke to derive aspirated plosive and affricate consonant graphemes in Korean alphabet: ㅏ → [k] → ㅗ → [kʰ], ㅗ → [c] → ㅗ → [tʃʰ].

According to E. Henderson, Sweet's organic Alphabet lacked, unfortunately, the royal support which ensured the adoption of Hunmin Jeongeum by the Korean nation, and printing difficulties obliged him in his later writings to use in his phonetic transcriptions the adapted forms of the Roman letters that linguists are familiar with today in the alphabet of the International Phonetic Association.

Of the phonetic alphabets so far devised, the IPA is no doubt the most widely used and at the same time, highly successful phonetic alphabet today. However, the IPA symbols, based mainly on Roman and Greek letters, have some serious disadvantages and drawbacks:

Roman [ a, i, e, o, u, p, t, k, s, z, l, x ] etc; Greek [ φ, β, v, θ, x ] etc; Modified [ o, e, æ, ø, ɪ, ʊ, η ] etc.

For one thing, the IPA symbols do not represent or reflect the shapes or movements of the organs of speech in the manner that the Korean letters do. Notice the formal resemblance of the shape of the back of the tongue blocking the soft palate to the Korean letter for the velar plosive sound ㅏ. Moreover, unlike in Korean alphabet, no formal interrelationship can be found in IPA between the phonetic symbols representing homorganic sounds such as p/b, v/d, k/q, s/z, f/v, etc. They are simply arbitrary and totally unrelated in shape. Notice the formal similarity among the three homorganic consonant letters in Korean in comparison to the formal dissimilarity between IPA [k] and [q].

Kor. ㅏ → [k] → ㅗ → [kʰ] → ㅜ → [k']
IPA. ㅓ → ㅗ

Consequently IPA symbols are much harder for beginners to learn and use than Korean letters.

3. Characteristics and Advantages of the Korean Alphabet

The Korean alphabet is unique in many respects and it certainly deserves to be more widely known and understood. The most characteristic features of Korean alphabet are as follows.

3.1 Phonetically Oriented

1) Consonant letters The shape of the basic letters was modelled on the actual shape of the articulatory organs involved in pronouncing the sound represented by the letters. The five basic consonant letters are created by the Korean King as follows:

ㅏ [k] represents the velar sound since it resembles the shape of the tongue blocking the throat.

ㅓ [n] represents the lingu al sound (dental/alveolar sound in modern phonetic terminology) since it resembles the tongue touching 'the upper jaw', i.e., upper teeth or teeth ridge.
\(<m\rangle\) represents the labial sound since it resembles the shape of the lips.
\(<\alpha\rangle\) [s] represents the dental sound since it resembles the shape of the teeth.
\(<\alpha\rangle\) represents the throat sound since it resembles the shape of the throat: [o/ə].

As pointed out by Henderson, this was exactly what Bell and Sweet had in mind when they tried to devise Organic and Visible Speech about five centuries later.

2) Vowel letters. Vowel letters were devised on the traditional oriental philosophical principles of In(negative) and Yang(positive). The King devised three basic letters \(<\cdot\rangle\), \(<\rightarrow\rangle\) and \(<\downarrow\rangle\), symbolizing respectively ‘heaven’, ‘earth’ and ‘man’ and assigning phonetic values to them as follows:

Three Basic Vowel Letters:
\(<\cdot\rangle\ [o]\ - \ "symbolizing heaven(round)"
\(<\rightarrow\rangle\ [u]\ - \ "symbolizing earth(flat)"
\(<\downarrow\rangle\ [i]\ - \ "symbolizing man”(upright)"

3.2 Systematic Derivation of Symbols

Most Korean graphemes are derived systematically from the basic consonant and vowel letters by addition of extra diacritical marks. For instance, the twelve remaining consonant letters were derived by adding to each of the five basic letters one or more additional strokes or symbols which indicated other relevant phonetic features or different manners of articulation at homorganic points of articulation.

