The Ohio State University

Working Papers in Linguistics No. 45

# Papers from the Linguistics Laboratory

Edited by

Stefanie Jannedy

The Ohio State University

Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210-1298 USA
lingadm@ling.ohio-state.edu

February 1995

## Information Concerning the OSUWPL

The Working Papers in Linguistics is an occasional publication of the Department of Linguistics of the Ohio State University and usually contains articles written by students and faculty of the department. There are generally one to three issues per year. Information about available issues appears below. Numbers 1, 5, 10 and 23 are out of print and no longer available. The table of contents can also be viewed on our WWW (World Wide Web) server:
:

http://ling.ohio-state.edu/Department/Osuwpl.html

There are two ways to subscribe to WPL. The first is on a regular basis: the subscriber is automatically sent and billed for each issue as it appears. The second is on an issue-by-issue basis: the subscriber is notified in advance of the contents of each issue, and returns an order blank if that particular issue is desired.

Volume

18 $5.00. 183 pp. (June 1975): Papers by Michael Geis, Sheila Geoghegan, Jeanette Gundel, G.K. Pullum, and Arnold Zwicky.

19 $5.00. 214 pp. (September 1975), edited by Robert K. Herbert: *Patterns of Language, Culture and Society: Sub-Saharan Africa* contains eighteen papers presented at the Symposium on African Language, Culture, and Society, held at The Ohio State University on April 11, 1975.

20 $5.00. 298 pp. (September 1975), edited by Robert K. Herbert: *Proceedings of the Sixth Conference on African Linguistics* contains twenty-seven papers presented at the Sixth Conference on African Linguistics, held at The Ohio State University on April 12-13, 1975.

21 $5.00. 252 pp. (May 1976), edited by Arnold M. Zwicky: *Papers on Nonphonology*. Papers by Steven Boer and William Lycan, Marian Johnson, Robert Kantor, Patricia Lee, and Jay Pollack.

22 $5.00. 151 pp. (February 1977), edited by Olga Garnia: *Papers in Psycholinguistics and Sociolinguistics*. Papers by Sara Garnes, Olga Garnica, Mary Louise Edwards, Roy Major, and John Perkins.

24 $5.00. 173 pp. (March 1980), edited by Arnold M. Zwicky: *Clitics and Ellipsis*. Papers by Robert Jeffers, Nancy Levin (OSU Ph.D. Dissertation), and Arnold Zwicky.

25 $5.00. 173 pp. (January 1981), edited by Arnold M. Zwicky: *Papers in Phonology*. Papers by Donald Churma, Roderick Goman (OSU Ph.D. Dissertation), and Lawrence Schourup.

26 $5.00. 133 pp. (May 1982), edited by Brian D. Joseph: *Grammatical Relations and Relational Grammar*. Papers by David Dowty, Catherine Jolley, Brian Joseph, John Nerbonne, and Amy Zaharlick.

27 $5.00. 164 pp. (May 1983), edited by Gregory T. Stump: *Papers in Historical Linguistics.* Papers by Donald Churma, G.M. Green, Leena Hazelkorn, Gregory Stump, and Rex Wallace.

28 $5.00. 119 pp. (May 1983), Lawrence Clifford Schourup, *Common Discourse Particles in English Conversation*, OSU Ph.D. Dissertation.

29 $5.00. 207 pp. (May 1984), edited by Arnold Zwicky and Rex Wallace: *Papers on Morphology.* Papers by Belinda Brodie, Donald Churma, Erhard Hinricks, Brian Joseph, Joel Nevis, Anne Steward, Rex Wallace, and Arnold Zwicky.

30 $5.00. 203 pp. (July 1984). John A. Nerbonne, *German Temporal Semantics: Three Dimensional Tense Logic and a GPSG Fragment*, OSU Ph.D. Dissertation.

31 $6.00. 194 pp. (July 1985), edited by Michael Geis: *Studies in Generalized Phrase Structure Grammar.* Papers by Belinda Brodie, Annette Bissantz, Erhard Hinricks, Michael Geis and Arnold Zwicky.

32 $6.00. 162 pp. (July 1986). *Interfaces.* 14 articles by Arnold M. Zwicky concerning the interfaces between various components of grammar.

33 $6.00. 159 pp. (August 1986). Joel A. Nevis, *Finnish Particle Clitics and General Clitic Theory*, OSU Ph.D. Dissertation.

34 $6.00. 164 pp. (December 1986), edited by Brian Joseph: *Studies on Language Change.* Papers by Riita Blum, Mary Clark, Richard Janda, Keith Johnson, Christopher Kupec, Brian Joseph, Gina Lee, Ann Miller, Joel Nevis, and Debra Stollenwerk.

35 $10.00. 214 pp. (May 1987), edited by Brian Joseph and Arnold M. Zwicky: *A Festschrift for Ilse Lehiste.* Papers by colleagues of Ilse Lehiste at The Ohio State University.

36 $10.00. 140 pp. (September 1987), edited by Mary Beckman and Gina Lee: *Papers from the Linguistics Laboratory 1985-1987.* Papers by Keith Johnson, Shiro Kori, Christiane Laeufer, Gina Lee, Ann Miller, and Riita Valimaa-Blum.

37 $10.00. 114 pp. (August 1989), edited by Joyce Powers, Uma Subramanian, and Arnold M. Zwicky: *Papers in Morphology and Syntax.* Papers by David Dowty, Bradley Getz, In-hee Jo, Brian Joseph, Yongkyoon No, Joyce Powers, and Arnold Zwicky.

38 $10.00. 140 pp. (July 1990), edited by Gina Lee and Wayne Cowart: *Papers from the Linguistics Laboratory.* Papers by James Beale, Wayne Cowart, Kenneth deJong, Lutfi Hussein, Sun-Ah Jun, Sook-hyang Lee, Brian McAdams, and Barbara Scholz.

39 $15.00. 366 pp. (December 1990), edited by Brian D. Joseph and Arnold M. Zwicky: *When Verbs Collide: Papers from the 1990 Ohio State Mini-Conference on Serial Verbs* contains eighteen papers presented at the conference held at The Ohio State University, May 26-27, 1990.

40 $15.00. 440 pp. (July 1992), edited by Chris Barker and David Dowty: *Proceedings of the Second Conference on Semantics and Linguistic Theory* contains twenty papers from the conference held at The Ohio State University, May 1-3, 1992.

41 $12.00. 148 pp. (November 1992), edited by Elizabeth Hume: *Papers in Phonology.* Papers by Benjamin Ao, Elizabeth Hume, Nasiombe Mutonyi, David Odden, Frederick Parkinson, and Ruth Roberts.

42 $15.00. 237 pp. (September 1993), edited by Andreas Kathol and Carl Pollard: *Papers in Syntax.* Papers by Christie Block, Mike Calcagno, Chan Chung, Qian Gao, Andreas Kathol, Ki-Suk Lee, Eun Jung Yoo, Jae-Hak Yoon, and a bibliography of published works in and on Head-Driven Phrase Structure Grammar.

43 $12.00. 130 pp. (January 1994), edited by Sook-hyang Lee and Sun-Ah Jun: *Papers from the Linguistics Laboratory.* Papers by Benjamin Ao, Islay Cowie, Monica Crabtree, Janet Fletcher, Ken de Jong, Sun-Ah Jun, Claudia Kurz, Gina Lee, Sook-hyang Lee, Ho-hsien Pan, and Eric Vatikiotis-Bateson.

44 $15.00. 223 pp. (April 1994), edited by Jennifer J. Venditti: *Papers from the Linguistics Laboratory.* Papers by Gayle M. Ayers, Mary E. Beckman, Julie E. Boland, Kim Darnell, Stefanie Jannedy, Sun-Ah Jun, Kikuo Maekawa, Mineharu Nakayama, Shu-hui Peng, and Jennifer J. Venditti.

45 $15.00. 169 pp. (February 1995), edited by Stefanie Jannedy: *Papers from the Linguistics Laboratory.* Papers by Julie E. Boland & Anne Cutler, K. Bretonnel Cohen, Rebecca Herman, Stefanie Jannedy, Keith Johnson & Mira Oh, Hyeon-Seok Kang, Jaan Ross & Ilse Lehiste, Ho-Hsien Pan, and Shu-hui Peng.

The following issues are available through either: The National Technical Information Center, The U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22151 (PB), or ERIC document Reproduction Service (ED) Center for Applied Linguistics, 161 N. Kent St., Arlington, VA 22209.

2 November 1968, 128 pp. (OSU-CISRC-TR-68-3). PB-182 596.

3 June 1969, 181 pp. (OSU-CISRC-TR-69-4). PB-185 855.

4 May 1970, 164 pp. (OSU-CISRC-TR-70-26). PB-192 163.

6 September 1970, 132 pp. (OSU-CISRC-TR-70-12). PB-194 829.

7 February 1971, 243 pp. (OSU-CISRC-TR-71-7). PB-198 278.

8 June 1971, 197 pp. (OSU-CISRC-TR-71-7). PB-202 724.

9 July 1971, 232 pp. (OSU-CISRC-TR-71-8). PB-204 002.

11 August 1971, 167 pp. ED 062 850.

12 June 1972, 88 pp. (OSU-CISRC-TR-72-6). PB-210 781.

13 December 1972, 255 pp. ED 077 268.

14 April 1973, 126 pp. ED (parts only)

15 April 1973, 221 pp. ED 082 566.

16 December 1973, 119 pp. ED (parts only)

## Information Concerning OSDL
## (Ohio State Dissertations in Linguistics)

Ohio State Linguistics Students are now making available dissertations written since 1992 by students in the linguistics department. For more information regarding available titles and abstracts as of January 1995, please visit our WWW (World Wide Web) Server at

http://ling.ohio-state.edu/Department/Dissertations.html

For more information and ordering procedures, please contact

osdl@ling.ohio-state.edu

or

OSDL (Ohio State Dissertations in Linguistics)
Department of Linguistics
Ohio State University
222 Oxley Hall
1712 Neil Avenue
Columbus, OH 43210-1289
USA

Currently available titles (February 1995):

Ao, Benjamin Xiaoping (1993) *Phonetics and Phonology of Nantong Chinese.* (Advisor: David Odden)

Chae, Hee-Rahk (1992) *Lexically Triggered Unbounded Discontinuities in English: An Indexed Phrase Structure Grammar Approach.* (Advisor: Arnold M. Zwicky)

Dai, John Xian-ling (1992) *Chinese Morphology and its Interface with the Syntax.* (Advisor: Arnold M. Zwicky)

Jun, Sun-Ah (1993) *The Phonetics and Phonology of Korean Prosody.* (Advisor:: Mary E. Beckman)

Lee, Gina M. (1993) *Comparative, Diachronic and Experimental Perspectives on the Interaction between Tone and the Vowel in Standard Cantonese.* (Advisor: Brian D. Joseph)

Welker, Katherine A. (1994) *Plans in the Common Ground: Toward a Generative Account of Conversational Implicature.* (Advisor: Craige Roberts)

## Announcement

Some Software that has been developed in the Linguistics Laboratory at OSU is now publically available over the internet. The programs include utility programs that implement various data reduction techniques for the analysis of Electro-palatographic data (Hardcastle, Gibbon & Nicolaidis, 1991), an implementation of Robust Linear Predictive Coding (Lee, 1988), a version of the Klatt formant synthesizer (Klatt & Klatt, 1990) that produces Windows TIFF format (.WAV) files, and an experiment-running program for method of adjustment vowel perception experiments (Johnson, Flemming & Wright, 1993).

To access this software, use FTP to connect to *ling.ohio-state.edu*. Log in with the username *anonymous* and enter your email address as the password. The files are located in sub-directories in the *phonetics* directory. See the *readme* file in the phonetics directory and the readme files in the various sub-directories for more information.


Hardcastle, W.J., Gibbon, F. & Nicolaidis, K. (1991) EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* **19**: 251-266.

Johnson, K., Flemming, E., & Wright, R. (1993) The hyperspace effect: Phonetic targets are hyperarticulated. *Language* **69**:505-528.

Klatt, D.H. & Klatt, L. (1990) Analysis, Synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**: 820-857.

Lee, C-H. (1988) On robust linear prediction of speech. *IEEE Trans. Acoust., Speech, and Signal Process.* **36**: 642-650.


Keith Johnson

## Foreword

This is the fifth issue of the *Papers from the Linguistics Laboratory* (see also OSUWPL No. 36, 38, 43, and 44). This volume contains several studies that take an experimental approach to various topics in languages like Estonian, Korean, Taiwanese, Kipare and Turkish, or discuss theoretical issues in psycholinguistics and acoustics. Some of the papers present work in progress or have been presented at international conferences, while others have been submitted for publication to refereed journals. I would like to thank Keith Johnson and Jennifer Venditti for their help during the compilation of this volume. The production of this volume has been supported by the National Institute on Deafness and other Communication Disorders under Grant No. 7 R29 DC01645-04 and by the Ohio State University Department of Linguistics.

Stefanie Jannedy
February 1995

## Phonetics and Psycholinguistics in the OSU Linguistics Laboratory

Gayle Ayers
Mary Beckman
Julie Boland
Kevin Cohen
Kim Darnell
Erin Diehm
Mariapaola D'Imperio
Donna Erickson
(*postdoc, Speech & Hearing Science*)
Jamie Green
Rebecca Herman
Stefanie Jannedy
Keith Johnson
Hyeon-Seok Kang
No-Ju Kim
Ilse Lehiste
Joyce McDonough
Chris McDougall
Julie McGory
(*Speech & Hearing Science*)
Jim Nemcek
Mira Oh
(*Yeo Joo Technical College, Korea*)
Seamus O'Honie
Ho-Hsien Pan
Panos Pappas
Shu-hui Peng
Roberto Perry-Carrasco
(*Spanish & Portugese*)
Robert Poletto
R. Ruth Roberts-Kohno
Jin Shunde
(*East Asian Lang. & Lit.*)
Liz Strand
Linda Thibault
Jennifer Venditti

# Ohio State University Working Papers in Linguistics No. 45

**Papers from the Linguistics Laboratory**

**Table of Contents**

*OSU Working Papers in Linguistics* 45 (1-10)

# Interaction with autonomy: Defining multiple output models in psycholinguistic theory[*]

**Julie E. Boland and Anne Cutler**
jboland@ling.ohio-state.edu
anne@mpi.nl

**Abstract:** There are currently a number of psycholinguistic models in which processing at a particular level of representation is characterized by the generation of multiple outputs, with resolution involving the use of information from higher levels of processing. Surprisingly, models with this architecture have been characterized as autonomous within the domain of word recognition and as interactive within the domain of sentence processing. We suggest that the apparent internal confusion is not, as might be assumed, due to fundamental differences between lexical and syntactic processing. Rather, we believe that the labels in each domain were chosen in order to obtain maximal contrast between a new model and the model or models that were currently dominating the field.

Models of psycholinguistic processing typically consist of a number of levels loosely corresponding to levels of linguistic analysis. Even where a model deals only with the operations of one level - e.g. word recognition or parsing - some assumptions about its relationship to the other levels will usually be spelled out. In part, this is because models virtually always take a stand on one side or the other of the Great Divide in psycholinguistic theorizing - interaction versus autonomy.

Consider models of syntactic processing. One of the defining issues is whether syntactic choices are made with the benefit of relevant semantic knowledge. For example, both sentence fragments in (1) are syntactically ambiguous between a main clause structure (... *the book.*) and a reduced relative structure (... *by the lawyer was informative.*). However, (1b) is not semantically ambiguous: it is much more plausible that the evidence is being examined than that the evidence is examining something. Thus the main clause structure ought to be blocked for (1b) if semantic information can be used to decide between syntactic alternatives. On the other hand, the main clause structure is simpler, so it might be preferred if only syntactic information could be considered. Widely cited work by

[*]This paper was written while J. E. Boland was a Visiting Fellow at the Max Planck Institute for Psycholinguistics. We thank Dennis Norris, Michael Tanenhaus, and Lee Osterhout for their insights and comments on an earlier draft of this manuscript. The ordering of the authors is strictly alphabetical. Julie Boland, 137 Townshend Hall, 1885 Neil Ave., Columbus, OH 43210. Anne Cutler, Max Planck Institute for Psycholinguistics, Post Box 310, NL-6500 AH Nijmegen.

Ferreira and Clifton (1986) suggested that syntactic decisions such as these were based solely upon structural simplicity, supporting Frazier's (1978) autonomous model of syntactic processing. According to Frazier's model, the parser always constructs the simplest structure allowed by the phrase structure rules of the grammar. This initial parse uses only the major syntactic category (noun, verb, etc.) of the input, and is later checked against detailed lexical and semantic information.

(1) a. The defendant examined ...

b. The evidence examined ...

However, recent work suggests that semantic influences can affect syntactic choices (e.g. Altmann, Garnham, & Dennis, 1992; Britt, 1994; Boland, Tanenhaus, Garnsey, & Carlson, 1994; Pearlmutter & MacDonald, 1992; Trueswell, Tanenhaus, & Garnsey, 1994). For example, Trueswell et al. found that although processing difficulty arose when sentences like (1a) were completed with a reduced relative structure, no such processing difficulty arose when sentences like (1b) were completed with a reduced relative structure. They argue that, contrary to Frazier's (1978, 1987) claims, detailed lexical information is used to constrain the syntactic alternatives, and semantic information is used to select among them. Similar arguments are put forth in Boland, Tanenhaus, Garnsey, and Carlson (1994), based on their work on *wh*-questions.

As might be expected, proponents of the lexicalist constraint-based approach have adopted a position on the question of interaction versus autonomy. They describe their approach as an interactive system, opting for an interactive architecture on the grounds that multiple constraints, some of them non-syntactic, govern the selection of the initial syntactic structure (e.g. MacDonald, Pearlmutter, & Seidenberg, in press; Tanenhaus & Trueswell, in press).

The lexicalist constraint-based view is in some ways similar to the incremental interactive theory first proposed in Crain and Steedman (1985) and further refined in Altmann and Steedman (1988). In this model, syntactic alternatives are constructed in parallel within the constraints of lexical specifications, and a single representation is selected by the semantic system, using principles of referential support, a priori plausibility, etc. Thus there is a bottom-up generation of alternatives, with selection of a single structure left for a later stage of processing.

As the label given to the model makes plain, Altmann and Steedman (1988) considered their model to be interactive, noting that their results "support the interactive hypothesis" (p. 192). However, they explicitly described it as only weakly interactive. "According to this [weak] version [of the interactive hypothesis], syntax autonomously proposes analyses, while semantics and context merely dispose among the alternatives offered." (p. 205) They contrasted their position with strongly interactive models, which generate only the most plausible structure(s), and with Frazier's (1978) autonomous model, which generates only the simplest structure.

The incremental interactive model clearly separates generation processes from selection processes. This distinction is not unique to the parsing literature, however; it is also a feature of many word recognition models. Models of visual word recognition began to adopt this approach in the 1970s. Becker (1976) proposed a "verification model" in which a rough physical analysis of the input extracts sensory features and compiles a set of candidate words having those features, and therefore compatible at least in part with the incoming stimulus; these words then are ranked in order of frequency and compared one by one against a stored sensory representation of the input; this comparison process is termed verification because it actually consists of the generation of predictions from the lexical representations of the candidate words and the verification (or otherwise) of these predictions against the input store. The "checking model" put forward by Norris (1986) likewise generates an initial candidate set on the basis of partially analyzed perceptual information. The set is continually updated as the perceptual analysis is refined, but in the meantime the candidates in the set are also checked for compatibility with the sentential or other semantic context constructed so far in the recognition process. There is in this case no intrinsic ordering within the candidate set; word frequency, contextual compatibility and perceptual information all operate in the same way, to increment individual candidate words' weightings and thus eventually to determine which candidate word first reaches a specified recognition criterion.

Models of spoken word recognition, too, may split the recognition process into separate stages. Norris' (1994) SHORTLIST model, as its name suggests, is one such; in this model the initial stage again generates multiple candidates compatible with the input, while in the second stage a process of competition (involving, again, adjustment of weightings for each candidate word) determines which of the shortlisted candidates eventually wins through to recognition. A principled difference between the visual and spoken word recognition situations is that in the former case it is reasonable to consider a rough initial analysis of the entire word as input to the candidate set generation, so that such features as word length may play a role at this stage; in the latter case the temporal dimension within which the input arrives means that the initial candidate set will be primarily determined by the initial portions of the stimulus, and word length cannot be initially apparent.

The cohort model of Marslen-Wilson and Welsh (1978), especially in its revised form (Marslen-Wilson, 1987), also allows for an initial stage in which only the perceptual input determines a subset of lexical entries. Selection among this set of activated candidates is then carried out by a later stage, which in the earlier version of the model was sensitive both to further accumulation of perceptual evidence and to contextual (syntactic and semantic) information, but in the later version operates on perceptual information alone, in parallel with a contextual integration stage.

As Marslen-Wilson (1987) points out, the concept of multiple output distinguishes such models from, for example, direct access models such as the logogen model (Morton, 1970) in which only one lexical entry will surmount a recognition threshold and be effectively accessed. Norris (1986) argued that incorporating multiple output makes word recognition models in effect more parsimonious, since multiple access is in any case required to deal with the phenomenon of lexical ambiguity. It is generally agreed that at least under certain conditions presentation of an ambiguous word will lead to momentary availability of its multiple senses (Conrad, 1974; Swinney, 1979; Tanenhaus, Leiman & Seidenberg, 1979); that is, the perceptual input alone does not suffice to identify a unique word candidate, and final selection must be made by comparison against the context (which, since the evidence suggests momentary availability of multiple meanings, must occur at a post-access stage). If the mechanism for simultaneous access of multiple candidates must exist in any case, to explain selection where the input *cannot* unambiguously determine the output of the recognition process, then architectural economy is best served by exploiting precisely that mechanism in all recognition processes, for unambiguous as well as for ambiguous words.

There is widespread agreement in the word recognition literature of recent years that Multiple Output approaches fundamentally embody autonomy of the lexical access process. Becker's (1976) model, to be sure, included a separate process of generation of a semantically appropriate candidate set of words (to be compared via the same verification process with the stored input), and hence allowed for semantic context to drive a lexical access process. But since then none of the Multiple Output models have allowed higher-level (syntactic or semantic) processes such freedom. Instead, autonomy is deemed to be preserved in such models in that the actual process of contacting a lexical entry is responsive solely to bottom-up perceptual information, and is not affected in any way by higher-level processing. Thus the checking model has "a completely bottom-up flow of information" and the "stages are completely autonomous" (Norris, 1986: 131); the revised cohort model constitutes "a fully bottom-up model where context plays no role in ... access and selection" (Marslen-Wilson, 1987: 71) because "both access and certain aspects of selection are autonomous processes, in the sense that they are driven strictly from the bottom-up" (Marslen-Wilson, 1987: 98); Shortlist is "a bottom-up autonomous model" (Norris, 1994: 231) in which all "top-down feedback ... is redundant" (Norris, 1994: 191).

Thus the psycholinguistic literature may appear, to a newcomer, to be prey to internal confusion. Both for parsing and for word recognition, Multiple Output models have been proposed which have basically the same architecture; but the position they adopt on the interaction/autonomy issue is not as similar as their structural similarities would seem to demand. In fact models in the two domains take up fundamentally incompatible positions: one is called interactive, the other autonomous. Researchers in each domain clearly agree on the criteria by which these labels are applied, and thus, within each area, there is no confusion; but we believe that the reasons for the asymmetry provide an interesting subject of scrutiny.

The description that best fits the Multiple Output architecture depends on what one considers to be the defining features of autonomy and interaction, and different definitions have been established in word recognition and parsing. In the parsing literature, use of higher-level information to resolve lower-level decisions constitutes interaction. Given this definition, Multiple Output models are clearly interactive because higher-level information is used in the selection process. On the other hand, one might not consider a process interactive unless higher-level information actually affects the way that alternatives are generated within the system, ruling out certain candidates *a priori*, irrespective of their compatibility with bottom-up information. Autonomy would imply that processing operations at a given level proceed in the same way irrespective of whatever counsel might be deducible from higher-level considerations. It is this type of autonomy that has characterized the debate within the domain of word recognition. This is also the definition that Fodor (1983) used in his argument for modularity in mental processing: "a system [is] autonomous by being encapsulated, by not having access to facts that other systems know about" (p. 73). In these terms, Multiple Output models are clearly autonomous.

Altmann and Steedman (1988) indeed pointed out that the architecture of their model "does not compromise the modularity hypothesis of Fodor (1983) in any way" (p. 192). In fact, in terms of this definition of autonomy, Multiple Output models of parsing are more autonomous than Frazier's (1978, 1987) model, which is autonomous only with regard to initial syntactic analysis. When the initial analysis is inconsistent with thematic information, syntactic reanalysis occurs within, or is guided by, a thematic processor (Rayner, Carlson, & Frazier, 1983; Ferreira & Henderson, 1991). Note that it is not enough for the thematic processor simply to send an error signal to restart the syntactic processor, because the syntactic processor would automatically construct the simplest structure once again. Models such as the incremental interactive model (Altmann & Steedman, 1988) or the concurrent model proposed by Boland (1993), which produce parallel outputs, do not have this limitation. If necessary, the syntactic processor would reproduce the parallel outputs exactly as it had the first time, and the external selection processes would make the correct selection, guided by the knowledge of the previous mistake. Thus the parallel parser generates structures completely autonomously during reanalysis as well as during initial analysis.

The fact that parsing models and word recognition models have maintained different definitions of autonomy provides only a superficial explanation for the inconsistent labeling of Multiple Output models. The question then becomes: why has the parsing literature used one definition and the word recognition literature another? One possible reason is that there exist fundamental differences between lexical and syntactic processing, which justify adopting different definitions of autonomy. Traditionally, word recognition has been viewed as a lookup process, i.e. the access of stored lexical representations. Parsing, on the other hand, has been viewed as a construction process, whereby representations are computed rather than being chosen from a store. Correspondingly, outputs of lexical processing have been assumed to coincide with the completion of the processing stage (i.e. recognition of the word), but outputs of syntactic processing have been taken to correspond to many incremental stages in the construction of a complete syntactic structure.

However, we believe that current models of both parsing and word recognition make the maintenance of such rigid distinctions no longer tenable. For instance, it is clear that processes which essentially involve simple lookup can do much of the work in parsing traditionally believed to require construction processes. There is abundant evidence that syntactic decisions make use of detailed lexical information that is accessed as part of word recognition. This research has focused primarily on verb-based information, such as subcategorization frames (e.g. McElree, 1993; Osterhout, Holcomb, & Swinney, 1994), verb control information (Boland, Tanenhaus, & Garnsey, 1990), and thematic roles (e.g. Britt, 1994; Mauner, Tanenhaus, & Carlson, in press; Stowe, 1989; Taraban & McClelland, 1988). Use of stored lexical information means that syntactic processing is more dependent upon access processes and less dependent upon construction processes than has often been assumed. MacDonald, Pearlmutter, & Seidenberg (in press) have taken the lexicalist approach to sentence processing even further, suggesting that the lexical entries of nouns, verbs, and words of other categories contain X-bar structures. The only construction that takes places in their model is the connecting of one X-bar structure to another.

On the other hand, models of word recognition - and, in particular, Multiple Output models - do not necessarily consist solely of lookup procedures. For instance, Norris' "checking model" (1986) of visual word recognition contains much more of a continuous element, in that the initial stage is continually outputting updated analyses to the checking stage. Likewise, the SHORTLIST model of spoken word recognition (Norris, 1994) provides for a continuous input from the initial generation stage to the competition/selection stage. In fact this continuous updating feature turns out to be an essential feature of SHORTLIST. In order to account for empirical data indicating that human listeners employ prelexical segmentation routines *in conjunction with* competition processes (McQueen, Norris & Cutler, 1994), the SHORTLIST model has been modified to include a prelexical segmentation procedure mimicking Cutler and Norris' (1988) Metrical Segmentation Strategy (Norris, McQueen & Cutler, in press). To achieve this, it proved essential that the updated output of the initial stage continually *replace* the previous output. Only with this replacement mechanism did the model exactly simulate the human empirical data. The continuous output feature of such models renders the notion of a simple lookup procedure, with its completion amounting to completion of the lexical stage of processing, inaccurate as a description of the word recognition process.

We do not mean to imply that there is agreement that lexical and syntactic processes are fundamentally alike; these issues remain the subject of hot debate, and the traditionally held differences may in fact have influenced the adoption of different definitions of autonomy in the two domains. But whatever the outcome of the debate, there is no longer any logical force behind the argument that lexical and syntactic processing are so different that an identical architecture motivates opposite theoretical descriptions in the two domains.

Instead, we believe that considerations outside the architecture of processing models have influenced how Multiple Output models have come to be labeled. Multiple Output models, both of word recognition and of parsing, were introduced after other models had already, in effect, defined the territory. In each case, the Multiple Output model posed a challenge to the existing model, and was correspondingly assigned an opposing label.

The dominant model in syntactic processing in the 1980s, when the syntactic models discussed above were first mooted, was undoubtedly Frazier's (1978) model. Moreover, Frazier's model was particularly known for its position in the dominant theoretical debate in psycholinguistics, in that it was declared to be strictly autonomous (although, as we pointed out above, it is at least reasonable to claim that her model is not in fact maximally autonomous). Because Frazier's model was labeled as autonomous, the opposing models - which were indeed very different in structure - came to be termed interactive. Quite reasonably, the proponents of Multiple Output models wished to promote their approach as a genuine theoretical alternative to the currently dominant approach. The most obvious way to do this was to adopt a contrasting position within the dominant theoretical debate.

In the word recognition literature, too, we believe that labels were influenced by considerations of contrast. Just as autonomy could be said to be making the running in syntactic modeling, and hence be the position with which contrast could most easily be drawn, so were there models in word recognition which were market leader in much the same way, and these models were interactive. In visual word recognition, the dominant model prior to the emergence of Multiple Output models was Morton's (1970) logogen model, in which higher-level information from the context contributed directly to the activation of lexical candidates just as bottom-up information from incoming input did. In spoken-word recognition, the logogen model was also a contender, but the first model specifically devoted to the auditory case, the cohort model of Marslen-Wilson and Welsh (1978) was likewise interactive in that syntactic and semantic context was deemed capable of controlling the availability of potential candidate words. And finally, TRACE (McClelland & Elman, 1986), the most influential model of spoken-word recognition since the mid-1980s, again embodies interactive use of higher-level information in the word recognition process. Thus adoption of an autonomous stance again allowed proponents of Multiple Output models to achieve maximal contrast with the currently dominant models.

The interaction/autonomy debate has functioned as an effective energizer for psycholinguistics in the last few decades; it may have stimulated more research than any other single issue. Placing one's contribution within this paradigm is *de rigeur*; but, as we have argued, the placement may not always be rigorously determined by architectural issues alone. Contrast with theoretical alternatives - that is, in effect, sociopolitical considerations - may play as large a role.

**References**


Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31,* 685-712.

Altmann, G. T. M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 556-566.

Boland, J. E. (1993). The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research, 22,* 135-152.

Boland, J. E., Tanenhaus, M. K., and Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29,* 413-432.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1994). *Verb argument structure in parsing and interpretation: Evidence from wh-questions.* Manuscript submitted for publication.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure: The parsing of prepositional phrases. *Journal of Memory and Language, 33,* 251-283.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition, 2*, 130-138.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.) *Natural Language Parsing.* Cambridge: Cambridge University Press.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Ferreira, F. & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30,* 725-745.

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic parsing strategies.* Ph.D. Thesis, University of Connecticut. Distributed by the Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing. A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading.* London: Erlbaum.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (in press). The lexical nature of syntactic ambiguity resolution. *Psychological Review*.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition,* 25, 71-102

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology,* 10, 29-63.

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (in press). Implicit arguments in sentence processing. *Journal of Memory & Language*.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory & Language, 32,* 536 - 571.

McQueen, J. M., Norris, D. G., and Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621-638.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models for Human Memory.* New York: Academic Press.

Norris, D. G. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93-136.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189-234.

Norris, D. G., McQueen, J. M., and Cutler, A. (in press). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 20, 786-803.

Pearlmutter, N. & MacDonald, M. C. (1992). Plausibility effects in syntactic ambiguity resolution. *Proceedings of the 14th annual meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning & Verbal Behavior, 22,* 358-374.

Stowe, L. (1989). Thematic structures in sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing.* Dordrecht: Kluwer.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-660.

Tanenhaus, M. K., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior, 18*, 427-441.

Tanenhaus, M. K. & Trueswell, J. (in press). Sentence comprehension. In J. Miller and P. Eimas (Eds.), *Handbook of Perception and Cognition: Vol. 11.* Academic Press.

Taraban, R. & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing. *Journal of Memory & Language, 27*, 597-632.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285-318.

*OSU Working Papers in Linguistics* 45 (11-35)

# Vocal tract evolution and vowel production*

**K. Bretonnel Cohen**
kcohen@ling.ohio-state.edu

**Abstract:** Evolution has left the anatomically modern human with a supralaryngeal airway which is qualitatively different from that of all other animals. The source-filter theory of speech (the dominant theory of speech production in modern phonetics) relates articulations to their acoustic outputs by means of the action of a "filter"—the supralaryngeal vocal tract—on a "source"—air coming from a vibrating larynx. Since the human "filter" is qualitatively different from that of other animals, we should expect that the acoustic outputs that a human could generate would be different from those that other animals can produce. However, no satisfactory account of the nature of these differences has yet been given. A computer model was used to calculate variations in vocal tract transfer functions in human and non-human supralaryngeal airways as the location of a vocalic constriction is varied. The results suggest that the nature of the acoustic differences between the human and non-human anatomies has to do with nonlinearities which characterize articulatory/acoustic relations for the anatomically modern human vocal tract more so than for the non-human supralaryngeal airway. These differences arise from the range of articulations which are producible in a vocal tract with a bend in one wall, as opposed to one where both walls are straight. For the high vowels /i/ and /u/, these nonlinearities lead to areas of formant stability in the human airway which are significantly larger than those in the non-human airway. For the low vowel, the results suggest that the non-human airway should not be able to produce a front/back contrast between low vowels. In contrast, modelling of the modern human vocal tract correctly predicts the possibility of a front/back contrast for low vowels, though without any increase in areas of formant stability. The extrinsic tongue musculature of the human was then compared with that of *Pan troglodytes*. Using a perturbation theory model to evaluate the acoustic effects of extrinsic tongue muscle activity, it was found that the ability to generate the vowel triangle is related to the functional potentials of the extrinsic tongue musculature. These are not acoustically significant in the non-human. In the human, they give rise to the ability to generate the extreme points of the vowel triangle.

* Mary Beckman provided ideas, support, criticism, encouragement, and more. Cathy Callaghan introduced me to the evolution of language. Ashok K. Krishnamurthy derived the formulae for the modelling study, and Tzyy-Ping Jung and Michael J. Collins programmed them in Matlab. Stan Ahalt and the Center for Cognitive Science provided financial support. John A. Negulesco of the College of Medicine reviewed an earlier version of the section on the evolution of the vowel triangle and examined many bisected heads with me. Leslie Kent provided valuable discussion at all hours of the day and night. Frederick Parkinson listened to the whole thing many times, and many others provided valuable comments at the Spring 1993 meeting of the Acoustical Society of America, the 1993 Linguistic Institute, and the 1994 meeting of the Linguistic Society of America.

# I. Introduction

## I.1 The issue

The source-filter theory of speech relates the articulations by which speech is produced to the acoustic outputs which they generate. This relationship is understandable as the effect of a "filter"—the supralaryngeal vocal tract—acting on a "source," which in the prototypical case consists of periodically disturbed air exiting from a phonating larynx.

The supralaryngeal airway of the adult human is not just quantitatively but qualitatively different from that of all other animals, including our archaic hominid ancestors. Since we have qualitatively different "filters," we should expect there to be qualitative differences in the nature of the acoustic outputs that we can generate. However, a satisfactory analysis of the nature of these differences has not previously been presented. I will show that they are related to (1) nonlinearities in the relationship between articulatory and acoustic parameters, and (2) the contrasts that can be generated by the two sorts of filters.

## I.2 The anatomy

The human vocal tract differs from that of other animals as a result of two trends in hominid evolution. These trends consist of flexion of the base of the skull and a decrease in the height of the larynx.

The basicranium, or base of the skull, is formed from the occipital, temporal, sphenoid, vomer, palatine, and maxillary bones (Jacob and Francone 1965:81). Together they form the bottom of the cranial vault, the roof of the mouth, and the superior boundary of the vocal tract. They serve as articulators (e.g. the alveolar ridge of the maxillary bone) and as the superior (in the anatomical sense of that word, meaning located towards the head) insertions of a number of the muscles of the vocal tract, including muscles of the velum, uvula, and pharynx (McMinn and Hutchings 1977:16).

All non-hominid animals have a relatively flat basicranium. In other words, the plane which is oriented along the base of the skull is relatively flat. Archaic hominids, like other animals, have a somewhat less flat, but still flat, basicranium. Over the course of the evolution of hominids from Australopithecus to Homo sapiens, basicranial flexion—the degree to which the basicranium is non-flat—increases gradually. With the appearance of Homo sapiens, it becomes bent, or flexed. For clear illustrations of the relevant structures, see the illustrations in, e.g., Lieberman (1975), (1984), and (1991). The chimpanzee is representative of the standard non-human mammalian basicranial shape. Note that the basicranial line of the human contains a sharply acute angle, while that of the chimpanzee does not.

All animals other than anatomically modern humans, then, have an unflexed skull base. All animals other than humans also have a larynx located high in the throat. In most mammals, the superior edge of the larynx is roughly parallel to the first cervical vertebra. In humans, the superior edge of the larynx is parallel to the 4th cervical vertebra. In the absence of soft tissue remains, the same sort of fossil record that is present for the development of basicranial flexion over the course of hominid evolution is not present for laryngeal descent. However, comparative zoological evidence demonstrates clearly that in all other animals extant today, laryngeal height is inversely correlated with the degree of basicranial flexion. That is, the lesser the degree of basicranial flexion, the higher is the larynx in the neck (Laitman 1984, Laitman and Reidenberg 1988). So, if our hominid ancestors did not have high larynges to go with their unflexed basicrania, then they differed from all other animals in this respect. There is no reason to think that this is the case; probably, then, they did have high larynges.

The combined effect of these changes has been to leave the modern human with a pharyngeal cavity, located at a right angle to the oral cavity, which is absent in other animals. This is illustrated schematically very nicely in Hoffman *et al.* (1989:106). Thus, the trends of basicranial flexion and laryngeal descent over the course of the evolution of the hominidae have left us with a supralaryngeal airway which is qualitatively different from that of all other animals.

### I.3 Lieberman's "abrupt discontinuity" analysis

Philip Lieberman pioneered the study of the relationship between the evolution of the human vocal tract and the phylogeny of human language. He has suggested that the acoustic significance of basicranial flexion and laryngeal descent lies in the differing articulatory capabilities of an airway consisting of a tongue located within an unbent tube as compared to an airway consisting of a tongue located within a bent tube. Lieberman (1975:115, 1984:278-280) interprets the modelling studies in Stevens (1972:57) as demonstrating that (quantal) vowel production requires abrupt discontinuities in cross-sectional area. This effect can be achieved—in the modern human vocal tract—by displacing the body of the tongue anteriorly and superiorly, as in the production of the high front vowel [i]; posteriorly, as in the production of the low back vowel [ɑ]; or superiorly and posteriorly, as in the production of the high back vowel [u]. According to Lieberman, this sort of abrupt discontinuity in cross-sectional area cannot be achieved in a straight-tube, standard non-human mammalian airway; rather, only gradual discontinuities can be achieved, with the tongue sloping gradually into and out of a constriction.