\[\text{e.g.} \quad \langle\cdot\rangle \rightarrow \langle\rightarrow\rangle \rightarrow \langle\uparrow\rangle \rightarrow \langle\alpha\rangle \]
\[\langle n\rangle \rightarrow \langle d\rangle \rightarrow \langle t\rangle \rightarrow \langle t'\rangle \]

It is interesting to note that the basic letter \(<\cdot\rangle\[n]\) symbolizing dental/alveolar articulation is shared by all of the derived homorganic letters. Likewise, the remaining eight vowel letters were derived by different combinations of the three basic letters:

\(<\downarrow\rangle\ [a] \leftarrow \langle l\rangle \ + \ \langle\cdot\rangle \\
\(<\downarrow\rangle\ [\alpha] \leftarrow \langle l\rangle \ \ + \ \langle\cdot\rangle \ \ + \ \langle\cdot\rangle \\
\(<\downarrow\rangle\ [\alpha] \leftarrow \langle l\rangle \ \ + \ \langle\cdot\rangle \\
\(<\downarrow\rangle\ [o] \leftarrow \langle l\rangle \ \ + \ \langle\rightarrow\rangle \\
\(<\downarrow\rangle\ [u] \leftarrow \langle l\rangle \ \ + \ \langle\rightarrow\rangle \ \ + \ \langle\rightarrow\rangle \ etc.

3.3 Functioning as Phonemic Symbols

Korean alphabet( Hunmin Jeongeum), although formulated on a purely phonetic basis, was a phonemic alphabet in its actual application. And there is sufficient evidence (e.g., the recognition of three positions, initial, medial, and final, in the syllable and the statements concerning the distribution of sounds at the three positions of a syllable, etc.) that the king had completed some kind of preliminary phonological analysis of Korean according to the phonemic principle not far removed from that of modern linguistics, even though he did not actually use the term ‘Phoneme’ as against ‘Phone’ or ‘Sound’.

3.4 Syllable Block Writing

Another important characteristic of Korean alphabet is found in the spelling principle decreed by the king, according to which letters were to be combined, in accordance with the prescribed rule, into syllable blocks and not in a linear succession as in European languages. In other words, graphemes are arranged syllabically in such a way that each syllable has a distinct geometrical shape. For instance, syllable like \(<\text{mak}\>\ ‘\text{curtain}’\ and \(<\text{nun}\>\ ‘\text{eye}’\ would be arranged in actual writing as follows:

\(<\text{mak}\>\ : \ \uparrow \leftarrow <\cdot\rangle\ [m], \ \uparrow [a], \ \uparrow [k]>
\(<\text{nun}\>\ : \ \uparrow \leftarrow <\cdot\rangle\ [n], \ \uparrow [u], \ \uparrow [n]>

With the three characteristics of Hunmin Jeongeum taken into consideration, the Korean alphabet of the fifteenth century may be defined as a phonemic alphabet based on phonetic principles and spelled syllabically.
4. HanGeul: An Organic Alphabet of Distinctive Features

It is worth noting that the Korean alphabet HanGeul has a kind of distinctive theory incorporated in it. In fact one can see that Hunmin Jeongeum of 1446 was created on the basis of practically the same kind of distinctive theory that was initiated and developed in twentieth century by Jakobson, Chomsky and Halle in the 20th century. The notion of distinctive feature and binary opposition is clearly demonstrated by the articulatory and auditory (acoustic) description given in Hunmin Jeongeum of the phonetic value of the three basic vowels, which may be tabulated as follows (cf. Table 1):

<table>
<thead>
<tr>
<th>Features</th>
<th>Vowels</th>
<th>&lt;i&gt;</th>
<th>[a-um]</th>
<th>[u]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulatory</td>
<td>Retraction</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>(Tongue)</td>
<td>Advance</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Auditory</td>
<td>Shallow</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(Voice)</td>
<td>Deep</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

Notice that the two distinctive features "tongue retraction", which refers to the tongue movement (articulatory) and "voice deep" (dark and deep vocal quality as against bright and acute one), which relates to the auditory (acoustical) impression are shown to interact to characterize and define the vocalic quality of each of the three vowels as well as the phonetic relations among them.