However, evidence from a variety of sources suggests that Lieberman is wrong. It may be the case that non-human airways cannot generate abrupt discontinuities in cross-sectional area. However, it is clearly the case that in the production of vocalic sounds, human speakers do not produce such discontinuities, either. Consider Figure 1, which shows (a) the sort of constriction whose lack of abrupt discontinuities in cross-sectional area, Lieberman claims, prevents the non-human vocal tract from producing the three "point" vowels, and (b) a tracing from a sagittal x-ray of a human speaker producing the vowel [u]. X-ray (e.g. Fant 1960, Perkell 1969), MRI (e.g. Moore 1992), and palatographic and ultrasound studies (e.g. Stone et al. 1992) all clearly show exactly what Lieberman posits for other animals, to the exclusion of humans: a tongue surface sloping gently into and out of a constriction. In fact, a variety of writers have commented on the gradual nature of vocalic constrictions, e.g.:

> Real constrictions, formed by the tongue in the vocal tract during natural speech, have a gradual shape.
>
> Mrayati *et al.* 1988:270

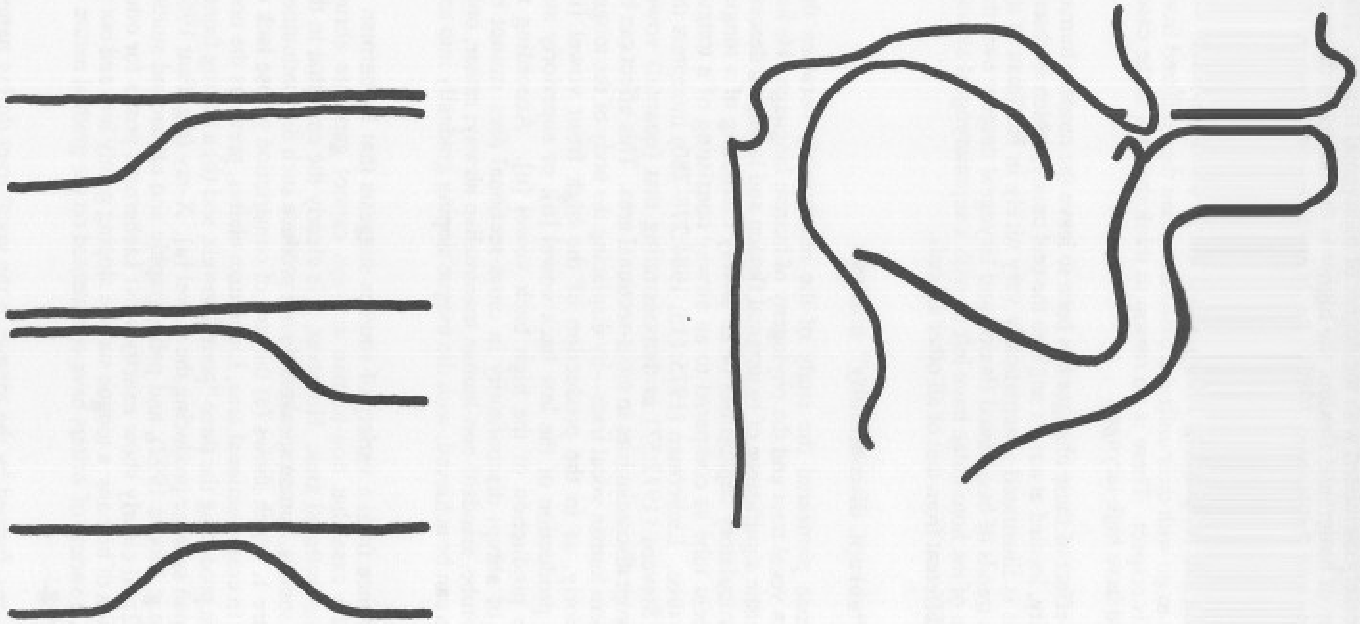


Figure 1. (a) Constrictions without abrupt discontinuities in cross-sectional area, adapted from Lieberman (1991).
(b) Tracing from a sagittal x-ray of a human speaker producing the vowel [u], adapted from Perkell (1969). The tongue, hard and soft palates, and posterior pharyngeal wall are shown.

> During the articulation of vowels, the tongue usually forms a constriction or region of minimum cross-sectional area....On either side of the constriction there is a gradual increase in area.
>
> Stevens and House 1955:485

Badin *et al.* talk at some length about this issue (1990:1297), though without providing quotable quotes. Furthermore, there is a history of modelling vocalic constrictions with gradual, nonabrupt constrictions which extends back through Fant's use of horn-shaped (i.e., tapered) tubes (1960:30 and following, 9 and following) through Stevens and House's use of a parabolic constriction (1955:486) as far as Chiba and Kajiyama (1941:82-83 and elsewhere).

So, whatever the acoustic consequences of the evolution of the human vocal tract may be, it seems clear that they cannot be related to the ability—or lack thereof—to produce abrupt discontinuities in cross-sectional area during vowel production. What, then, are they?

### I.4 An "increased nonlinearity" analysis

Gunnar Fant's Acoustic theory of speech production (1960) explains the characteristics of the speech signal in terms of the output of a filter—the supralaryngeal vocal tract—acting upon a laryngeal source. The acoustic theory of speech production makes certain predictions about the way that the transfer function (i.e., the acoustic output, expressed as a set of formant frequencies, of a given vocal tract configuration) of the vocal tract should vary as a constriction is moved from location to location. These predictions are expressed in graphs called nomograms. Ladefoged and Bladon (1982) tested these predictions by producing sustained vocoids while varying only the place of constriction, using mirrors, bite blocks, and ultrasound to keep lip aperture, area of constriction, etc. constant. They noted that the formant structures produced by actual speakers varied from those predicted by Fant's nomograms. Specifically (among other things), at very forward (i.e., close to the lips, or far from the glottis) locations for an [i]-like constriction, the second formant frequency did not fall, as Fant's nomograms predicted. Rather, the second formant frequency stayed relatively stable.

Ladefoged and Bladon hypothesized that this effect was related to the fact that within the range of locations for a high front vowel constriction, the second formant is a back-cavity resonance. They suggest that "because of the curvature of the vocal tract, the length of this cavity does not increase when the constriction moves closer to the alveolar ridge" (p. 194). Rather, past the bend in the vocal tract, the back cavity length remains constant. While front cavity length decreases, there is a concomitant increase not in the length of the back cavity, but in its diameter. Though they were referring to the curve at the alveolar ridge, we were inspired by their comment to consider the effect of the curvature of the vocal tract as a whole, specifically the different effects of changing the location of a vocalic constriction in a tube bounded by a straight wall—analogous to the non-human supralaryngeal airway—versus the effect of changing the location of a vocalic constriction in a tube bounded on one side by a bent wall—analogous to the anatomically modern human supralaryngeal airway.

A geometric relationship such as that described by Ladefoged and Bladon can exist for a tube with a right-angle bend in it. However, it cannot exist for an unbent tube. Rather, in an unbent tube the only possible relationship between the lengths of the front and back cavities is that of a trade-off: as the back cavity length increases, the front cavity length decreases. And, in neither cavity is the diameter affected by variations in the length of the other.

Ladefoged and Bladon noted that "it is difficult to relate acoustic behaviors such as... formant discontinuities to the articulatory states which produced them, namely moving the tongue progressively in small steps along the upper surface of the vocal tract" (p. 192). This is certainly true, if our expectation is that articulatory/acoustic relationships should be linear. However, that is not necessarily the case. A non-linear relationship such as this is just the sort that would be predicted by Stevens' Quantal Theory of speech.

One of the hypotheses proposed here is that one of the acoustic consequences of the anatomical changes which occurred in the evolution of the modern human vocal tract can be described as a trend toward increasing the nonlinearity of the relationship between articulations and the associated acoustic output in the production of vowels. Stevens's (1972, 1989) modelling studies predict the existence of nonlinearities in articulatory/acoustic relations. Specifically, he claims that certain vowels are articulated in locations where there is a nonlinear relationship between the location of a constriction and the associated formant frequencies. For example, within the range of locations in which a high front vowel is produced, the second formant frequency is stable over a range of values for location. Stevens demonstrates these nonlinearities by means of a model of vowel production in which the location of a constriction is varied from the back to the front of the vocal tract by trading off the lengths of the front and back cavities.

This is precisely the relationship between front and back cavity lengths of which an unbent vocal tract is capable. It does in fact yield a nonlinear relationship between location of a constriction and second formant frequency over some range of values for constriction location. One wonders, then, if a bent-tube vocal tract has the same sort of nonlinear relationship between articulatory and acoustic parameters. If it turned out that there were no differences in the outputs generatable by straight-tube (non-human) and bent-tube (qualitatively different modern human) vocal tracts, that finding would be embarrassing for the source-filter theory. If it turned out that the bent-tube vocal tract is more linear than the straight-tube vocal tract in its articulatory/acoustic relationships, that would be an embarrassment for Stevens' Quantal Theory and would constitute an absolute refutation of the thesis of this paper. If, on the other hand, it turned out that the bent-tube vocal tract is associated with a less linear articulatory/acoustic relationship—e.g., if there were a larger area of stability for the second formant in the range of locations in which high front vowels are articulated—that would support the thesis that the human vocal tract has evolved in the direction of increased articulatory/acoustic nonlinearities.

## II. Method

I tested this hypothesis by modelling vowel production in straight and bent vocal tracts. Transfer functions were calculated with a transmission line analogue model. The model was tested against the comparable nomograms in Stevens

(1989), and against actual production data as part of a separate study (Beckman *et al.* (1995).

I modelled vowel production in a straight-tube, non-human vocal tract as in Stevens (1989). The location of the constriction is varied by trading off front and back cavity length: i.e., as the length of the back cavity increases, the length of the front cavity decreases.

I modelled vowel production in a bent-tube, modern human vocal tract by varying the sizes of the front and back cavities in a way which more accurately reflects the geometry of the modern human vocal tract. I used the same ratios of cavity sizes and constriction characteristics as in Stevens 1989. This sacrifices some realism, but has the advantage of allowing direct comparison with Stevens's classic model. For [i] and [u], I varied back cavity length from the glottis to the point corresponding to the bend in the vocal tract by trading back and front cavity lengths. However, past the bend, I varied the constriction location by decreasing front cavity length while keeping back cavity length constant and increasing back cavity diameter. This is expressed schematically in Figure 2 (b) below. The motivation for this this approach is pictured in Figure 2 (a). If a tongue body shape like that pictured in Figure 2 is moved forward, the effect on the back cavity is an increase not in distance along the vertical axis—i.e., back cavity length—but in distance along the horizontal axis, i.e. back cavity diameter.

The novelty of my approach lies in its strategy for varying the location of the constriction. This strategy differs from that of other models in that it explicitly recognizes the different effects on cavity area of changes of location in the anterior portion of the vocal tract as compared to changes of location in the posterior portion of the vocal tract. For a review of other models, all of which fail to take these differing effects into account, see Section VI, *Postscript I: vocal tract models*.

## II.1 Changes in overall length

My strategy for modelling changes in place of articulation in the modern human vocal tract results in a net decrease in the overall length of the vocal tract. This occurs as the constriction is advanced from the mid-point to the front of the vocal tract. This might make one wonder if the effect of second formant frequency stability is an artifact of the decrease in overall length. This can be shown not to be the case. Reduction of overall length with the same ratio of front and back tube lengths results in a pattern of increase in all frequencies. Our findings show an increase in the third formant frequency only, with the second formant remaining stable and the first formant essentially unaffected. This point will be repeated below after presentation of the modelling results.

The reduction in length in our model is in the amount of 3 centimeters. There is justification from physiological studies for some reduction in length in moving from a back to a front vowel articulation, if not in the same magnitude. Adjusting the modelling strategy so that the reduction in length more closely approximates a physiologically reasonable amount results in a pattern of second formant frequency behavior which is substantially similar to those observed in my model of the modern human vocal tract, the differences being in the range (in articulatory and acoustic space) of the stable region. Thus, the patterns of formant behavior generated by the model cannot be said to be due to inappropriate reduction of the vocal tract length. Furthermore, my model has the advantage of generating the actual patterns of formant behavior observed in human speakers (compare my

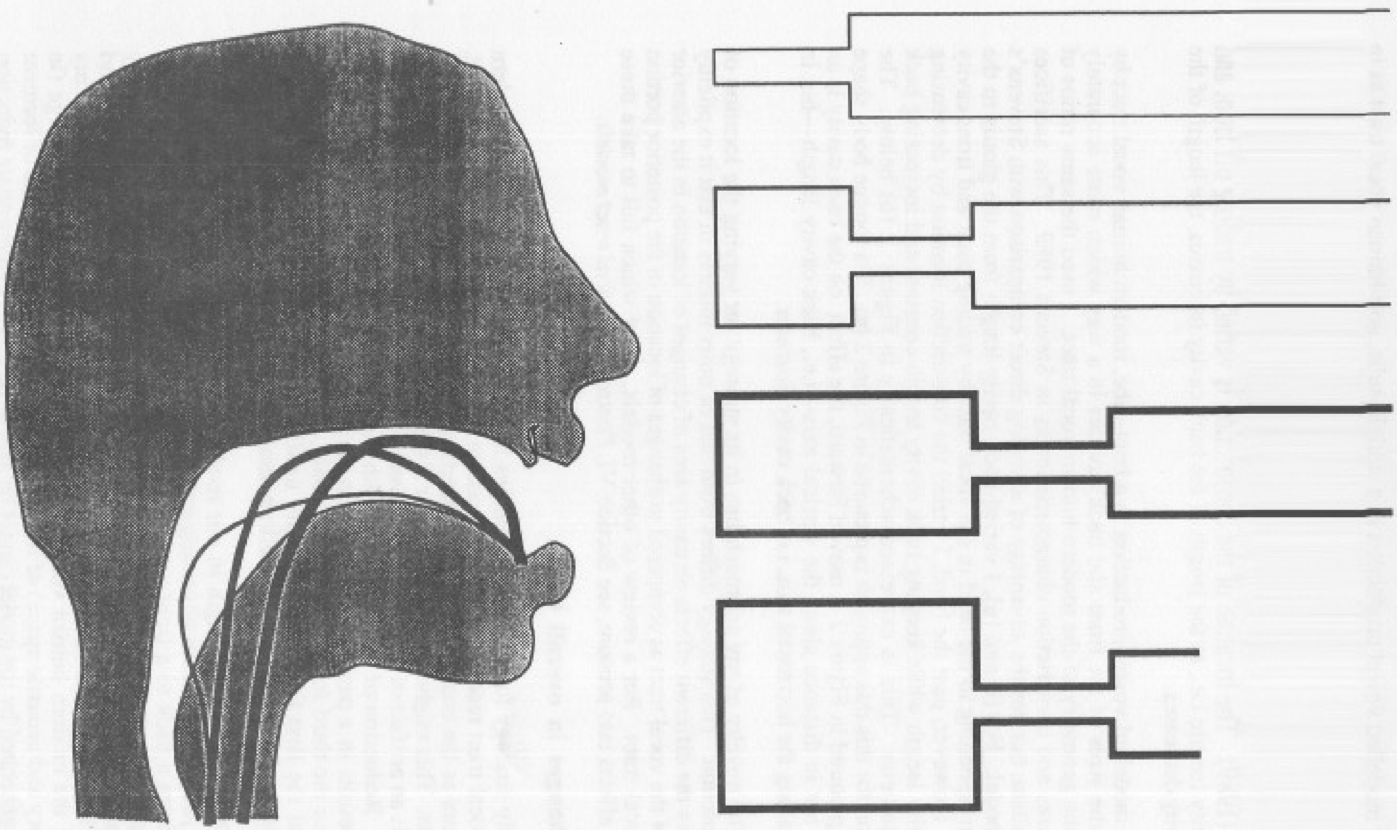


Figure 2. (a) Range of tongue shapes from back to front constriction.
(b) Schematic representation of variation in tube lengths and areas.

findings with, e.g., Ladefoged and Bladon) with fewer control parameters than an isolongitudinous model (one which maintains overall length) requires; it does a better job of generating them than do the isolongitudinous models, in the case of the third formant frequency. Kent *et al.* point out that "regardless of the individual approach taken, the basic goal in [modelling studies of articulatory/acoustic relations] has been to reduce the number of degrees of freedom" (1991:268). In that respect, the model I have used for the modern human vocal tract is superior to an isolongitudinous model.

## III. Results

Figure 3 shows the effects of varying the location of an [i]-like constriction within a modern human vocal tract (solid line) versus a non-human supralaryngeal airway (broken line). The plot for the straight-tube, non-human configuration duplicates the relevant nomogram in Stevens (1989:12). The results for the bent-tube, modern human configuration differ in that this configuration yields a much larger area of stability for the second formant within the range of locations in which a high front vowel is articulated. With the non-human configuration, the second formant is even nominally stable only from perhaps 7.5 cm to 8.5 cm from the glottis. In contrast, in the modern human model, the second formant frequency varies hardly at all from 7.5 cm to 11 cm from the glottis.

For [u], I varied chamber lengths and widths as for [i]. A short tube was added to the anterior end for the lip-rounding component of [u], and the cavities anterior and posterior to the constriction were of equal width for locations in the posterior portion of the vocal tract.

The results for an [u]-like constriction are shown in Figure 4. Note that in the modern human model, over a range of values from around 8 cm to 10 cm from the glottis, the third formant remains high and distant from the second formant. In the non-human model, a backness percept can be expected only over the range of values for constriction location from about 4 cm to 8 cm from the glottis, where the second formant-third formant distance is large due to the second formant being low. At least half of this range is too far back to be at all realistic, and over most of it the first formant frequency is rather high for a high vowel. (The frequency of the first formant is inversely correlated with vowel height.) The modern human vocal tract allows for a backness percept over a wider range of values for constriction location, including more realistic values for location centered around 8 cm from the glottis and a range of locations where the first formant frequency is lower (as is appropriate for a high vowel). A large second formant-third formant distance is maintained not by keeping the second formant low but by keeping the third formant high even as the second formant begins to rise.

Modelling the production of [a] is an interesting challenge, and highlights the difficulty in describing low vowels within a constriction-based model of vowel production. [a]-like vowel configurations differ from non-low vowel configurations in that non-low vowels have a constriction and two cavities—one anterior and one posterior to the constriction. Constriction location is varied by changing the relative sizes of the two cavities; the characteristics of the constriction itself, i.e. its length and diameter, remain constant. In contrast, for a low vowel, there are not two cavities whose areas can be manipulated independently of the constriction. Rather, for a low back vowel, the "constriction" consists of the small

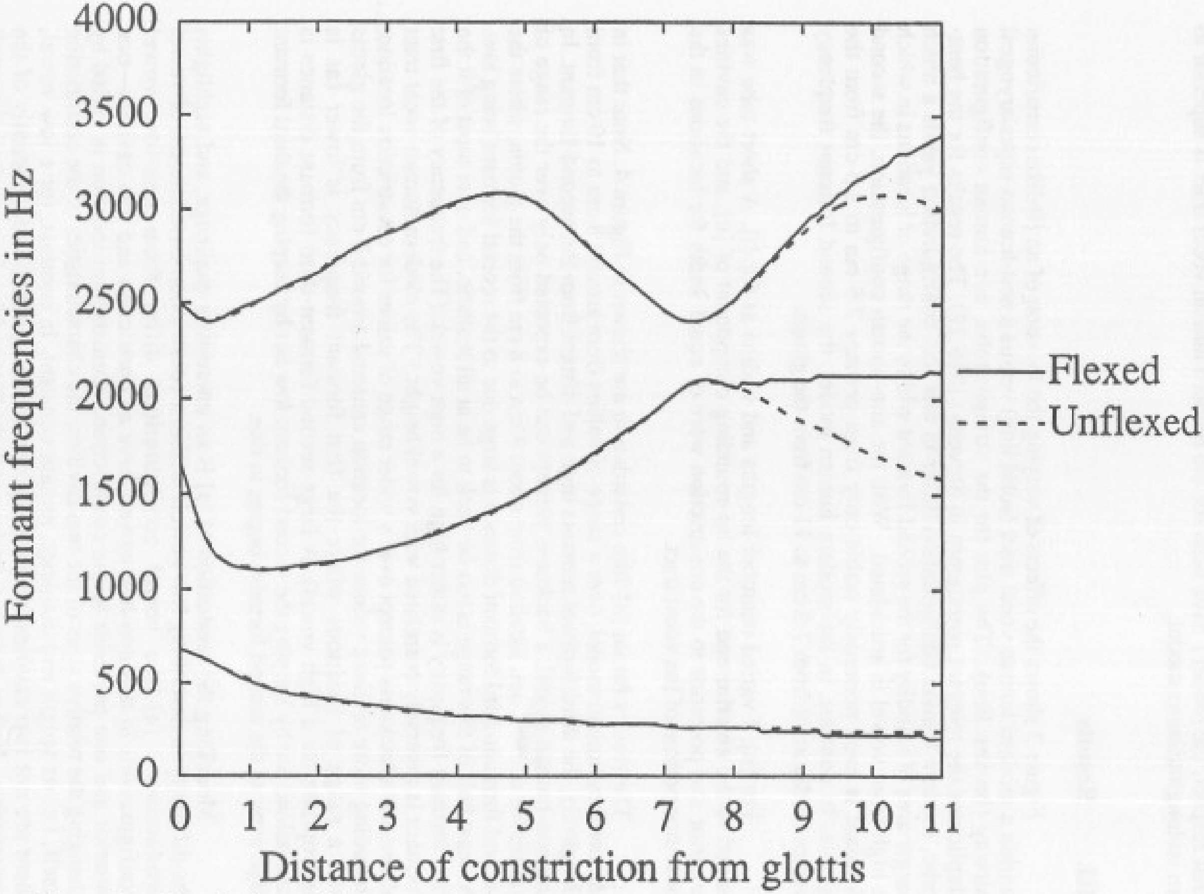



Figure 3. Results for an [i]-like constriction.