4.1. The Korean Alphabet and the Prosodic Analysis

The current Korean alphabet, and in an even greater extent, fifteenth- century Hunmin Jeongeum, is not a haphazard collection of isolated letters. It is, in a sense, not only a phonetic alphabet based exclusively on detailed phonetic observations of the articulatory organs, but also a remarkably neat system composed of interrelated elements (letters). In particular, it is interesting to note the striking similarity between the manner in which the Korean alphabet is systematized and the theoretical tenet of the 'Prosodic Analysis' as initiated and developed by the London School, namely, a multidimensional approach characterized by the establishment of phonemic units on the one hand, and the abstraction and assignment of prosodic features to and over the phonemic units on the other. This can be exemplified by the Korean consonant letters. Of the 19 consonant letters, the following 17 letters are chosen for the purpose of this discussion (cf. Table 2):

<table>
<thead>
<tr>
<th>Features</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
<th>Voiceless glottalized</th>
<th>Voiced nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi-labial plosives</td>
<td>ι /p/</td>
<td>ʰ /pʰ/</td>
<td>ʰʰ /pʰʰ/</td>
<td>ʷ /m/</td>
</tr>
<tr>
<td>alveolar plosives</td>
<td>ι /t/</td>
<td>ʰ /tʰ/</td>
<td>ʰʰ /tʰʰ/</td>
<td>ʷ /n/</td>
</tr>
<tr>
<td>velar plosives</td>
<td>ι /k/</td>
<td>ʰ /kʰ/</td>
<td>ʰʰ /kʰʰ/</td>
<td>ʷ /ɡ/</td>
</tr>
<tr>
<td>affricates</td>
<td>ι /c/</td>
<td>ʰ /cʰ/</td>
<td>ʰʰ /cʰʰ/</td>
<td>ʷ /s/</td>
</tr>
<tr>
<td>alveolar fricatives</td>
<td>ι /s/</td>
<td>ʰ /sʰ/</td>
<td>ʰʰ /sʰʰ/</td>
<td>ʷ /l/</td>
</tr>
</tbody>
</table>

These 17 consonant letters can be described in prosodic terms as composed of five phonematic units, each representing a different place of articulation: (i) bi labial, (ii) alveolar, (iii) velar, (iv) post-alveolar (affricate) and (v) alveolar-fricative, and four prosodic features, (i) voiceless unaspirated, (ii) voiceless aspirated, (iii) voiceless glottalized, and (iv) voiced nasal. The analysis can be simplified considerably by symbolizing the phonematic units by the 'voiceless unaspirated' letters <H, ι, ι, ι, ι>, thereby reducing the number of prosodies from four to three. With the prosody 'voiceless unaspirated' treated as an unmarked term automatically ascribable to the five phonematic units. Symbolizing, by superscripts, the three prosodies as: ι[θ] – 'aspiration prosody'; ι[θ] – 'glottal prosody'; ι[θ] –
'nasal prosody', the 17 consonant letters may be represented as consisting of five phonematic unit and a zero or one of the three prosodies as follows (cf. Table 3):

<table>
<thead>
<tr>
<th>Prosody</th>
<th>Zero</th>
<th>Aspiration</th>
<th>Glottal</th>
<th>Nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi-labial plosive</td>
<td>(\text{-}(/p/))</td>
<td>(\text{ aspiration}=/p/')</td>
<td>(\text{ glottal}=/m/p/')</td>
<td>(\text{ nasal}=/m/)</td>
</tr>
<tr>
<td>alveolar plosive</td>
<td>(\text{-}(/t/))</td>
<td>(\text{ aspiration}=/t/')</td>
<td>(\text{ glottal}=/n/t/')</td>
<td>(\text{ nasal}=/n/)</td>
</tr>
<tr>
<td>velar plosive</td>
<td>(\text{-}(/k/))</td>
<td>(\text{ aspiration}=/k/')</td>
<td>(\text{ glottal}=/\text{n}/k/)</td>
<td>(\text{ nasal}=/\text{n}/)</td>
</tr>
<tr>
<td>affricates</td>
<td>(\text{-}(/c/))</td>
<td>(\text{ aspiration}=/c/')</td>
<td>(\text{ glottal}=/\text{x}/c'/)</td>
<td>(\text{ nasal}=/\text{x}/s/)</td>
</tr>
<tr>
<td>fricatives</td>
<td>(\text{-}(/s/))</td>
<td>(\text{ aspiration}=/s/')</td>
<td>(\text{ glottal}=/\text{a}/s'/)</td>
<td>(\text{ nasal}=/\text{a}/s'/)</td>
</tr>
</tbody>
</table>

In the respect to this kind of intricate internal structure of Korean consonant graphemes that is effected by distinctive features like "aspiration" and "glottalization", British linguist Sampson has coined a new term "Featural (Writing) System" for the Korean alphabet. According to Sampson, therefore, the Korean alphabet is to be classified as forming a unique type of alphabetic system, quite distinct from the ordinary alphabetic systems such as Roman or Cyrillic.

4.2. Limitations of HanGeul as Phonetic Symbols

Although the Korean alphabet has been widely acclaimed by phonicians and linguists as an excellent phonetic writing system and no doubt a highly successful writing system for the Korean language, it leaves much to be desired before it can be utilized as a truly international phonetic alphabet capable of representing minute phonetic differences of human speech sounds. For instance, there is no way to represent the voiced/voiceless distinction or to distinguish labial and labio-dental articulation by Korean letters. Thus the following pairs of English words are indistinguishable in the HanGeul writing:

- /p/ fine/pine → 파인[pain]
- /t/ vote/boat → 보우트[bout]
- /d/ they/day → 데이[dei]
- /r/ lice/rice → 라이스[rais]

On the other hand, the Korean alphabet has a definite advantage over IPA in that it has basic letters available for representing unaspirated, slightly aspirated, and strongly aspirated consonants separately in Korean such as /\text{m}/ (voiceless unaspirated), /\text{n}/ (voiceless slightly aspirated) and /\text{x}/ (voiceless strongly aspirated). The IPA would need to utilize additional diacritical marks to represent the relevant distinctions:

- e.g. Korean: \(<\text{m}>'\) - \(<\text{m}>'\) - \(<\text{x}>'\)
- IPA: \(<\text{p}>'\) - \(<\text{p}>'\) - \(<\text{p}>'\).

It is necessary, therefore, to implement the current Korean alphabet to make it a really versatile international phonetic alphabet.

5. Principles of the International Korean Phonetic Alphabet

The International Korean Phonetic Alphabet, first published in 1917 was devised by the present writer by applying the organic principle much more extensively than King Sejong had done. Accordingly, the IKPA symbols are just as simple and easy to learn and memorize as the Korean alphabet, but at the same time they are much more consistent and logical than the IPA symbols which are unsystematic and arbitrary except in one respect, i.e., retroflex symbols, which are consistently marked by a hook attached to the relevant letters. The organic principles applied in devising the International Korean Phonetic Alphabet can be summarized as follows.

1) Mobilizing All HanGeul Letters. All HanGeul letters are mobilized in the making of the Korean Phonetic Alphabet except those representing diphthongs such as \(<\text{a}>[\text{a}]\), \(<\text{a}>[\text{a}]\), \(<\text{a}>[\text{a}]\). In addition, the 4 extinct letters of Hurmin Jeongeum of the 15th century
have all been revived and given definite phonetic values. For instance, the triangle is introduced as a symbol representing voiced alveolar fricative [z].

2) Devising Indispensable Symbols. New symbols have been devised by adding one or more of the following marks to the relevant basic letters, consistent with the principles of Hunmin Jeongeum. Some new symbols are derived by deleting a stroke from relevant Hangul letters.