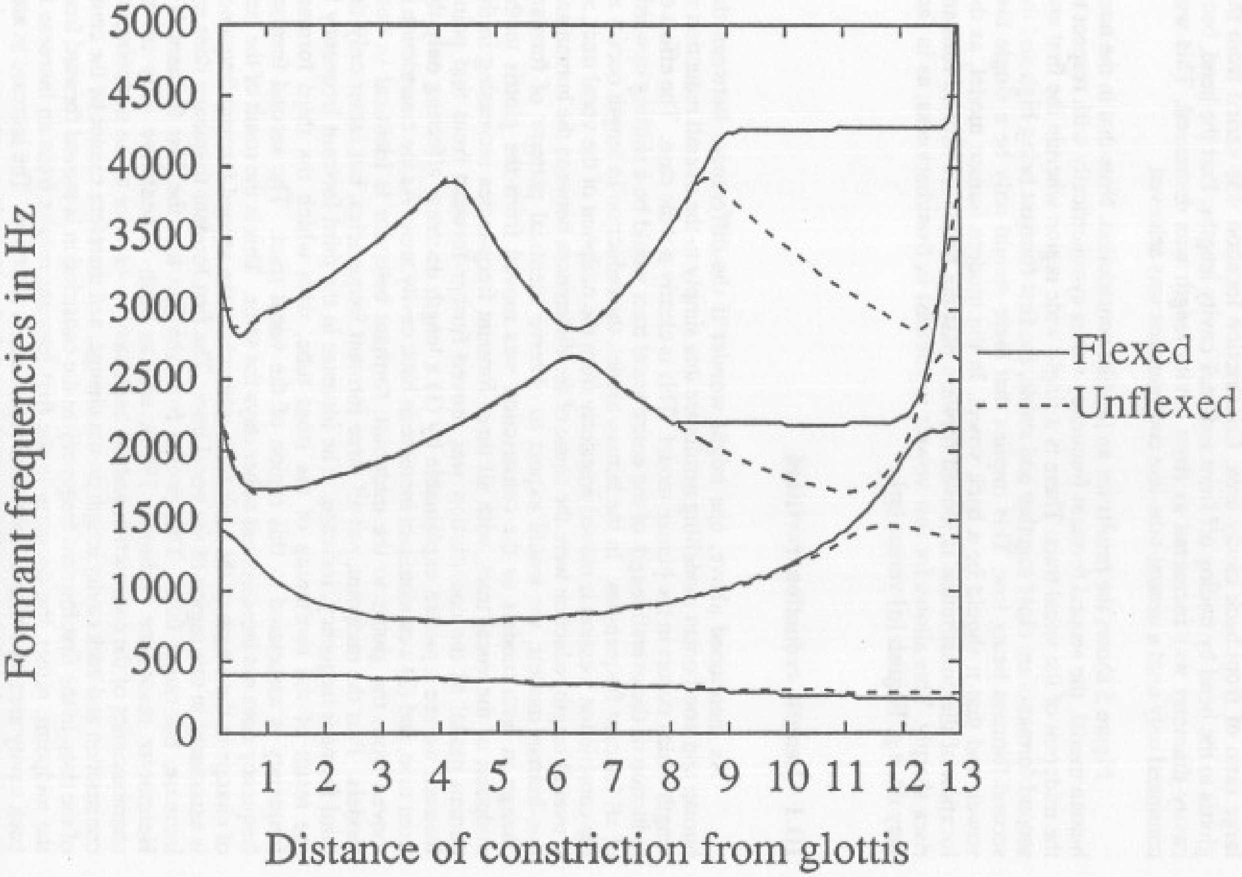



Figure 4. Results for an [u]-like constriction.

back cavity cross-sectional area. For a low front vowel, the vocal tract configuration approaches that of a neutral, i.e. completely non-constricted, tube.

Production of an [a]-like vowel was modelled with two cavities, with a large ratio of front:back cavity area. Constriction location was varied from the glottis to the bend by trading off front and back cavity lengths. Past the bend, back cavity diameter was increased as front cavity length was decreased. This was continued only until a neutral-tube-like configuration was achieved.

Figure 5 shows the results for an [a]-like constriction. Note that in the non-human model, the second formant frequency varies symmetrically with respect to the midpoint of the vocal tract. There is a single wide region wherein the first and second formants are close together and stable, the first formant being high and the second formant being low. This implies that there should only be a single low vowel, and that it should be a back vowel. In the modern human model, as the location of the constriction is moved forward past the bend, the second formant rises sharply. This allows for low vowels to contrast in front/backness, as in fact they do, e.g. English [ɑ] versus [æ].

### III.1 Length reduction revisited

As mentioned above, one might wonder if the differences between the human and non-human modelling results are due simply to the overall reduction in length which occurs in the human model. This is clearly not the case. The effect of reduction of the overall length of the entire vocal tract would be a shifting upwards of all formant frequencies. In the human model, the reduction in length occurs as the constriction location is moved anteriorly from the midpoint of the vocal tract, so if overall length reduction were the cause of the differences between the human and non-human models, we would expect to observe identical patterns of formant change in both models as the constriction was moved from the glottis to the midpoint of the vocal tract, with all three formant frequencies increasing in the human model as the constriction was moved further forward from that point. Instead, we see a pattern explainable by (1) a length decrease affecting only the front tube, and (2) a concomitant increase in back cavity area. As the constriction is moved from the glottis to the midpoint, formant behavior is identical in both models. Past the midpoint, not all three formant frequencies, but rather only the third formant frequency, increase. The increase in the third formant frequency is the result of the shortening of the front tube, with which the third formant frequency is associated in this region of the vocal tract. The second formant frequency does not increase, but rather stays the same. This is the result of the lack of change in the length of the back tube, with which the second formant frequency is associated in this region of the vocal tract. The first formant frequency does not increase, but rather falls. This results from the fact that the first formant is a Helmholtz resonance (Johnson 1994), and as such is sensitive only to the characteristics of the constriction and the back cavity. Anterior to the midpoint, the constriction and back cavity length do not change, and therefore cannot be the cause of the drop in the first formant frequency as the constriction is moved forward from the midpoint: rather, the decrease in the first formant results from an increase in back cavity area, to which a Helmholtz resonance is sensitive. The increase in area in the human model is equivalent to the increase in length in the non-human model, so the behavior of the first formant frequency is the same in both cases.

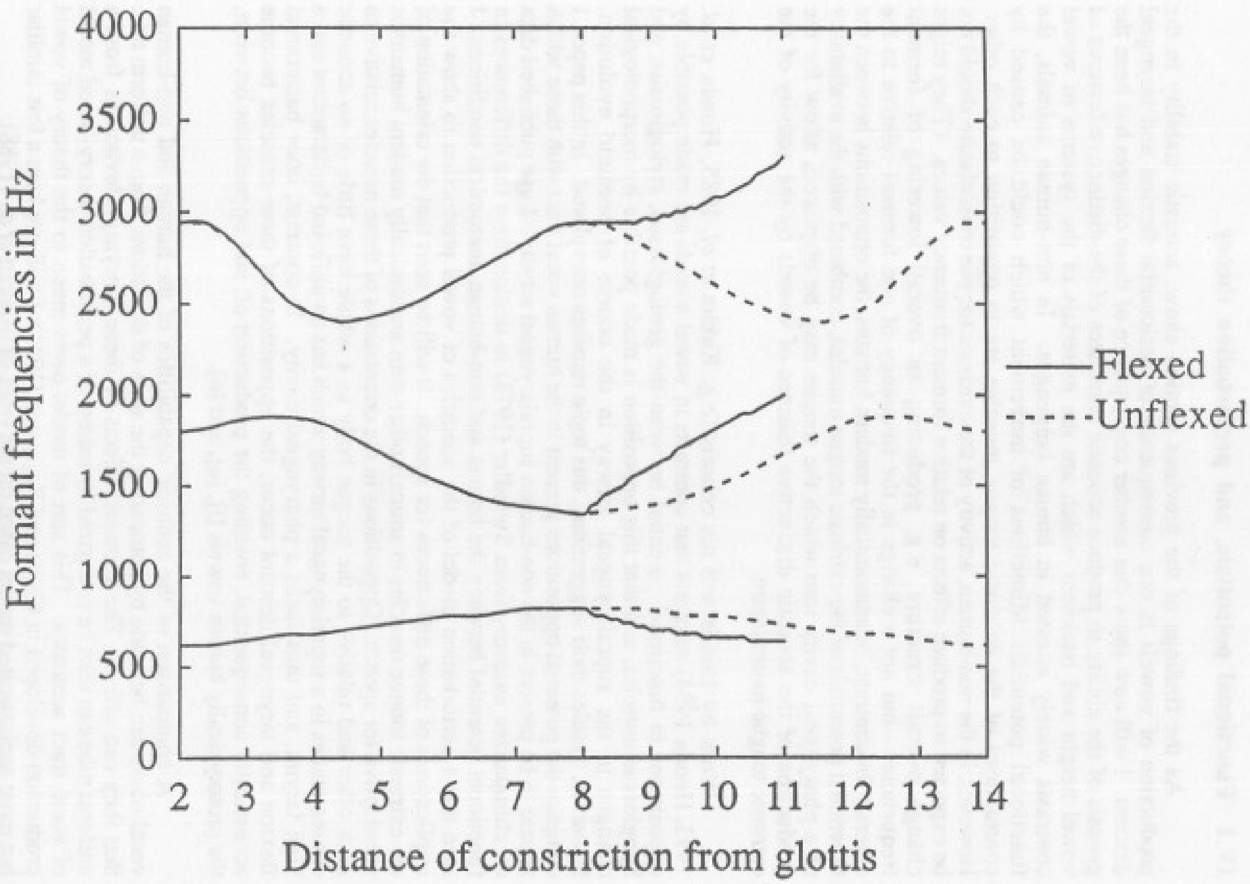



Figure 5. Results for an [a]-like constriction.

## IV. The evolution of the vowel triangle

### IV.1 Functional potentials, and perturbation theory

As the findings of the previous section show, acoustic stability in the production of vowels is one consequence of basicranial flexion and laryngeal descent. I will now show that another consequence of these changes has been the genesis of the ability to produce acoustic correlates of the distinctive features of vowel height and backness which are the materials of the systems of vowel contrasts widely attested in human languages. In non-human animals, the functional potentials (directions of movement which could be caused by contraction) of the extrinsic tongue muscles are in opposition to each other. However, in the non-human, activity of the extrinsic tongue musculature should not be expected to produce effects on relative formant frequency values. (They might change overall structure, e.g. producing an overall lowering of formant frequencies—but not a change in the frequency of one formant relative to the others.) In contrast, in anatomically modern humans, the oppositions between the functional potentials of the extrinsic tongue muscles, combined with the availability of a pharyngeal cavity into which the tongue may be displaced, allow for the production of the acoustic distinctive features of vowels by the activity of the extrinsic tongue musculature.

Work by Honda and his cohorts (e.g. Kakita *et al.* 1985, Honda *et al.* 1993, Honda 1994) suggests that contrasts in vowel sounds are made possible by oppositions in functional potential between the genioglossus, styloglossus, and hyoglossus muscles, and that this opposition is made possible by morphological changes in the supralaryngeal airway in the course of hominid evolution. However, besides their suggestions, this topic remains unexplored. In this paper, I compare the potential oppositions present in the human vocal tract with those which appear to be present in the non-human supralaryngeal airway. I use published data on chimpanzee anatomy from Swindler (1973) to demonstrate the differences in functional potential between the human and non-human anatomical conditions. I then use a perturbation model of the acoustics of vowel production to show the implications of these differences for speech. It will be seen that the orientation of the extrinsic tongue muscles in animals other than anatomically modern humans is preadaptive for speech. Oppositions in the orientations of these muscles relative to each other and relative to the tongue body as a whole have little or no acoustic consequences in a supralaryngeal airway which has an unflexed basicranium and a high larynx, and thus lacks a pharyngeal cavity. In contrast, once basicranial flexion and laryngeal descent occur, the oppositions of these muscles become acoustically consequential, enabling the production of, and oppositions between, the prototypically human vowels [i], [u], and [ɑ].

A comparison of the acoustic capabilities of the human and non-human vocal tracts can be made by comparing the sorts of deviations from a uniform tube that they can effect. The acoustic affects of these (varying) deviations from a uniform tube can then be compared by means of a perturbation-theory-based model of vocal tract acoustics. This sort of model owes much to the theory of vowel production developed in Chiba and Kajiyama (1941); I will use here a less familiar but more sophisticated model based on the work of Mrayati *et al.* (1988).

Mrayati *et al.*'s perturbation theory model is based on consideration of the effect of small, localized changes of cross-sectional area in the human vocal tract. Many other theoretical models use a transmission line analogue to model the vocal tract as a series of two tubes, with the length of each tube determining particular individual formant frequencies. In contrast, Mrayati *et al.*'s model works by consideration of the effects (on the resonance of the vocal tract as a whole) of changes in cross-sectional area on the potential and kinetic energy of air flowing through localized sections of the vocal tract. Following the work of Fant and Pauli, Mrayati *et al.* use these parameters to calculate the direction and amount of change in a formant's frequency from the area and the total, kinetic, and potential energies of a region (or subsection) of the vocal tract.[1] This approach to the acoustics of speech nicely models the formant transitions in CV (consonant-vowel) sequences, is fruitful in the consideration of the classic acoustic-articulatory inversion problem, and is a nice addition to the work of Stevens, in which the effect of changes in area, as opposed to changes in location, of a constriction receives rather short shrift. It also provides a nice account of the compensatory effects of changes in area at the lips and glottis. (I have some problems with the required orthogonality of changes in area elsewhere in the vocal tract. Such changes are clearly shown by Mrayati *et al.* to be acoustically orthogonal, but whether they can be articulatorily orthogonal in areas of the vocal tract other than its ends is not so true, I don't think.)

One finding of their work is that under the conditions which characterize (central) vowel production, i.e. relatively large "constriction" area and with the presence of acoustic coupling between the ante- and post-constriction tubes, there exist four regions of the vocal tract which have unique effects on formant structure, such that a constriction in one of these regions will produce a characteristic effect on each of the first two resonant frequencies of the vocal tract (i.e., formants). (The number of these "distinctive regions," as Mrayati *et al.* call them, is actually related to the number of formants being considered, so that for three formants, there are eight regions, for four formants, there are fourteen, etc.) These regions are illustrated schematically in Figure 6 (adapted from their Figure 2) with respect to a schematized tube. (Note that Figure 6 shows three formants, and therefore eight regions are identified on it.) (Mrayati *et al.*'s nomograms show the effect of *increasing* the area of a region, rather than the effects of constricting an area, as we are accustomed to seeing. I have adapted them to show the effects of a constriction within a region by inverting the original Figure 2.) Note that there are regions which produce a low F1 and high F2 (labelled C), low F1 and low F2 (labelled A), and high F1 and low F2 (labelled C-bar). These correspond to the vocal tract regions constricted in the production of the vowels [i], [u], and [a].

## IV.2 The extrinsic tongue musculature

In both *Homo sapiens* and the chimpanzee, *Pan troglodytes,* the extrinsic tongue muscles consist of the genioglossus muscle, the styloglossus muscles, the hyoglossus muscles, and the palatoglossus muscles. (Negulesco 1993:63). All of the extrinsic muscles originate externally to the tongue, but have insertions within it. (In this they differ from the intrinsic tongue muscles, which have both their origins and their insertions within the tongue.) The palatoglossus muscle has not

---

[1] Section VII, *Postscript II: Mrayati et. al's perturbation theory* gives a detailed explanation of the model.

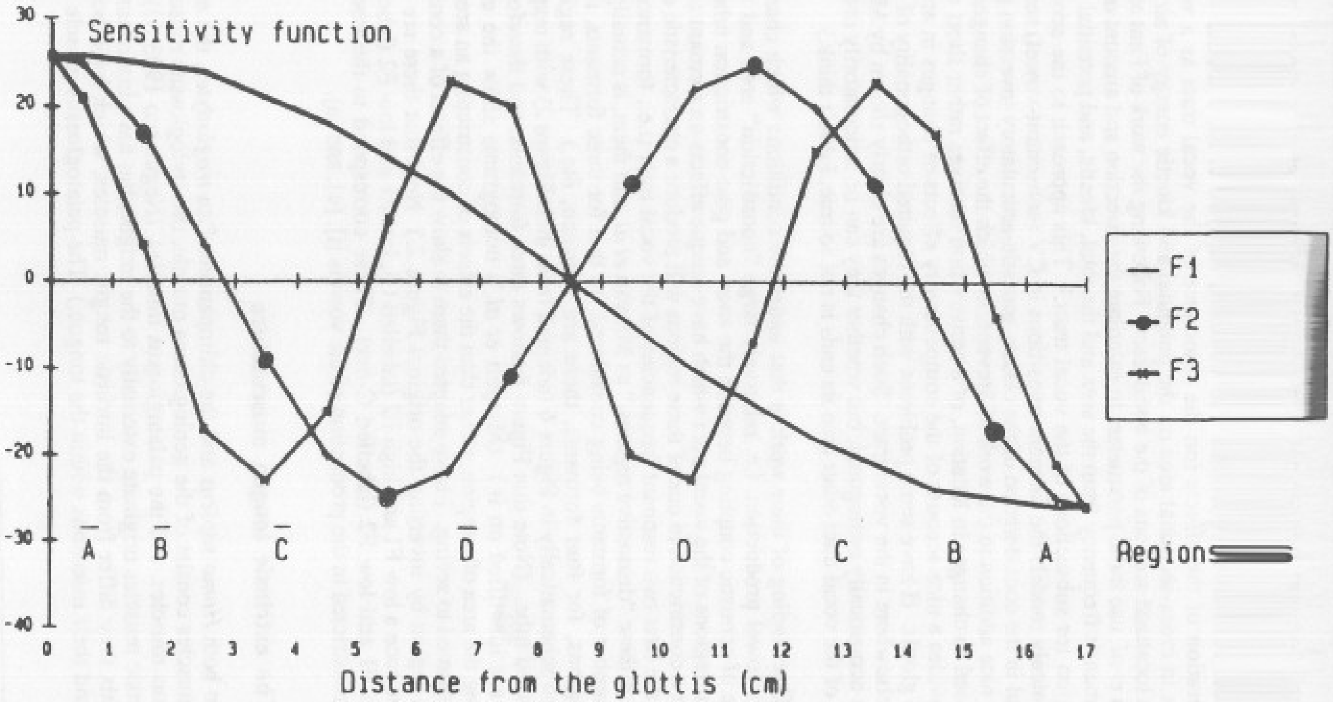



Figure 6. Distinctive regions in a perturbation model of vocal tract acoustics, adapted from Mrayati *et al.* (1988).

been shown to be of importance in the production of vowel sounds, and will not be discussed further here.

### IV.2.a Genioglossus

In both the human and Pan, the fibers of the posterior portion of the genioglossus run roughly anterior-posterior. The genioglossus originates on the upper mental spines of the mandible and appears to form the main body of the tongue, when viewed in sagittal section. (I should note here that although the functional division of the genioglossus in the human is well-established on the basis of electromyographic data, such a division has not, to my knowledge, been demonstrated for Pan.) In modern Homo, the root of the tongue forms the anterior border of the pharyngeal cavity. In contrast, in Pan, the tongue is located completely within, and in fact nearly fills, the oral cavity.

The function of the genioglossus muscle in vowel production is to produce the constriction for high front vowels. Radiographic (Fant 1960, Perkell 1969) and MRI (Moore 1992) studies have demonstrated that the production of [i] is characterized by displacement of the tongue body superiorly and anteriorly in the oral cavity, so that the tongue dorsum is positioned beneath the hard palate, posterior to the alveolar ridge. Electromyographic studies (e.g. Smith 1971) of tongue muscle activity during speech production have demonstrated that this is accomplished by contraction of the posterior portion of the genioglossus muscle.

The location of the constriction corresponds to region C in Mrayati *et al.*'s model. The acoustic effect of a constriction at this point is a lowering of the first formant frequency and an elevation of the second formant frequency. This produces the formant pattern associated with the vowel [i].

In the chimpanzee, the orientation of the fibers of the posterior genioglossus suggests that the functional potential of the genioglossus m. would probably be to displace the tongue body anteriorly and superiorly, if the mandible were lowered to allow it room to move within the oral cavity. However, it is not at all clear that the chimpanzee tongue could be displaced forward, as it seems to be attached at its "root," rather than having a root which is a free-moving structure, as in modern Homo. Thus, genioglossus contraction has no obvious acoustic effect in the chimpanzee.

### IV.2.b Hyoglossus

The hyoglossus originates from the greater horn of the hyoid and inserts into the body of the tongue. It appears to have the functional potential to move the tongue body posteriorly and inferiorly in both species. However, while Homo has space into which to move the tongue body in this direction—i.e., the pharyngeal cavity whose presence is the result of basicranial flexion and laryngeal descent—Pan does not.

The function of the hyoglossus muscles in vowel production is to produce the constriction for low (back?) vowels. The production of [ɑ] is characterized by displacement of the body of the tongue inferiorly and posteriorly. This is accomplished by contraction of the hyoglossus mm. (and possibly the pharyngeal constrictors). The location of the constriction corresponds to region C-bar in

Mrayati *et al.*'s model. The acoustic effect of a constriction at this point is a rasing of the first formant and lowering of the second formant. This produces the formant pattern associated with the vowel [ɑ].

In the chimpanzee, the hyoglossus muscles appear to have the functional potential to move the tongue body posteriorly and inferiorly. However, in the absence of a pharyngeal cavity, there appears not to be a space into which to move to the tongue body. Thus, hyoglossus contraction has no obvious effect in the chimpanzee.

### IV.2.c Styloglossus

The styloglossus muscle is present in both species, originating on the styloid process of the temporal bone and inserting into the tongue body. However, its orientation relative to the genioglossus m. and the tongue body as a whole differs between the two species. In modern Homo, the styloglossus is oriented at an angle to the genioglossus, such that its functional potential is to move the tongue body posteriorly and superiorly. Styloglossus contraction causes the tongue body to move perpendicular to the long axis of the tube, producing a constriction around the area of the velum.

Radiographic studies demonstrate that the production of [u] is characterized by displacement of the tongue body superiorly and posteriorly. Electromyographic studies (Smith 1971:65) demonstrate that this is accomplished by contraction of the styloglossus mm. The location of the constriction corresponds to region D-bar in Mrayati *et al.*'s model. A vocalic constriction at this point has the effect of lowering the second formant frequency.

The location of the constriction is near a pressure minimum for both the first and second formant. The model therefore suggests that [u] should have low first and second formants, as it does. In the human, styloglossus muscle contraction produces a sound with a high first formant and low second formant by moving the tongue body perpendicular to the vocal tube, near its mid-point. In contrast, in Pan, styloglossus contraction does not move the tongue body perpendicular to the axis of the vocal tract, but rather parallel to it. If such movement is in fact possible in the absence of a cavity posterior to the tongue and oral cavity, the effect would be expected to be that of an overall lowering of formant frequencies. However, since the styloglossus muscles do not run perpendicular to any region of the chimpanzee airway, styloglossus contraction would not have the effect of changing the cross-sectional area of any region of the vocal tract, and thus it would not be expected to have an effect on any one formant frequency.

## IV.3 Discussion

Vowels in human languages contrast in terms of two distinctive features: vowel height, and vowel backness. These distinctive features have acoustic correlates: the first formant frequency, for vowel height, and the second formant frequency, for vowel backness.

Distinctive features may be thought of as having some linguistic value by virtue of the presence of an opposition between their opposite values. The evidence reviewed above shows that production of these acoustic correlates of the distinctive features for vowels can be related to the activity of the extrinsic tongue musculature. In the human, the opposition between the functional potentials of the genioglossus muscle and the hyoglossus muscles is related to the opposition between high and low first formant values, i.e. the acoustic correlates of low vowel height and high vowel height. The opposition between the genioglossus and the styloglossus is related to the opposition between high and low second formant values, i.e. the opposition between vowel frontness and vowel backness. These acoustic oppositions are generatable in the human, but not in the non-human, because the anatomical oppositions to which they are relatable are present in the human, but not in the non-human.

It will be noted that the effect of styloglossus contraction, causing a constriction in region D-bar, on the first formant is not to decrease it, as must be the case for a high vowel. The model portrays a connection between height and backness, such that front vowels are predicted to be high, and back vowels are predicted to be low. This fact makes it difficult to derive /u/ from extrinsic tongue muscle activity alone, but should not be seen as a weakness for Mrayati *et al.*'s model. Rather, the model suggests a reason why lip rounding, which *does* cause first formant lowering, is involved in the production of high back vowels, and indeed is redundant to [+back], in the vast majority of the world's languages. Consulting tables of formant values in a variety of sources (e.g. Ladefoged 1993:197 for American English; Shalev, Ladefoged, and Bhaskararao 1993:91 for Toda; Blankenship, Ladefoged, Bhaskararao, and Chase 1993:129 for Khonoma Angami; Bradlow 1993:24 for English and Spanish) it will be seen that high back vowels generally have lower F1's than do front vowels of the same height. This cross-linguistic pattern is not surprising, given the association between front vowel articulations and low first formant values which this model suggests.

The comparative anatomical evidence from Pan troglodytes suggests that potential opposition in functional potential of the extrinsic tongue musculature existed already in the archaic hominids: in Pan, the genioglossus could potentially move the tongue body in a posterior-anterior direction, while the styloglossus could potentially move it in an anterior-posterior direction. Similarly, the hyoglossus could potentially move the tongue body inferiorly, while the genioglossus and styloglossus could move it superiorly. Basicranial flexion changed the origins, or bony attachment points, of these muscles, and thus their functional potentials relative to each other. Laryngeal descent and the resultant presence of a pharyngeal cavity posterior and inferior to the oral cavity allowed for movement of the tongue in dimensions not previously possible—posteriorly in response to styloglossus contraction, inferiorly in response to hyoglossus contraction. The effect of basicranial flexion and laryngeal descent, then, is to allow these preadapted oppositions to have acoustic consequences, where before they did not.

## V. Conclusion

The production of vocal language requires the ability to generate contrasts and the ability to produce contrasting sounds with stability. Our data show that the abilities to produce distinctions in vowel height and vowel backness are direct results of the evolutionary changes which have lead to the anatomically modern human having a supralaryngeal airway which is qualitatively different from that of all other animals. Perhaps more importantly, they suggest that the relationship between articulations and acoustic outputs in the modern human is characterized by regions of acoustic stability. These regions make it more possible to produce speech: if there is a larger range of values for some articulatory parameter (constriction location, in these cases) which will yield some desired acoustic output, then one can be less precise in attempting to produce it. Note that the differences between the acoustic outputs of the two sorts of supralaryngeal airways arise not from the ability, or lack thereof, to produce any one particular sort of articulation, or any one particular vowel. Rather, they arise from the *ranges* of articulations producible in the two sorts of airways. Nor is it being claimed that the differences in acoustic outputs results from any *acoustic* effect of the bend in the airway: rather, they arise from the different range of *articulations* which result from the presence of a bend in one of the boundary walls of the airway.

It should be noted that although I disagree with Lieberman's interpretation of the acoustic significance of human vocal tract evolution, my findings are not incompatible with the theory of human language evolution which he has developed, and about which I am agnostic. His theory requires that there be some qualitative difference between human and non-human supralaryngeal airways, and substitution of the contents of this paper for the analogous section in any of his publications would harm his overall theory not a whit.

Nonlinearities in articulatory-acoustic relations have been the topic of much discussion in past years. Some have felt them not to be relevant to the description of speech, e.g. Ladefoged and Lindau (1989). Others have looked for explanations from some more general system, independent of the specific characteristics of speech, as do Abry *et al.* (1989). They are the expected finding within the framework of a theory that looks to nonlinearities in articulatory/acoustic relations to explain the workings of the phonetic/phonological component of the grammar. The work of Stevens (1972, 1989), Wood (1982, 1986), Beckman *et al.* (1995), and others suggests that human speakers utilize these sorts of articulatory/acoustic relations. My research suggests that they characterize articulatory/acoustic relationships in human, but not in non-human, airways.

## VI. Postscript I: vocal tract models

Stevens and House (1955) varied the location of a constriction by changing the distance from the glottis of the high point of a parabola. Models such as those of Maeda (1990) and Lindblom and Sundberg (1971) are based on measurements of the dorsal surface of the tongue relative to the palate and posterior pharyngeal wall during the production of vowels. Constriction location is varied by extrapolating the points through which the tongue surface would pass when moving between the actual measured locations. Stevens (1972) and (1989) are based on a model in which the location of a constriction is varied by trading off the lengths of the cavities anterior and posterior to the constriction. Carré and Mrayati (1990) vary constriction location by means of successive transversal changes of cross-sectional area in a series of concatenated tubes. All such models have in common the characteristic of treating the variation of the location of a constriction in any one part of the vocal tract as a task just like variation of the location of a constriction in any other part of the vocal tract. They fail to capture the generalization that varying the location of a constriction affects the relative sizes of the cavities anterior and posterior to the constriction differently in the front and back regions of the vocal tract. Jackson (1988) provides a comprehensive review of models of vowel production, none of which take into account this aspect of the geometric relationships between the front and back regions of the vocal tract.

## VII. Postscript II: Mrayati *et al.*'s perturbation theory

Mrayati *et al.*'s perturbation theory, and the theory of distinctive regions and modes which is derived from it, is based on consideration of the effects of cross-sectional area on the total, potential, and kinetic energy of small, local areas of the vocal tract. They use these parameters to calculate the effect on the resonances of the vocal tract as a whole of small, local changes in area. In these calculations, the total energy of the vocal tract is considered to be a constant, as is reasonable for the lossless case, i.e. if the loss of energy through damping by the soft tissue of the vocal tract, radiation out of the mouth, etc., is ignored.

In this theory, potential and kinetic energy are analogous to the familiar measures of a flowing gas: pressure and velocity. The vocal tract will resonate at frequencies related to the pressure waves whose wavelength is optimum for a vocal tract of a given length. These waves of different frequencies have pressure and velocity maxima and minima at different points along the length of the vocal tract. The familiar figures in Chiba and Kajiyama (1941) show the locations of the velocity maxima (marked $N_n$) and minima (at the points where lines cross). Pressure maxima are not labelled; they occur at the points of the velocity minima.

The relationship between potential and kinetic energy and cross-sectional area is that (my emphasis):

(1) "**potential energy** density **is proportional to the area** and to the square of the sound pressure" (Mrayati *et al.* 259)
(2) "**kinetic energy** density **is** proportional to the square of the flow velocity and **inversely proportional to the area**" (Mrayati *et al.* 259)

So, the effect of a constriction (i.e., a reduction in area) is to increase the kinetic energy and to decrease the potential energy. (If this seems counterintuitive, it's because we're talking here not about the effect of a change of volume on a closed cylinder in which a gas is *contained*, which is what we're used to thinking about in introductory chemistry courses, but about the size of an open tube through which a gas is *travelling*.) The effect of producing a constriction differs in different areas of the vocal tract because the initial conditions differ in different areas of the vocal tract, as illustrated in Chiba and Kajiyama (1941).

The equation given in Mrayati *et al.* for calculating the effect of a change in area on a formant frequency is

$$\frac{\Delta F}{F} = \sum_{n=1}^{N} \frac{KE_n - PE_n}{TE_n} \frac{\Delta A_n}{A_n}$$

where

| | | |
|---|---|---|
| F | = | formant frequency |
| KE | = | kinetic energy |
| PE | = | potential energy |
| TE | = | total energy |
| A | = | area |

Consider the case where kinetic energy is large and potential energy is small: the value of the expression *KE–PE* will be large and positive. If the change in area is a decrease (rather than an increase), then the value of $\Delta A_n$ will be negative, the value of the expression $\Delta A_n/A_n$ will be negative, and the change in the frequency of the formant will consist of a large decrease.

Such is the situation at the lips. Each of the resonant frequencies of the vocal tract has a velocity maximum and a pressure minimum at this point. Since potential energy is proportional to (the square of) the sound pressure, and pressure is at a minimum at this point, the potential energy is low. Kinetic energy is proportional to (the square of) flow velocity, and velocity is at a maximum at this point, so the kinetic energy is high. Therefore, the model predicts that the effect of a constriction at the lips should be a lowering of all of the first three resonant frequencies of the vocal tract, as is in fact the case.

As a further example, consider the effect of a constriction in the palatal area. Examining the figures in Chiba & Kajiyama (1941), we see that for the first resonant frequency, velocity is quite high, though not quite as high as it is at the lips. The second resonant frequency has a velocity minimum and a pressure maximum in this area. The situation regarding the first formant will be similar to that described above, except that the effects of the constriction will be just slightly less than at the lips, since velocity at this point is slightly less than at the lips. For the second resonant frequency, since kinetic energy is proportional to velocity and velocity is at a minimum at this point, the kinetic energy is low. Since potential energy is proportional to pressure and pressure is at a maximum at this point, the potential energy is high. Thus, the value of the expression *KE–PE* will be large and negative. If the change in area is a constriction, i.e. a reduction in area, then the term $\Delta A_n$ will be negative, the value of the expression $\Delta A_n/A_n$ will be negative, and the effect of the constriction will be a large increase in the frequency of the second formant. The model thus predicts that a constriction in the palatal area should produce (a) an F1 which is quite low, but not quite as low as that produced by a constriction at the lips, and (b) a high F2. Both of these predictions are correct.

## References


Abry, Christian; Louis-Jean Boe; and Jean-Luc Schwartz. (1989) Plateaus, catastrophes and the structuring of vowel systems. Journal of phonetics (17)47-54.

Badin, Pascal Perrier; Louis-Jean Boë; and Christian Abry. (1990) Vocalic nomograms: acoustic and articulatory considerations upon formant convergences. Journal of the Acoustical Society of America 87(3)1290-1300.

Beckman, Mary E.; Tzyy-Ping Jung; Sook-hyang Lee; Kenneth DeJong; Ashok K. Krishnamurthy; Stanley C. Ahalt; K. Bretonnel Cohen; and Michael J. Collins. (1995) Variability in the production of quantal vowels revisited. Journal of the Acoustic Society of America 97(1)471-490.

Blankenship, Barbara; Peter Ladefoged; Peri Bhaskararao; and Nichumeno Chase. (1993). Phonetic structures of Khonoma Angami. UCLA working papers in phonetics (84)127-142.

Bradlow, Ann R. (1993). Language-specific and universal aspects of vowel production and perception: a cross-linguistic study of vowel inventories. Ithaca, NY.: Cornell University.

Carré, Rene and M. Mrayati. (1990) Articulatory-acoustic-phonetic relations and modeling, regions and modes. Speech production and speech modelling. Hardcastle, William J. and Alain Marchal, eds. Netherlands: Kluwer Academic Publishers.

Chiba, T., and M. Kajiyama. (1941) The vowel: its nature and structure. Tokyo: Kaiseikan.

Fant, Gunnar. (1960) Acoustic theory of speech production. Netherlands: Mouton.

Hoffman, Paul R.; Gordon H. Schuckers; and Raymond G. Daniloff. (1989) Children's phonetic disorders: theory and treatment. Boston: College-Hill Press.

Honda, Kiyoshi. (1994) Somatoneural relation in the auditory-articulatory linkage. ATR Human Information Processing Research Laboratories manuscript.

Honda, Kiyoshi.; H. Hirai; and N. Kusakawa. (1993) Modeling vocal tract organs based on MRI and EMG observations and its implication on brain function. Annual bulletin of the Research Institute in Logopedics and Phoniatrics (27)37-49. Tokyo.

Jackson, Michel T.T. (1988) Phonetic theory and cross-linguistic variation in vowel articulation. UCLA working papers in phonetics 71.

Jacob, Stanley W. and Clarice A. Francone. (1965) Structure and function in man. Philadelphia: W.B. Saunders.

Johnson, Keith A. (1994) Acoustic phonetics for linguists. Ohio State University manuscript.

Kakita, Yuki; Osamu Fujimura; and Kiyoshi Honda. (1985). Computation of mapping from muscular contraction patterns to formant patterns in vowel space. Phonetic linguistics: essays in honor of Peter Ladefoged, pp. 133-144. Victoria Fromkin, ed. Academic Press.

Kent, Raymond D.; Bishnu S. Atal; and Joanne L. Miller. (1991) (Commentary on) vocal tract and acoustic relationships. Papers in speech communication: speech production, pp. 267-268. R.D. Kent *et al.*, eds. Woodbury, New York: Acoustical Society of America.

Ladefoged, Peter. (1993). A course in phonetics. Harcourt Brace Jovanovitch.

Ladefoged, Peter, and Anthony Bladon. (1982) Attempts by human speakers to reproduce Fant's nomograms. Speech communication (1)185-198.

Ladefoged, Peter, and Mona Lindau. (1989) Modeling articulatory-acoustic relations: a comment on Stevens' "On the quantal nature of speech." Journal of Phonetics (17)99-106.

Laitman, Jeffrey T. (1984) The anatomy of human speech. Natural history 8/84:22-26.

Laitman, Jeffrey T. and J.S. Reidenberg. (1988) Advances in understanding the relationship between the skull base and larynx with comments on the origin of speech. Human evolution 3(1-2)99-109.

Lieberman, Philip. (1975) On the origins of language. The Macmillan series in physical anthropology. New York: Macmillan.

------. (1984) The biology and evolution of language. Cambridge: Harvard University Press.

------. (1991) Uniquely human: the evolution of speech, thought, and selfless behavior. Cambridge: Harvard University Press.

Lindblom, Bjorn, and J. Sundberg. (1971) Acoustical consequences of lip, tongue, jaw, and larynx movements. JASA (50)1166-1179.

Maeda, Shinji. (1990) Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. Speech production and speech modelling. Hardcastle, William J. and Alain Marchal, eds. Netherlands: Kluwer Academic Publishers.

McMinn, R.M.H., and R.T. Hutchings. (1977) Color atlas of human anatomy. Chicago: Year Book Medical Publishers, Inc.

Moore, Christopher A. (1992) The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. Journal of speech and hearing research 35(5)1009-1023.

Mrayati, M.; R. Carré; and B. Guérin. (1988). Distinctive regions and modes: a new theory of speech production. Speech Communication (7)257-286.

Negulesco, John A. (1993) The anatomy of the head and neck. Department of Anatomy, College of Medicine, Ohio State University.

Perkell, Joseph S. (1969) Physiology of speech production: results and implications of a quantitative cineradiographic study. Cambridge: MIT Press.

Shalev, Michael; Peter Ladefoged; and Peri Bhaskararao. (1993). Phonetics of Toda. UCLA working papers in phonetics (84):89-126.

Smith, T. S. (1971) A phonetic study of the function of the extrinsic tongue muscles. UCLA Working papers in phonetics (18). Los Angeles.

Stevens, Kenneth N. (1972) Quantal nature of speech. Human communication: a unified view, pp. 51-66. E.E. David, Jr. and P.B. Denes, eds. New York: McGraw-Hill.

------. (1989) On the quantal nature of speech. Journal of phonetics (17)3-45.

Stevens, Kenneth N. and Arthur S. House. (1955) Development of a quantitative description of vowel articulation. JASA (27)484-493.

Stone, Maureen; Alice Faber; Lawrence J. Raphael; and Thomas H. Shawker. (1992) Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. Journal of Phonetics (20)253-270.

Swindler, D. R. and C. D. Wood. (1973). An atlas of primate gross anatomy. Seattle: University of Washington Press.

Wood, Sidney. (1982) X-ray and model studies of vowel articulation. Lund University working papers (23).

------. (1986) The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. JASA 80(2)391-401.

*OSU Working Papers in Linguistics* 45 (36-55)

# Final Lowering in Kipare*

**Rebecca Herman**
rherman@ling.ohio-state.edu

**Abstract:** Data from Kipare, a Bantu language which has both lexical tone and final lowering, were examined to determine the optimal model of final lowering in a tone language. The models evaluated are the register model, in which final lowering is accounted for by manipulating tone categories, and the boundary tone model, in which final lowering is accounted for by a separate tone on the same tier as lexical tones but associated to a higher prosodic position. Aspects of final lowering in Kipare examined in this study include which lexical tone categories show final lowering, whether the lowered tones coincide with tones from another lexical tone category, the locality of the effect, the relation of contrasting utterance types to utterance types showing final lowering, and what the domain of final lowering is.

## 1. Introduction

In many languages, fundamental frequency decreases utterance-finally. For example, final lowering is documented in Japanese by Pierrehumbert and Beckman (1988), in English by Liberman and Pierrehumbert (1984), in Dutch by Gussenhoven and Rietveld (1988), in Danish by Thorsen (1985), in Yoruba by Laniran (1992), and in Kikuyu by Clements and Ford (1981). Ladefoged (1982) generalizes that "In nearly all languages the completion of a grammatical unit such as a normal sentence is signaled by a falling pitch."

In tone languages, tone functions contrastively. Tonal categories such as high (H) and low (L) must be posited to account for the contrasts. In such cases, it is possible to analyze a decrease in fundamental frequency utterance finally as a change from one tonal category to a lower one. Such analyses are prevalent in the literature on tonal phenomena in African languages.

For example, Clements and Ford (1981) describe "automatic downstep" in Kikuyu, where the last low tone (or low tone sequence) of a noninterrogative sentence is downstepped. They analyze this as a rule inserting a downstep. The rule is formulated as: $\emptyset \dashrightarrow\ !\ /\ \_\_\ L_q\ (H_q)\ ]_S$, where "q" stand for "any number of." A further rule suspends this effect in positively oriented questions by deleting the downstep.