1) Adding the voice bar [ / ], [ - ] or [ \ ] to derive voiced symbols from voiceless ones.
   e.g. <   > + [ \ ] → <   > [ a ]
2) Adding a < - >-shaped hook symbolizing the front of the tongue bunching up to derive palatalized symbols from non-palatalized ones.
   e.g. <   > + [ L ] → <   > [ j ]
3) Adding < j>-shaped hook symbolizing the tongue tip curling upward to derive retroflex sounds from non-retroflex ones.
   e.g. <   > + [ j ] = <   > [ t ]
4) Adding a small circle under or over a letter to derive fricative symbols from homorganic plosives.
   e.g. <   > + [ s ] = <   > [ s ]
5) Adding a small hook to derive uvular symbols from velars.
   e.g. <   > + [ r ] = <   > [ q ]
6) Deleting a stroke from plosives to derive homorganic fricative symbols.
   e.g. <   > - [ - ] = <   > [ f ]
7) Adding a stroke to trill sounds to derive homorganic lateral sounds.
   e.g. <   > + [ v ] = <   > [ v ]
8) Deleting and adding strokes to derive fricative symbols from plosives.
   e.g. <   > - [ - ] + [ \ ] → <   > [ v ]
9) Vowel symbols are also derived from the basic letters by applying the same principle as shown in the derivation of consonant symbols(cf. IKPA chart)
10) Two semi-vowel symbols are derived from the relevant vowels by modifying their shape(cf. IKPA chart).

3) Representing Homorganic Sounds Systematically. The symbols of homorganic sounds that are articulated at the same place of articulation are designed in such a way that they are all marked by a basic articulatory phonetic feature. This will no doubt have mnemonic value for learners and readers.
   e.g.   [k] →  [k'] →   [k']

4) Devising Diacritical Marks to Enrich the Phonetic Representation A number of diacritical marks have been devised to represent various shades of speech sounds such as voicing, palatalization, retroflexion, etc.

6. Conclusion and Illustration

6.1 The advantages of the International Korean Phonetic Alphabet can be summarized as follows:

1) IKPA represents the shape and/or articulating movement of the organs of speech, i.e., organic phonetic alphabet.
2) IKPA represents the place and manner of articulation in a consistent and systematic manner.
3) Homorganic symbols share the common elements in IKPA.
4) IKPA is easy to learn, teach and memorize.
5) IKPA may serve as a universal phonetic alphabet for all human beings.

In sum, IKPA is more than a mere phonetic alphabet consisting of arbitrary symbols. The IKPA symbols visualize or mirror the actual speech organs or their action and thus tell us exactly what sort of an articulatory action is involved in producing sounds. IKPA is in reality the articulatory phonetic theory itself, which is self-explanatory for specialists and laymen alike. It is in this sense that the IKPA is rightly called "A Universal Visible Speech".
6.2 Illustration of IKPA

1) Korean: 꼰က’יר기 가다 [kok’irige kanda]
4) Japanese: ‘ vara トマ to ‘ vara トマ “Narita to Haneda’
5) French: Bon jour, ma cherie”
6) Thai: สวัสดีครับ [sawadi k(r)ap]
7) Lahu: 괴 쿤라 뱽 뱽 뱽 뱽 괴 쿤라 뱽 뱽 뱽 뱽 괴 쿤라 뱽 뱽 뱽 뱽 [gka kaulib qaive jo]

Reference

## 국제한글음성문자

### International Korean Phonetic Alphabet

#### 당소리/Consonants

<table>
<thead>
<tr>
<th>Place Manner</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental and Alveolar</th>
<th>Retroflex</th>
<th>Post-Alveolar</th>
<th>Alveolo-Palatal</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>ㅂ ㅃ ㅍ</td>
<td>ㅌ ㄸ ㄵ</td>
<td>ㄵ ㄸ ㄟˇ</td>
<td>ㄹ ㄧㄵ</td>
<td>ㄲ ㄳ ㄴ</td>
<td>ㄵ ㄶ ㄧㄵ</td>
<td>ㄹ ㄧㄵ</td>
<td>ㄹ ㄧㄵ</td>
<td>ㄹ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
</tr>
<tr>
<td>Affricate</td>
<td>ㅌ ㄸ ㄧㄵ</td>
<td>ㄹ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
</tr>
<tr>
<td>Fricative</td>
<td>ㅂ ㅃ ㅍ</td>
<td>ㅌ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
</tr>
<tr>
<td>Nasal</td>
<td>ㅁ ㄭ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
</tr>
<tr>
<td>Lateral</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
</tr>
<tr>
<td>Rolled</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
</tr>
<tr>
<td>Flapped</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
<td>ㄴ</td>
</tr>
<tr>
<td>Approximant</td>
<td>ㄸ ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
<td>ㄧㄵ ㄧㄵ</td>
</tr>
</tbody>
</table>