In a related case, Elugbe (1977) describes final low raising in statements and lack thereof in questions in Isoko and Urhobo. This is a counterexample to Ladefoged's generalization about falling pitch signaling "a normal sentence." In these languages, it is questions which are distinguished by final lowering. No

* The bulk of funding for this project was supplied by the Linguistics Department at the Ohio State University. The paper was completed with support from an NSF Graduate Fellowship. I would like to thank Frida Tungaraza for her generosity with her time and for her tireless, good-humored work. This paper would not have been possible without her. Thanks also to M. Beckman, K. Johnson, S.-A. Jun, and D. Odden for guidance with this project. I am very grateful to C. McDougall for support and help. Finally, thanks to S. Jannedy and J. Venditti for technical assistance in the Lab.

matter what the function of final lowering, though, it is still analyzed as a change to a lower tone category. Thus, questions are analyzed with a final floating low which causes highs to fall but merges with low.

Roberts (1992) describes phrase-final lowering in Sukuma whereby a high tone is lowered to extra-low phrase-finally. This is analyzed as a rule changing the category H phrase-finally to the category XL.

The question is whether analyses such as those just described, which manipulate phonological tones, provide the optimal account of final lowering. Such analyses use rules like H --> !H / __# and !H --> L / __#. Based on evidence from contour tones, register shifts, "key raising," and natural classes of tones, it has been argued (Clements (1981), Inkelas and Leben (1990), and Snider and van der Hulst (1993)) that the tonal categories in such rules should be reinterpreted. Rather than indivisible entities, the categories may be defined in terms of feature combinations in the register model. In the register model, there are two tiers of tones, the pitch and the register tones. The representations of various tones in the register model are shown in Table 1.

| | H | !H | L | !L |
|---|---|---|---|---|
| register tier | h | l | h | l |
| primary tier | H | H | L | L |

Table 1. The representations of tones in the register model.

Assume a language in which H, !H, and L are each lowered to the next category down in final position. In such a system, the lowering of H to !H utterance-finally would involve changing an underlying register h to register l. The lowering of !H to L utterance-finally would involve changing an underlying primary H to primary L and an underlying register l to register h. And the lowering of L to !L would involve changing register h to register l. Although an account of final lowering is possible using such a system, it seems problematic, since a single phenomenon is analyzed in a non-uniform way.

The predictions of the register model for final lowering in a system in which each tone in final position lowers to the next tone category down are:

(i) The fundamental frequency values of a tone in final position should coincide with the fundamental frequency values of the next lower tone in medial position (assuming they are preceded by the same tonal sequence). For example, the values of a H tone finally should coincide with the values of a !H medially, all else being equal, since they have the same representation. This is schematized in (1), where (a) and (b) are predicted to be identical.

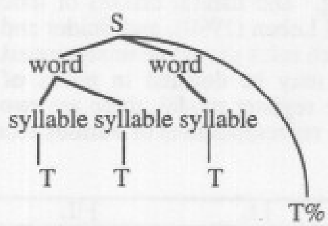
(1) (a) ... /H/ #
(b) .../!H/ ...

(ii) Final lowering should affect only the final tone-bearing unit, since only the final tone-bearing unit changes its representation. Alternatively, the change could spread to previous tone-bearing units, in which case it would be expected to affect each tone-bearing unit in the same way.

(iii) Contrasts between representations manifested via these tones should be localized in meaning to the domain to which they are associated.

An alternative analysis of final lowering, put forth by Pierrehumbert and Beckman (1988), uses boundary tones. In this model, delimitative peripheral tones are attached directly to higher nodes, and are produced at about the same time as the initial or final lexical tones (p. 127). The boundary tones are linked to higher nodes in the prosodic tree, so they are available to affect the values of tones linked at lower levels. They are, in effect, local to the whole phrase since they are viewed as a property of the phrase (p. 75). This is schematized in (2), where T% represents the boundary tone and each T represents a lexical tone. S in this case represents a sentence, although Pierrehumbert and Beckman use the intonational phrase as the prosodic node to which boundary tones are attached.

(2)




The predictions of the boundary tone model of final lowering are:

(i) All tone categories should lower by the same amount when affected by a boundary tone. Medial !H and final H would not be expected to coincide, since they would have different representations. Rather, what might be expected is a relationship between medial and final tones of the same category. This relationship should then hold across all tonal categories. So perhaps what would be expected is a proportion of the sort "Final H is to an immediately preceding H as final !H is to an immediately preceding !H." For example, Silverman (1987) quantifies final lowering as the compression of the pitch range to be 30% narrower than the immediately preceding range. One potential problem with this prediction is that low tones are scaled differently than high tones (as discussed in Beckman and Pierrehumbert, 1992).

(ii) Final lowering has the greatest effect on extreme final tone-bearing units, but since it has scope over the entire utterance, it may also affect previous tone-bearing units to a lesser extent. Thus, the boundary tone "begins to show its influence on tones at some distance from the end" (p. 162).

(iii) Contrasts between representations manifested via different boundary tones should be contrasts in utterance types.

Kipare is a Bantu language which has both lexical tone and final lowering. It has been observed to have lowering of high tones to downstepped highs utterance-finally, lowering of downstepped high tones to lows utterance finally (Odden, 1986 a and b), and lowering of lows to a lower value utterance-finally (Odden, personal communication). Odden (1986a) notes that the rule lowering downstepped high tone to low tone must be ordered before the rule lowering high tones to downstepped high tones, implying that an underlying high tone does not lower to downstepped high tone utterance finally, then lower further to low tone. Rather, this implies that either lexical high lowers to downstepped high utterance finally, or else lexical downstepped high lowers to low utterance finally. However, downstepped high derived from lexical high does not lower further to become low. Thus Kipare provides a test case with which to evaluate the nature of final lowering in a tone language.

## 2. Methods

The speaker was a female Kipare speaker from the Same district, Kilimanjaro region, in Tanzania. She has lived in the United States for seven years. Her primary language is Kipare, but she also speaks English and KiSwahili.

Investigations of final lowering were conducted by first recording the subject reading sentences. Each sentence was recorded five times. The order in which the sentences were read was randomized (by shuffling 3-by-5 index cards on which the sentences were written) to control for position in the list of sentences. Recordings were done in a double-walled sound booth at the Ohio State University Linguistics Laboratory using a TEAC V-427C cassette deck and a head-mounted microphone. Then, the recordings were digitized and pitch traces were created using *Waves*™ (Entropic Research Laboratory, Inc., 1993) . The fundamental frequency was measured from the pitch traces in the center of each target vowel, as shown in Figure 1.

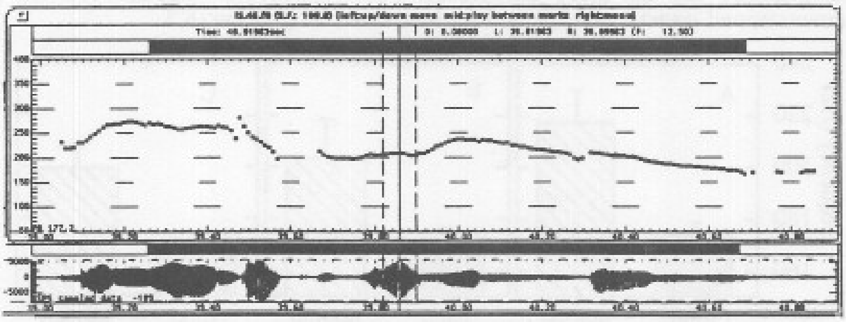


Figure 1. The dashed lines indicate where the target vowel is in the pitch trace. The solid line indicates where fundamental frequency was measured in the target syllable [gwà]. The sentence shown here is [mì nìéndá kùgwà hényùmbá yò dù], "Me, I would like to fall only today."

The corpus was designed to allow examination of the fundamental frequency of high-toned syllables, downstepped high-toned syllables, and low-toned syllables in final and in non-final positions. In order to rule out segmental effects, similar syllables, all involving the sequence [wa], were used to test all of these categories. The word [kùnwá], "to drink," provided the high tone with the syllable [nwá], the word [ná!nwá], "I drank," provided the downstepped high tone with the syllable [!nwá], and the word [kùgwà], "to fall," provided the low tone with the syllable [gwà]. Since the downstep is grammatically conditioned by the subject prefix [ná] in [ná!nwá], "I drank," it was impossible to have identical syllables immediately preceding the target syllable. However, the corpus was designed to minimize such variation.

In order to verify that all three tonal categories of Kipare, namely H, !H, and L, do lower, each word was placed in various positions within carrier sentences: preceded by 0, 2, 3, 4, 5, and 6 syllables; followed by 0, 2, 3, 4, and 5 syllables; and every combination of these. (It was impossible to find appropriate single syllable carrier words in either preceding or following environments.) The preceding and following tone patterns were not uniformly high or uniformly low (again, due to problems finding utterances in the language with all H or all L tones).

Instead, the tone patterns built on the tone pattern of the next shortest sentence. So two syllables preceding the verb had the tone pattern HL, three syllables preceding the verb had the tone pattern LHL, four syllables had the tone pattern LLHL, five syllables had the tone pattern HLLHL, and six syllables had the tone pattern HLLLHL. Similarly for following tone patterns, where two syllables following the verb had the tone pattern HL, three syllables had the tone pattern HLH, four syllables had the tone pattern HLHL, and five syllables had the tone pattern HLHLL. See Appendix 1 for a complete list of sentences.

### 3. Do all tonal categories lower utterance-finally?

When comparing the mean fundamental frequency of all recorded tokens which were utterance-final with the mean of all recorded tokens which were non-final, it is clear that the fundamental frequency for all three tonal categories is lower utterance finally than utterance medially. Unpaired t-tests show that the means of the utterance final tokens are different from the means of the utterance medial tokens at $p<.0001$ for all three categories. For high-toned targets, $t=-17.8$, for downstepped targets, $t=-11.3$, and for low-toned targets, $t=-18.3$.

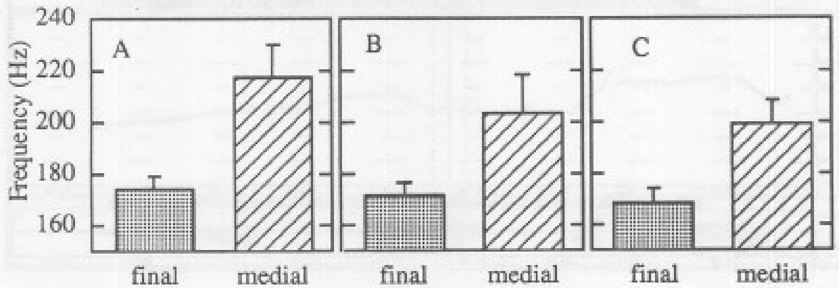



Figure 2. Plot of mean fundamental frequency values with standard deviation bars for utterance final and utterance medial tokens.
A: The F0 of target syllables for all tokens with H targets.
B: The F0 of target syllables for all tokens with !H targets.
C: The F0 of target syllables for all tokens with L targets.

| | | no. of tokens | mean F0 | stand. dev. |
|---|---|---|---|---|
| H toned target | final | 30 | 174.23 Hz | 5.05 |
| | medial | 120 | 217.19 Hz | 12.97 |
| !H toned target | final | 30 | 171.52 Hz | 4.83 |
| | medial | 120 | 203.45 Hz | 15.30 |
| L toned target | final | 30 | 168.07 Hz | 6.40 |
| | medial | 120 | 199.49 Hz | 8.80 |

Table 2. Number of tokens, mean F0, and standard deviations for all three tonal categories in final and medial positions.

As Figure 2 A-C show, not only does the phenomenon of final-lowering occur in this Kipare speaker's speech, but also all three tonal categories show final lowering. This could be explained by either the register model or the boundary tone model. However, the distribution of "super-low" tones in the register model must be stipulated (via a rule inserting register l in a specific position) since !L tones only occur utterance finally. In the boundary tone model, the distribution of "super-low" tones is understood as the effect of boundary tones only occurring at the edges of prosodic categories.

## 4. Do categories coincide?

Given the register model, the expectation is that high tones occurring utterance-finally would have the same values as downstepped high tones occurring medially since they have identical representations. In order to test this, the same data set was used as was used in section 3, only making different comparisons. Underlying H tones occurring utterance finally were compared with !H tones (derived via having a H toned grammatical marker before a H toned syllable) occurring medially. The prediction is that /H#/ and /!H.../ should have comparable fundamental frequencies when preceded by identical tonal contexts, since both are represented with a H on the primary tier and a l on the register tier. Similarly, !H tones occurring utterance finally were compared with L tones occurring utterance medially. The prediction is that /!H#/ and /L.../ should also have comparable fundamental frequencies when preceded by identical tonal contexts, since both are represented with a L on the primary tier and a l on the register tier.

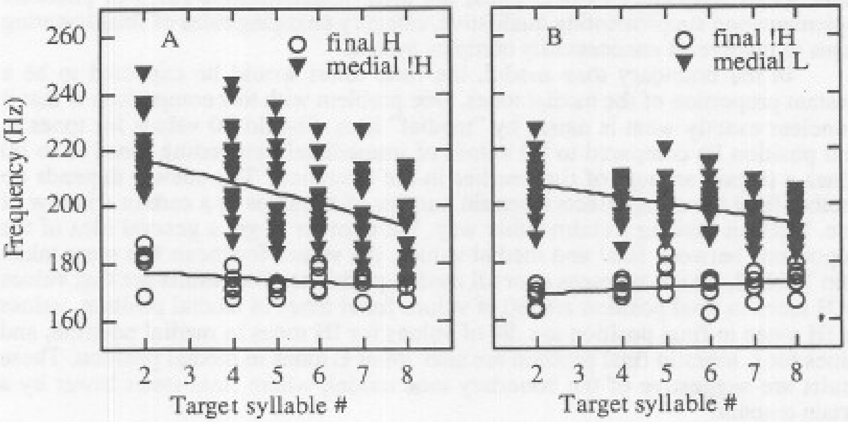



Figure 3.
A: Final high tones compared with medial downstepped high tones
B: Final downstepped high tones compared with medial low tones

It is not the case that final H tones coincide with medial !H tones, as shown in Figure 3A. In Figure 3A, the values of high tones utterance-finally are shown together with the values of downstepped high tones utterance medially. The X-axis shows which syllable number in the utterance the target syllable is. For example, in the utterance [kùnwá], the target syllable [nwá] is syllable number 2 and is in final position. If high tones changed to downstepped high tones utterance-finally, the two groups would be expected to overlap. However, they do not. Similarly, the expectation is that downstepped high tones occurring finally would have the same values as low tones occurring medially, since they have identical representations. Again, this is not the case, as shown in Figure 3B.

In the register model where medial downstepped tones and final high tones have identical representations, another tone change might be cited as the cause of this lack of overlap. For example, many African tone languages have a rule raising the last non-low tone before a low tone. Yoruba (Laniran, 1992) is one such language. Hausa (Inkelas, Leben, and Cobler, 1986) is another. Kipare has an independently attested pre-L raising of this sort. (Odden, personal communication.) However, such a pre-L raising cannot be the explanation for the lack of overlap in these data since none of the target medial !H tones in the corpus preceded a L tone.

Or, instead of another tone change to account for this lack of overlap, perhaps "phonetic implementation" could be cited. That is, the fundamental frequency targets are identical at some level but then phonetic effects "bump down" the final tones. This expresses the same idea as is found in Pierrehumbert (1980), where she argues that the hierarchical representation would not be an alternative to replace tone mapping rules, but must be supplemented by rules expressing the same regularities. The question is, if this sort of "phonetic implementation" is a necessary supplement to the register model, then why introduce the additional level of structure involved in the register tier at all? Why not read the phonetic values directly from the string of tones? As Pierrehumbert (1980) argues, having a level of representation between the underlying sequence of tones and the fundamental frequency contour is superfluous, since it is possible to map tones directly into fundamental frequency values. Thus, the idea of quantitative rules of phonetic implementation supplementing qualitative, category changing rules of final lowering seems to involve an unnecessarily complex system.

In the boundary tone model, the final tones would be expected to be a constant proportion of the medial tones. One problem with this comparison is that it is unclear exactly what is meant by "medial" here. Should F0 values for tones in final position be compared to F0 values of immediately preceding tones or to F0 values a certain amount of time earlier in the utterance? The answer depends on whether final lowering affects a certain number of syllables or a certain window of time. Without making a claim either way, but in order to get a general idea of the proportions between final and medial values, the values for mean F0s were taken from Table 2, which averages over all medial positions. The results are that values for H tones in final position are .80 of values for H tones in medial position, values for !H tones in final position are .84 of values for !H tones in medial position, and values for L tones in final position are also .84 of L tones in medial position. These results are suggestive of the boundary tone model, where final tones lower by a certain amount.

## 5. Thinking locally or acting globally?

As shown above, the fundamental frequency lowers dramatically utterance-finally. However, the question is whether this lowering is just a local effect, affecting only a single tone-bearing unit, or whether the phenomenon affects previous tone-bearing units as well. In order to test this, the same data set as in section 3 was examined. This time, target syllables at varying distances from the end of the utterance were compared. Distance from the end of the utterance was measured in terms of number of syllables.

The closer the target syllable is to the end of the utterance, the lower the fundamental frequency is. Figures 4 A-C show mean values of the fundamental frequency (in Hz) plotted against the number of syllables following the target syllable in the utterance. In general, the fundamental frequency drops lower and lower when followed by less and less syllables, plunging at extreme utterance-final position. Targets followed by a single syllable were not available for comparison (due to problems finding appropriate words in the language). However, in all three graphs the drop in frequency between 2 and 0 is clearly greater than twice the drop in frequency between any other two points along the X-axis. This leads to the speculation that if utterances where the target syllable was followed by a single syllable were available, the mean frequency would be slightly higher than halfway between the mean for 0 and the mean for 2.

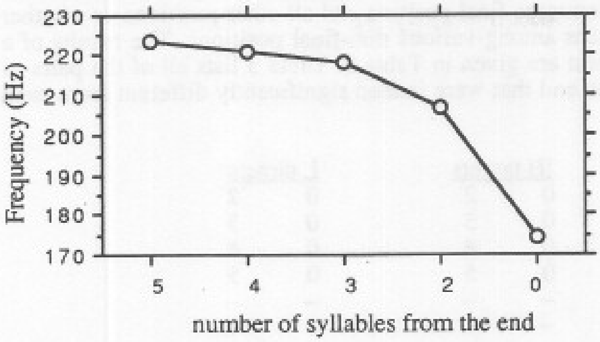



Figure 4A. Mean values of F0 for H target syllables

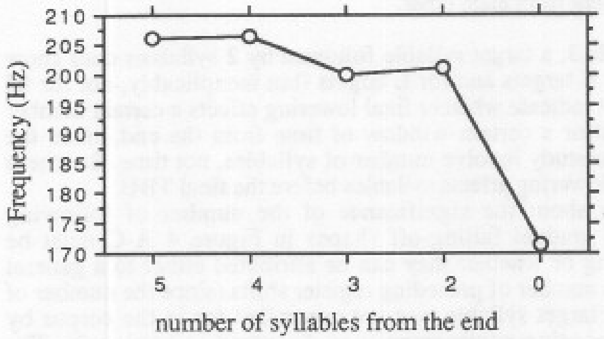



Figure 4B. Mean values of F0 for !H target syllables

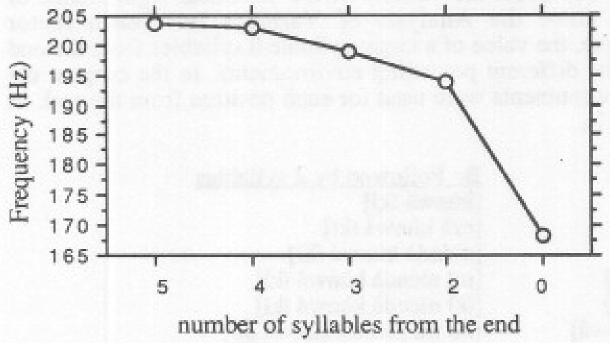



Figure 4C. Mean values of F0 for L target syllables

A 2-factor Analysis of Variance (ANOVA) was performed using the number of syllables preceding the target, the number of syllables following the target, and the fundamental frequency as factors. The ANOVA shows that the number of following syllables is statistically significant at p=.0001 for H, !H, and L in determining the fundamental frequency of the target syllable.

The question is whether this statistical significance is just a result of the large difference between extreme final position and all other positions or whether this also reflects differences among various non-final positions. The results of a post-hoc Tukey-Kramer test are given in Table 3. Table 3 lists all of the pairs of syllable positions from the end that were indeed significantly different from each other at a .05 level.

| H targets | | !H targets | | L targets | |
|---|---|---|---|---|---|
| 0 | 2 | 0 | 2 | 0 | 2 |
| 0 | 3 | 0 | 3 | 0 | 3 |
| 0 | 4 | 0 | 4 | 0 | 4 |
| 0 | 5 | 0 | 5 | 0 | 5 |
| 2 | 3 | -- | -- | -- | -- |
| 2 | 4 | -- | -- | 2 | 4 |
| 2 | 5 | -- | -- | 2 | 5 |

Table 3. Pairs of syllable positions from the end that were significantly different from each other.