#### 홍소리/Vowels

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Rounded</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>닫힌(Close)</td>
<td>ㄑ ㄑ ㄑ ㄑ</td>
<td>ㄧ ㄧ ㄧ ㄧ</td>
<td>ㄧ ㄧ ㄧ ㄧ</td>
<td>ㄧ ㄧ ㄧ ㄧ</td>
</tr>
<tr>
<td>반닫힌(Half-close)</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
</tr>
<tr>
<td>반열lanmış(Half-open)</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
</tr>
<tr>
<td>열린(Open)</td>
<td>ㄕ ㄕ ㄕ ㄕ</td>
<td>ㄒ ㄒ ㄒ ㄒ</td>
<td>ㄒ ㄒ ㄒ ㄒ</td>
<td>ㄒ ㄒ ㄒ ㄒ</td>
</tr>
</tbody>
</table>

Devised by H.B.Lee (1971~)
Plenary Session 4
09:00 - 10:00, Friday, October 8, 2004
From X-ray or MRU Data to Sounds through Articulatory Synthesis: towards an Integrated View of the Speech Communication Process
Jacqueline Vaissière, Laboratoire de Phonétique et de Phonologie

Abstracts
This tutorial presents an integrated method to simulate the transfer from X-ray (or MRI) data to acoustics and finally to sounds. It illustrates the necessity of an articulatory model (hereby Maeda’s model) so as to:
1. Construct realistic stimuli (sounds that human beings could really produce) for psychoacoustic experiments.
2. "hear" what kind of sounds the vocal tract of a man or a woman, of a new-born or a monkey could produce and inversely, what vocal shapes could produce a sound with given acoustic characteristics.
3. Study the correlation between the observed subtle articulatory and acoustic differences and the choices of preferred prototypes in the realisation and perception of the same API symbol by native speakers of different languages.
4. Modelise vowels and consonants in context, and differentiate between transitional gestures which are necessary in a co-articulation process, but not essential in order to differentiate phonemes.
5. Simulate the acoustic and perceptual consequences of the articulatory deformation realized by the singers (e.g. singing formant), or in case of pathological voices. Emphasis is put on the work done in our laboratory, and more generally by different teams in France (Grenoble, Aix-en-Provence, Strasbourg and Nancy).

Biographical Sketch
After a thesis on speech synthesis about the introduction of prosodic parameters into diphone speech synthesis for French, the latter being prepared in 1970 at the IBM Research Center, La Gaude (France), and the Centre d'Etudes pour la Traduction Automatique, Grenoble, Jacqueline Vaissière joined, as a visiting scientist, the Speech Communication Group at MIT, directed by Ken Stevens. After her journey in the US, in 1975, she became affiliated for 15 years with le Centre National d'Etudes des Telecommunications, Lannion (France), where she worked on automatic speech recognition and automatic directory services. She was elected Professor of Phonetics in 1990 at the University - Sorbonne Nouvelle in Paris. She is currently director of the Laboratoire de phonétique et de Phonologie, associated with the Centre National de la Recherche Scientifique (CNRS). She is also director of the Ecole Doctorale, « Langage et Langues » in the same university. She has worked as a consultant at the Compagnie Générale d'Electricité, in 1973, at Bell Laboratories (Murray Hill, New Jersey, USA) in 1983, and spend 9 months as an invited scientist at MIT, Speech Communication Group, USA in 1985, and at the Advanced Telecommunication Research, Kyoto, Japan, both summers of 1992 and 1993.