According to Table 3, a target syllable followed by 2 syllables does show final lowering effects for H targets and for L targets (but inexplicably, not for !H targets). These data do not indicate whether final lowering affects a certain number of syllables from the end or a certain window of time from the end, since the measurements used in this study involve number of syllables, not time. But these results do show that final lowering affects syllables before the final TBU.

Another question about the significance of the number of following syllables is whether the gradual falling-off shapes in Figure 4 A-C must be attributed to final lowering or whether they can be attributed either to a general declination trend or to the number of preceding register shifts (since the number of downsteps preceding the target syllable was not controlled for in the corpus by having all L or all H preceding environments, as discussed in section 2). The number of preceding syllables is in fact significant at p=.0001 for H, !H, and L tones, indicative of declination. However, the statistical significance of position from the beginning does not affect the validity of the statistical significance of position from the end since the Analysis of Variance tests each factor independently. For example, the value of a target syllable 0 syllables from the end is calculated across all the different preceding environments. In the corpus, the same set of preceding environments were used for each position from the end, as shown in the example in (3).

| (3) | A. Final Position | B. Followed by 2 syllables |
|---|---|---|
| | [kùnwá] | [kùnwá íkì] |
| | [ézà kùnwá] | [ézà kùnwá íkì] |
| | [nìéndà kùnwá] | [nìéndà kùnwá íkì] |
| | [mì nìéndà kùnwá] | [mì nìéndà kùnwá íkì] |
| | [íkì nìéndà kùnwá] | [íkì nìéndà kùnwá íkì] |
| | [íkì mì nìéndà kùnwá] | [íkì mì nìéndà kùnwá íkì] |

Although there may be register shifts in the preceding environment, there will be the same number of register shifts being calculated in at each position from the end.

In sum, the shape of the figures in Figure 4 A-C is indicative of final lowering as separate from declination or register shifts.

In the register model, only the extreme final syllable would be affected by final lowering, but in the boundary tone model, the effect is expected to "reach in" from the end of the sentence. The results obtained here show that final lowering affects more than just the last syllable, which does not match the predictions of the register model.

One further problem related to this phenomenon is that in a sequence of downstepped H tones, all of the tones lower to L utterance finally whereas when there is a HH sequence, only the last H lowers to !H utterance finally (Odden, 1986a and b). This suggests that these are two different processes (and so would pose a problem to any model attempting to provide a unified account). However, the data examined here involve only monosyllabic verbs, so there is no evidence bearing on this problem in this study.

## 6. Is there a contrast?

Final-lowered utterances are not the only possible utterance type in Kipare. There is also a contrasting utterance type with final raising of pitch that expresses doubt, disbelief or incredulity. Sentences segmentally identical to the declaratives used earlier and with the same lexical tone patterns were elicited, but in contexts intended to express disbelief. (Only sentences with third-person subjects were used in this section, since first-person subjects seemed pragmatically odd in such contexts.) Figure 5 shows a comparison of an incredulous sentence (upper panel) with a declarative sentence (lower panel). See Appendix 2 for a complete list of utterances used in this section of the study.

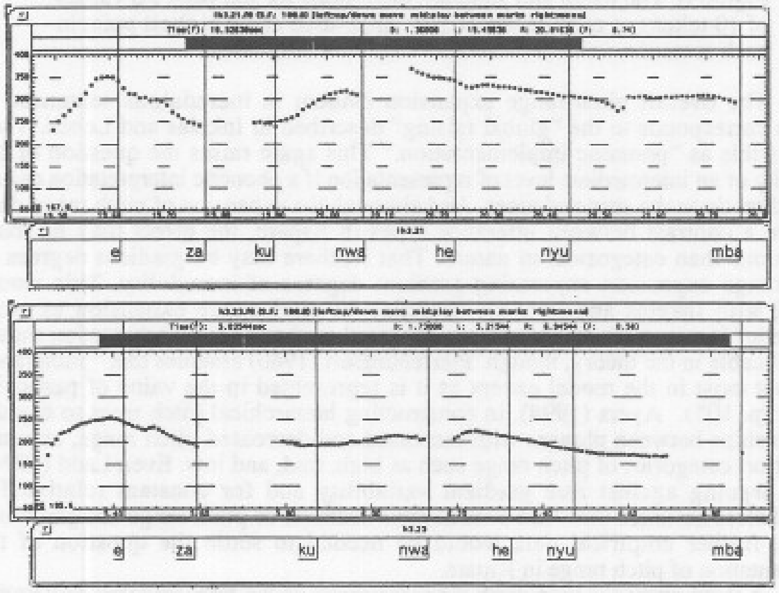



Figure 5. Upper panel: incredulous. Lower panel: declarative.
[ézà kùnwá hényùmbá] "He/she comes to drink in the house."

The expression of incredulity via manipulation of pitch is not unique to Kipare. Hirschberg and Ward (1992) show that for the L*+H L H% (using Pierrehumbert's notation, elsewhere referred to as rise-fall-rise) contour in English, the primary factor distinguishing an incredulity interpretation from an uncertainty interpretation is pitch range. Jun and Oh (1994) show that in Korean, although not all speakers use the same strategy for expressing incredulity, larger pitch range is one of the factors manipulated to distinguish between incredulity questions and wh-questions. Miura and Hara (in press) show that in Osaka Japanese, sentence-initial F0 lowering and sentence-final F0 raising are some of the characteristics of rhetorical questions, which express disbelief or incredulity.

Not only does the shape of the pitch trace differ in sentences expressing incredulity from declaratives, but also the pitch range is greatly expanded.

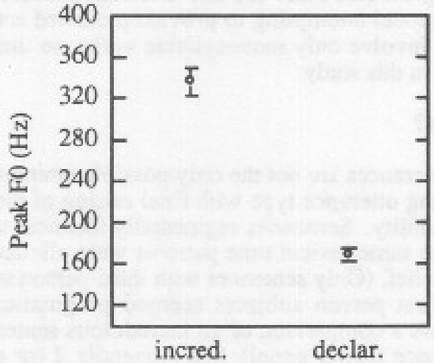



Figure 6. The mean and standard deviations for the peak F0 values of 10 tokens of each utterance type, measured at the highest point in each sentence.

The overall pitch-range expansion evident in incredulous sentences in Kipare corresponds to the "global raising" described in Inkelas and Leben. They dismiss this as "phonetic implementation." This again raises the question of the necessity of an intermediate level of representation if a phonetic interpretation can be given directly to the string of tones. And although the expansion of pitch range does provide a contrast between utterance types in Kipare, the effect may be more continuous than categorical in nature. That is, there may be gradient degrees of pitch range expansion expressing gradient degrees of incredulity. This would accord with Inkelas and Leben's depiction of pitch range expansion as more "phonetic" in nature. Even phenomena relegated to phonetic implementation should be explicable in the theory, though. Pierrehumbert (1980) assumes that "pitch range does not exist in the model except as it is represented in the value of particular tones" (p. 105). Ayers (1994), in constructing hierarchical pitch trees to capture relationships between phrases with increased and decreased pitch range, assumes no a priori categories of pitch range such as high, mid, and low. Even Ladd (1994), while arguing against free gradient variability and for constant relative F0, nevertheless assumes that "the overall modifications of pitch range are gradient." Again, further empirical data would be needed to settle the question of the representation of pitch range in Kipare.

A third utterance type, with the same shape as the final-lowered declarative sentences but with the expanded pitch range seen in the incredulous sentences, expresses yes/no questions. Such sentences would actually elicit a "yes" or "no" response from the hearer. All three types are shown in Figures 7 and 8.

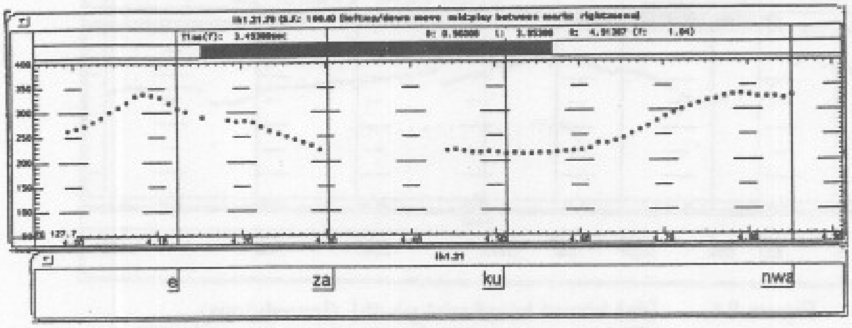



Figure 7A. [ézà kùnwá] (incredulous)
"He/ she comes to drink?!"

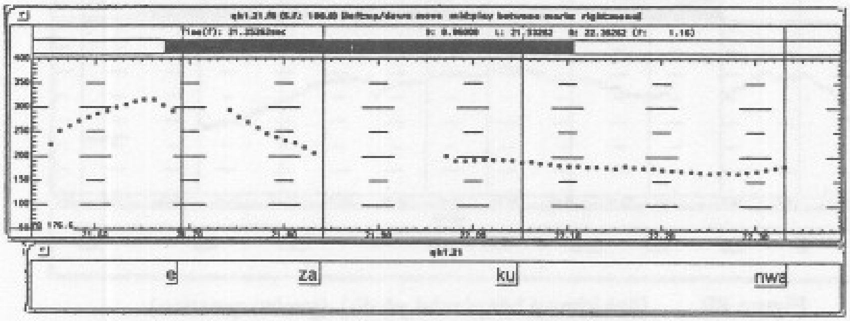



Figure 7B. [ézà kùnwá] (yes/no question)
"Does he/she come to drink?"

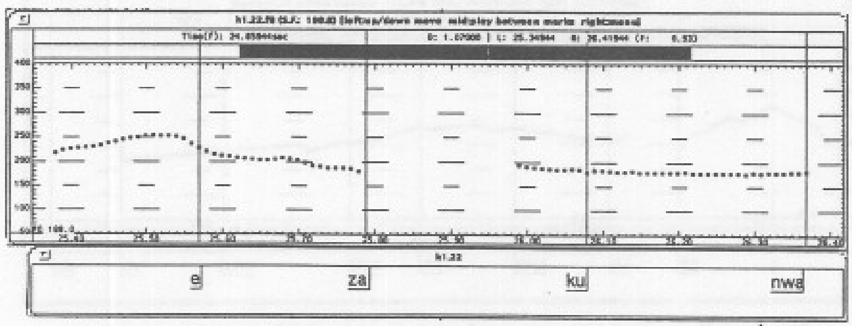



Figure 7C. [ézà kùnwá] (declarative)
"He/she comes to drink."

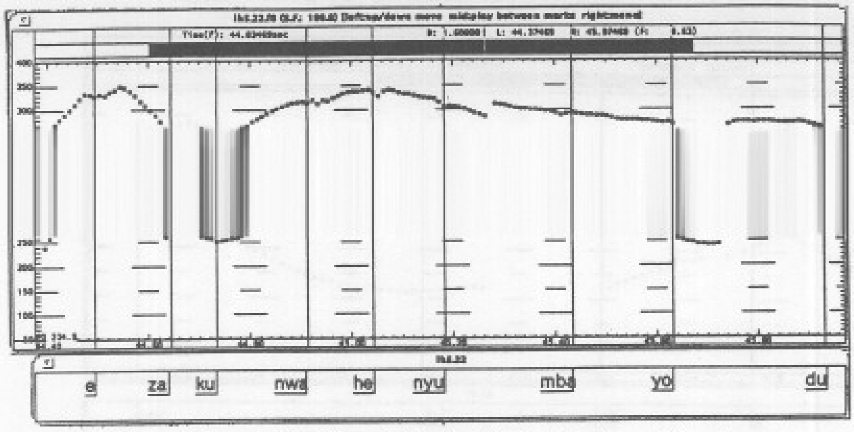



Figure 8A. [ézà kùnwá hényùmbá yò dù] (incredulous)
"He/she comes to drink in the house only today?!"

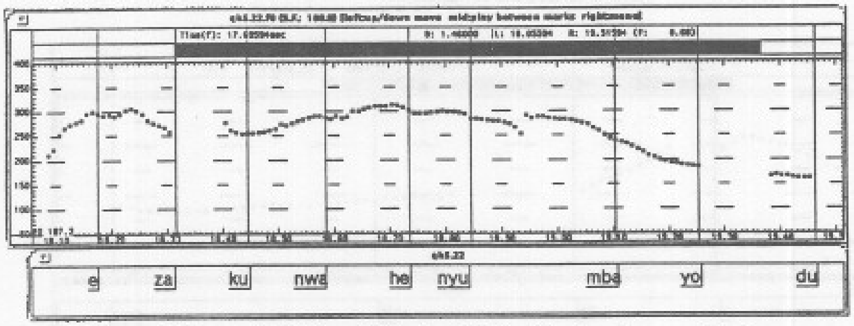



Figure 8B. [ézà kùnwá hényùmbá yò dù] (yes/no question)
"Does he/she come to drink in the house only today?"

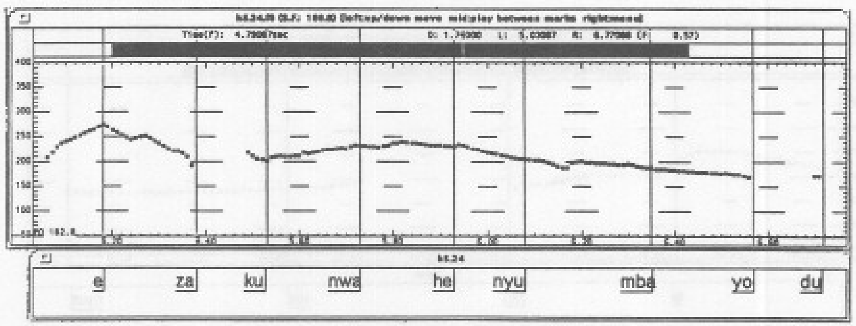



Figure 8C. [ézà kùnwá hényùmbá yò dù] (declarative)
"He/she comes to drink in the house only today."

Again, using these three types of tunes is not unique to Kipare. Shen (1990) describes three tunes of Mandarin Chinese:

"Tune I: starting with a mid key, moving upward to a mid-high key at the highest peak, falling to a low register at the ending point.
Tune II: starting with a mid-high key, moving upward to a high key at the highest peak, dropping, but not too low, ending in the high or mid-high register.
Tune III: starting with a mid-high key, moving upward to a high key at the highest peak, stepping down and ending with a low key." (page 26)

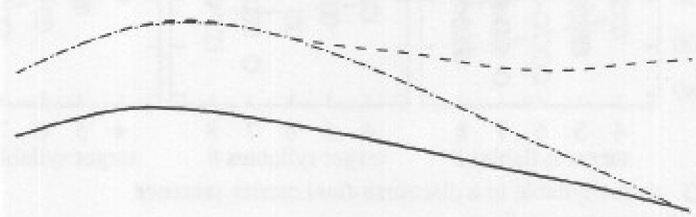


Figure 9. Shen's figure 2.5, showing the three tunes of Mandarin Chinese (used with permission of the author)

These three tunes in Mandarin involve the same combinations of elements as the three tunes in Kipare: non-expanded pitch range with final lowering, expanded pitch range with final lowering, or expanded pitch range with final raising. In Mandarin too, according to Shen, the three tunes mark different utterance types. "Tune I for assertive, Tune II for question ending in a high register, and Tune III for interrogative ending in a low register" (page 27).

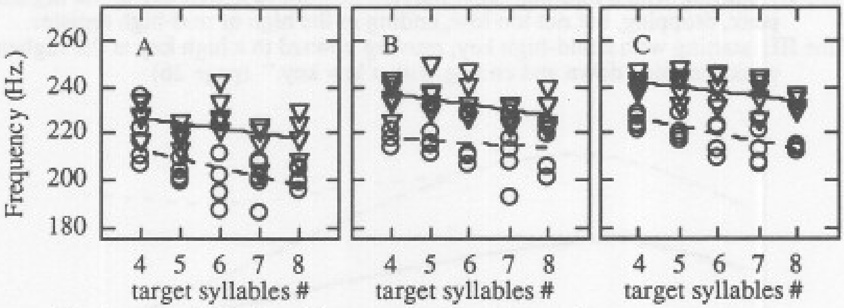
Furthermore, Shen cites Hermann's (1942) observation that high pitch is used for interrogatives in many languages, with the understanding that high pitch can be "not only a rising terminal but can also mean a relatively high overall pitch level" (page 12). The evidence from Kipare, where expanded pitch range signals yes/no questions, provides further confirmation of this observation.

So the use of various tunes to signal different utterance types is a common phenomenon cross-linguistically, in both lexical tone languages and in other languages.

Regarding the issue of final lowering, the contrast in Kipare between declarative sentences and yes/no questions on the one hand and sentences expressing incredulity on the other (manifested with rising vs. falling pitch utterance finally) is a contrast between utterance types rather than a lexical contrast. So it does not seem appropriate to analyze the contrast as an effect of a contrast in lexical tones. Instead of a tone or tones associated with an individual tone-bearing-unit, as implied by the register model, final lowering and raising may be viewed as the effect of a tone associated with the entire utterance and thus having scope over the whole sentence, as implied by the boundary tone model.

### 7. What is the domain of final lowering?

In section 3 through section 6, each sentence was produced in isolation. So the "final" lowering observed was assumed to be sentence-final lowering. However, Silverman (1987) suggests that final lowering is a property of the end of a prosodic paragraph rather than the end of a phrase. In order to explore the domain of final lowering in Kipare, the sentence containing the target syllable was followed by another sentence. See appendix 3 for the list of sentences used in this part of the study.




○ target syllable in a discourse-final carrier sentence (dashed regression line)

▽ target syllable in a carrier sentence followed by 1 sentence (solid regression line)

Figure 10.
A: The target syllable is utterance-final.
B: The target syllable is followed by 3 syllables.
C: The target syllable is followed by 4 syllables.

There were 25 tokens spoken in a sentence in isolation and 25 tokens spoken in a sentence followed by another sentence for each of A, B, and C in (18).

As shown in figure 10, the fundamental frequency values of the target syllables were lower in discourse final position than in sentence final (but not discourse final) position. This is indicated by the difference between the two regression lines in each graph, where the lower one is for sentences spoken in isolation and the lower one is for sentences followed by another sentence. Hirschberg and Pierrehumbert (1986) suggest that a decrease in final-lowering effects can indicate the absence of a boundary between discourse segments. Grosz and Hirschberg (1992) also note the importance of pitch range in signaling topic structure. Since the carrier sentence in this corpus was designed to be conceptually continuous with the sentence containing the target syllable, it could be that the decrease in final lowering effects can be attributed to marking the semantic or conceptual connection between the two sentences.

Thus Silverman's suggestions about the domain of final lowering are borne out in Kipare. The domain of final lowering seems to be bigger than just the sentence, although further study would be needed to determine exactly what the extent of the domain is. No matter what the exact domain for final lowering, the fact remains that either the register model or the boundary tone model must be enriched so as to be able to refer to more than just the sentence level in describing final lowering.

## 8. Conclusion

This study compares the register model with the boundary tone model in analyzing final lowering in Kipare. Several factors have been examined. All tonal categories lower, but do not coincide with the next category down. This does not accord with the prediction of the register model. The final tones do seem to be a constant proportion of preceding tones, which does match the prediction of the boundary tone model. The effects of final lowering are non-local, affecting more than just the extreme final tone-bearing unit. This does not match the predictions of the register model, but it does match the predictions of the boundary tone model. Contrasts exist at the utterance level. This would be best handled by the boundary tone model, since the boundary tone has scope over the entire utterance. The domain of final lowering is greater than the sentence, which cannot be handled by either model. Overall, these factors are suggestive of the effects of a boundary tone having scope over the entire utterance rather than the effects of a register tone associated with a single tone-bearing unit.

## 9. Works Cited

Ayers, Gayle M. (1994). Discourse Function of Pitch Range in Spontaneous and Read Speech. Ohio State University Working Papers in Linguistics. 44: 1-49.

Beckman, Mary and Janet Pierrehumbert. (1992). Comments on Chapters 14 and 15. Papers in Laboratory Phonology II: Gesture, Segment, Prosody. G.J. Docherty and D.R. Ladd, eds. Cambridge University Press: Cambridge. 387-397.

Clements, George N. (1981). The Hierarchical Representation of Tone Features. Harvard Studies in Phonology, vol. II.

Clements, G.N. and K.C. Ford. (1981). On the Phonological Status of Downstep in Kikuyu. Phonology in the 1980s. D. L. Goyvaerts, ed. 309-358.

Elugbe, Ben Ohi. (1977). Some Implications of Low Tone Raising in Southwestern Edo. Studies in African Linguistics, supp.7. 53-62.

Entropic Research Laboratory, Inc. (1993). *Waves+ 5.0.* AT&T Bell Laboratories. Washington, D.C.

Grosz, Barbara and Julia Hirschberg. (1992). Some Intonational Characteristics of Discourse Structure. ICSLP 92 Proceedings. University of Alberta. 429-432.

Gussenhoven, C. and A.C.M. Rietveld. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. Journal of Phonetics. 16: 355-369.

Hermann, Eduard. (1942). Problem der Frage. Nachrichten v.d. Akademie der Wissenschaft in Göttingen, Philologische Historische Kasse, 121-408. (Cited in Shen, 1990).

Hirschberg, Julia and Janet Pierrehumbert. (1986). The Intonational Structuring of Discourse. Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics. Columbia University, New York. 136-144.

Hirschberg, Julia and Gregory Ward (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. Journal of Phonetics. 20:241-251.

Inkelas, Sharon and William R. Leben. (1990) Where phonology and phonetics interact: the case of Hausa intonation. Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. J. Kingston and M. Beckman, eds. Cambridge University Press: Cambridge. 17-34.

Inkelas, Sharon, William R. Leben, and Mark Cobler. (1986). The Phonology of Intonation in Hausa. Proceedings of NELS 17.

Jun, Sun-Ah and Mira Oh. (1994). A Prosodic Analysis of Three Sentence Types With "Wh" Words in Korean. Proceedings of the 1994 International

Conference on Spoken Language Processing. The Acoustical Society of Japan. 323-326.
Ladd, D. Robert. (1994). Constraints on the gradient variability of pitch range, or, Pitch Level 4 Lives! Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form. P. Keating, ed. Cambridge University Press: Cambridge. 43-63.
Ladefoged, Peter. (1982). A Couse in Phonetics, second edition. Harcourt Brace Jovanovich, Inc.
Laniran, Yetunde Olabisi. (1992). Intonation in Tone Languages: the Phonetic Implementation of Tones in Yoruba. PhD Dissertation. Cornell University.
Liberman, Mark and Janet Pierrehumbert. (1984). Intonational Invariance Under Changes in Pitch Range and Length. Language Sound Structure. M. Aronoff and R. Oehrle, eds. MIT Press: Cambridge, MA.
Miura, Ichiro and Noriyo Hara. (in press). Production and Perception of Rhetoricl Questions in Osaka Japanese. To appear in Journal of Phonetics.
Odden, David. (1986a) On the Role of the Obligatory Contour Principle in Phonological Theory. Language, 62:2. 353-383.
Odden, David. (1986b) Three dialects of Kipare. Current Approaches to African Linguistics, 3. 257-280.
Pierrehumbert, Janet B. (1980). The Phonology and Phonetics of English Intonation. PhD Dissertation. MIT.
Pierrehumbert, Janet B. and Mary E. Beckman. (1988) Japanese Tone Structure. Linguistic Inquiry Monograph 15. MIT Press.
Roberts, R. Ruth. (1992). A Non-Metrical Theory of Sukuma Tone. Ohio State University Working Papers in Linguistics. 41:135-148.
Shen, Susan. (1990). The Prosody of Mandarin Chinese. University of California Press.
Silverman, Kim. (1987). The Structure and Processing of Fundamental Frequency Contours. PhD Dissertation. University of Cambridge.
Snider, Keith and Harry van der Hulst. (1993) Issues in the Representation of Tonal Register. in The Phonology of Tone; the Representation of Tonal Register. Harry van der Hulst and Keith Snider, eds. Mouton de Gruyter: Berlin. 1-27.
Thorsen, Nina. (1985). Intonation and text in Standard Danish. Journal of the Acoustical Society of America. 77: 1205-1216.

**Appendix 1: Corpus of Utterances used for section 3, section 4, and section 5**

| | |
|---|---|
| [kùnwá] | "to drink" |
| [ézà kùnwá] | "he/she comes to drink" |
| [nìéndà kùnwá] | "I would like to drink" |
| [mì nìéndà kùnwá] | "(I) I would like to drink" |
| [íkì nìéndà kùnwá] | "now I want to drink" |
| [íkì mì nìéndà kùnwá] | "now (I) I want to drink" |
| | |
| [ná!nwá] | "I drank" |
| [áhà ná!nwá] | "but I drank" |
| [mì áhà ná!nwá] | "but me, I drank" |
| [yò mì áhà ná!nwá] | "today I have drunk" |
| [íkì mì áhà ná!nwá] | "but I have drunk" |
| [íkì yò mì áhà ná!nwá] | "but today I have drunk" |
| | |
| [kùgwà] | "to fall" |

| | |
|---|---|
| [íkì kùgwà] | "to fall now" |
| [nìéndà kùgwà] | "I would like to fall" |
| [mì nìéndá kùgwà] | "(I) I would like to fall" |
| [íkì mì nìéndà kùgwà] | "now (I) I want to fall" |
| | |
| [kùnwá íkì] | "to drink now" |
| [ézà kùnwá íkì] | "he/she comes to drink now" |
| [nìéndà kùnwá íkì] | "I would like to drink now" |
| [mì nìéndà kùnwá íkì] | "(I) I would like to drink now" |
| [íkì nìéndà kùnwá íkì] | "now I want to drink now" |
| [íkì mì nìéndà kùnwá íkì] | "now (I) I want to drink now" |
| | |
| [ná!nwá íkì] | "I drank now" |
| [áhà ná!nwá íkì] | "but I drank now" |
| [mì áhà ná!nwá íkì] | "but me, I drank now" |
| [yò mì áhà ná!nwá íkì] | "today I have drunk now" |
| [íkì mì áhà ná!nwá íkì] | "but I have drunk now" |
| [íkì yò mì áhà ná!nwá íkì] | "but today I have drunk now" |
| | |
| [kùgwà íkì] | "to fall now" |
| [íkì kùgwà íkì] | "(now) to fall now" |
| [nìéndà kùgwà íkì] | "I would like to fall now" |
| [mì nìéndá kùgwà íkì] | "(I) I would like to fall now" |
| [íkì mì nìéndà kùgwà íkì] | "now (I) I want to fall now" |
| | |
| [kùnwá hényùmbá]] | "to drink in the house" |
| [ézà kùnwá hényùmbá] | "he/she comes to drink in the house" |
| [nìéndà kùnwá hényùmbá] | "I would like to drink in the house" |
| [mì nìéndà kùnwá hényùmbá] | "(I) I would like to drink in the house" |
| [íkì nìéndà kùnwá hényùmbá] | "now I want to drink in the house" |
| [íkì mì nìéndà kùnwá hényùmbá] | "now (I) I want to drink in the house" |
| | |
| [ná!nwá hényùmbá] | "I drank in the house" |
| [áhà ná!nwá hényùmbá] | "but I drank in the house" |
| [mì áhà ná!nwá hényùmbá] | "but me, I drank in the house" |
| [yò mì áhà ná!nwá hényùmbá] | "today I have drunk in the house" |
| [íkì mì áhà ná!nwá hényùmbá] | "but I have drunk in the house" |
| [íkì yò mì áhà ná!nwá hényùmbá] | "but today I have drunk in the house" |
| | |
| [kùgwà hényùmbá] | "to fall in the house" |
| [íkì kùgwà hényùmbá] | "to fall in the house now" |
| [nìéndà kùgwà hényùmbá] | "I would like to fall in the house" |
| [mì nìéndá kùgwà hényùmbá] | "(I) I would like to fall in the house" |
| [íkì mì nìéndà kùgwà hényùmbá] | "now (I) I want to fall in the house" |
| | |
| [kùnwá hényùmbá yò] | "to drink in the house today" |
| [ézà kùnwá hényùmbá yò] | "he/she comes to drink in the housetoday" |
| [nìéndà kùnwá hényùmbá yò] | "I would like to drink in the house today" |
| [mì nìéndà kùnwá hényùmbá yò] | "(I) I would like to drink in the house today" |
| [íkì nìéndà kùnwá hényùmbá yò] | "now I want to drink in the house today" |
| [íkì mì nìéndà kùnwá hényùmbá yò] | "now (I) I want to drink in the house today" |
| | |
| [ná!nwá hényùmbá yò] | "I drank in the house today" |
| [áhà ná!nwá hényùmbá yò] | "but I drank in the house today" |

| | |
|---|---|
| [mì áhà ná!nwá hényùmbá yò] | "but me, I drank in the house today" |
| [yò mì áhà ná!nwá hényùmbá yò] | "today I have drunk in the house today" |
| [íkì mì áhà ná!nwá hényùmbá yò] | "but I have drunk in the house today" |
| [íkì yò mì áhà ná!nwá hényùmbá yò] | "but today I have drunk in the house today" |

| | |
|---|---|
| [kùgwà hényùmbá yò] | "to fall in the house today" |
| [íkì kùgwà hényùmbá yò] | "(now) to fall in the house today" |
| [nìéndà kùgwà hényùmbá yò] | "I would like to fall in the house today" |
| [mì nìéndá kùgwà hényùmbá yò] | "(I) I would like to fall in the house today" |
| [íkì mì nìéndà kùgwà hényùmbá yò] | "now (I) I want to fall in the house today" |

| | |
|---|---|
| [kùnwá hényùmbá yò dù] | "to drink only today" |
| [ézà kùnwá hényùmbá yò dù] | "he/she comes to drink only today" |
| [nìéndà kùnwá hényùmbá yò dù] | "I would like to drink only today" |
| [mì nìéndà kùnwá hényùmbá yò dù] | "(I) I would like to drink only today" |
| [íkì nìéndà kùnwá hényùmbá yò dù] | "now I want to drink only today" |
| [íkì mì nìéndà kùnwá hényùmbá yò dù] | "now (I) I want to drink only today" |

| | |
|---|---|
| [ná!nwá hényùmbá yò dù] | "I drank only today" |
| [áhà ná!nwá hényùmbá yò dù] | "but I drank only today" |
| [mì áhà ná!nwá hényùmbá yò dù] | "but me, I drank only today" |
| [yò mì áhà ná!nwá hényùmbá yò dù] | "today I have drunk only today" |
| [íkì mì áhà ná!nwá hényùmbá yò dù] | "but I have drunk only today" |
| [íkì yò mì áhà ná!nwá hényùmbá yò dù] | "but today I have drunk only today" |

| | |
|---|---|
| [kùgwà hényùmbá yò dù] | "to fall only today" |
| [íkì kùgwà hényùmbá yò dù] | "(now) to fall only today" |
| [nìéndà kùgwà hényùmbá yò dù] | "I would like to fall only today" |
| [mì nìéndá kùgwà hényùmbá yò dù] | "(I) I would like to fall only today" |
| [íkì mì nìéndà kùgwà hényùmbá yò dù] | "now (I) I want to fall only today" |

## Appendix 2: Corpus of utterances used for section 6

| | |
|---|---|
| [ézà kùnwá] | "he/she comes to drink" |
| [ézà kùnwá íkì] | "he/she comes to drink now" |
| [ézà kùnwá hényùmbá] | "he/she comes to drink in the house" |
| [ézà kùnwá hényùmbá yò] | "he/she comes to drink in the house today" |
| [ézà kùnwá hényùmbá yò dù] | "he/she comes to drink " " " only today" |

## Appendix 3: Corpus of utterances used for section 7

The sentence following the carrier sentence containing the target syllable was:
[úkò nzé hénà kìrúkè] "it is hot outside."

The target sentences used were:

| | |
|---|---|
| [ézà kùnwá] | "he/she comes to drink" |
| [nìéndà kùnwá] | "I would like to drink" |
| [mì nìéndà kùnwá] | "(I) I would like to drink" |
| [íkì nìéndà kùnwá] | "now I want to drink" |
| [íkì mì nìéndà kùnwá] | "now (I) I want to drink" |

| | |
|---|---|
| [ézà kùnwá hényùmbá] | "he/she comes to drink in the house" |
| [nìéndà kùnwá hényùmbá] | "I would like to drink in the house" |

[mì nìéndà kùnwá hényùmbá] "(I) I would like to drink in the house"
[íkì nìéndà kùnwá hényùmbá] "now I want to drink in the house"
[íkì mì nìéndà kùnwá hényùmbá] "now (I) I want to drink in the house"

[ézà kùnwá hényùmbá yò] "he/she comes to drink in the house today"
[nìéndà kùnwá hényùmbá yò] "I would like to drink in the house today"
[mì nìéndà kùnwá hényùmbá yò] "(I) I would like to drink in the house today"
[íkì nìéndà kùnwá hényùmbá yò] "now I want to drink in the house today"
[íkì mì nìéndà kùnwá hényùmbá yò] "now (I) I want to drink in the house today"

*OSU Working Papers in Linguistics* 45 (56-84)

## Gestural Phasing as an Explanation for Vowel Devoicing in Turkish*

Stefanie Jannedy

jannedy@ling.ohio-state.edu

**Abstract**: Recent work in phonetics has suggested that vowel devoicing or schwa deletion, observed in various languages, is a gradient process. This study provides evidence for the previously undocumented process of high vowel devoicing in Turkish. The prosodic and segmental factors *rate, stress, preceding environment, following enviornment, vowel type, and syllable type* were investigated. The factors are described, evaluated and ranked according to the results of a multiple regression (Variable Rule) analysis. Where applicable, results are contrasted with findings for i.e., Japanese and Korean. Furthermore, VOT (voice onset time) measurements of the three voiceless stops [p t k] were obtained, as well as duration measurements of vowels in open and closed syllables where vowels are significantly longer in Turkish. Generally, most devoicing occurred when the vowel was shorter (i.e., as a result of faster rates of speech, lack of stress, in closed syllables, ect.). These findings accord well with predictions made by a model assuming gradual gestural overlap of adjacent consonantal and vocalic gestures. It will be attempted to explain the findings with differences in phasing between articulatory gestures.

### I. Introduction

In Turkish a syllable containing any of the four high vowels [i y ɨ u] can be realized without any audible traces of voicing. The phenomenon is demonstrated in Figure 1, which shows a contrasts between two words produced by the same speaker, one containing a fully realized vowel, the other containing a fully devoiced vowel. As becomes clear from these spectrograms, the endpoint of a continuum of vowel devoicing can be interpreted as vowel deletion. On the left we see a spectrogram and waveform of the word *tüfek* 'gun, rifle' spoken in a slow rate of speech. The vertical striations at the bottom of the spectrogram are the individual glottal pulses, showing that this first vowel is voiced. The presence of the vowel is also reflected in the waveform. The spectrogram on the right shows the same word produced in a normal rate of speech. Here, the vowel has completely disappeared, there are no voicing traces left so that this vowel is analyzed as completely devoiced. This phenomenon is previously undocumented for Turkish, but resembles a process noted for several other languages, including Svabian (Griffen, 1983), Canadian French (Cedergren & Simeneau, 1985; Cedergren, 1986), Korean (Jun & Beckman, 1993, 1994) and Japanese (McCawley, 1968; Jun & Beckman, 1993, 1994).

* Versions of this paper were presented at the 168th meeting of the Acoustical Society of America in Austin, Texas and the Fifth Australian Conference on Speech Science and Technology in Perth, Australia. I would like to thank Mary Beckman, Marios Fourakis, Keith Johnson, Brian Joseph, Sun-Ah Jun, No-Ju Kim, Joyce McDonough, Jennifer Venditti, participants at these two conferences and members of the Phonetics-Phonology reading group at the Ohio State University for comments and suggestions.

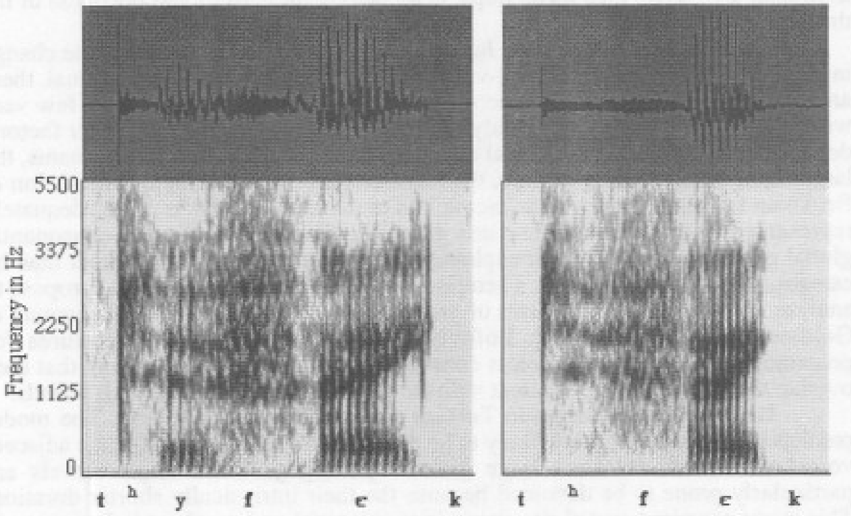



Figure 1: Spectrograms and waveforms of completely voiced and completely devoiced /y/ in [tyfek] 'gun, rifle'

Vowel devoicing has been discussed most thoroughly for Japanese, where it has traditionally been described in terms of a categorical feature changing rule. McCawley (1968:127), for example gives the following SPE-type rule by which the high vowels /i/ and /u/ in Japanese become devoiced between voiceless consonants or wordfinally after voiceless consonants:

$$(1) \quad \begin{bmatrix} \text{-cns} \\ \text{+voc} \\ \text{+dif} \end{bmatrix} \longrightarrow \text{[-voi] in env.} \begin{bmatrix} \text{[-voi]} \ \_\_\_\_ \ \text{[-voi]} \\ \_\_\_\_ \ \# \end{bmatrix}$$

However, there are reasons to suspect that such categorical phonological rules are inadequate to describe vowel devoicing in Japanese and other languages also previously studied for the phenomenon. Beckman & Shoji (1984) showed that in Japanese the initial syllable in the minimal pair /ʃikaN/ and /ʃyukaN/ is not completely neutralized in production. Rather, the fricative spectrum of the preceding "esh" [ʃ__] retains information about the vowel's quality to various degrees so that due to coarticulation, a contrast is preserved. In cases where no coarticulatory information is preserved in the consonant-vowel transitions, the end of the continuum of vowel devoicing can be interpreted as vowel deletion. The gradient nature of this effect is shown in production and confirmed by results of a perception experiment where listeners' identification responses show a high correlation with the amount of vocalic coloring of the fricative.

Jun & Beckman (1993) studied the behavior of the three high vowels [i ɨ u] of Korean, occurring as the first vowel in CVCV syllables with combinations of voiceless aspirated, lenis and fortis stops preceding and following the initial high vowel in the first open syllable. Duration measurement of the preceding aspirated and lenis stop consonants plotted against the duration of the vowel show that vowels preceded by aspirated consonants are generally shorter than vowels preceded by lenis stops. The authors also provide data that shows that the amount of completely and partially devoiced vowels was highest when it was preceded by aspirated, then lenis, then fortis stops, in ascending order of glottal openness of the three Korean stop types.

These previous studies on Japanese and Korean have shown that the change in specification from [+voice] to [-voice] is not categorical but rather gradual, there are intermediate stages found where a partially voiced vowel has only a few very weak glottal pulses. The probability of devoicing a vowel is, among other factors, dependent on the size of the glottal opening gesture of neighboring consonants, the larger the glottal opening gesture, the more frequently devoicing occurred. Jun & Beckman suggest that vowel devoicing in Korean and Japanese is more adequately represented in terms of more or less glottal overlap of the adjacent consonantal glottal gestures and thus, better explained in terms of gestural hiding rather than by categorical phonological rules. Therefore, Jun & Beckman (1993, 1994) propose to analyze vowel devoicing in terms of gradual glottal gestural overlap (Browman & Goldstein, 1990; Munhall & Löfqvist, 1992) where the glottal gestures for preceding and following voiceless consonants are phased in such a way that they overlap to a greater or lesser extent with the voicing gestures for the high vowels.

High vowel devoicing in Turkish can be explained similarly. The model predicts that vowels are more likely to be devoiced if they are short and the adjacent voiceless consonants have large glottal opening gestures. High vowels are particularly prone to be devoiced because the their intrinsically shorter duration. This study examines vowel devoicing in word initial and word medial position in Turkish and provides evidence for this previously undocumented process. The prosodic and segmental factors *rate, stress, preceding environment, following environment, vowel-* and *syllable type* are considered in the current study. The factors are described, evaluated, and ranked according to the results of a multiple regression (VARBRUL) analysis. Where applicable, results are contrasted with findings for other languages such as Montreal French, Korean or Japanese. Furthermore, VOT (voice onset time) measurements of the three voiceless stops [p t k] were obtained, as well as duration measurements of non-high vowels in open and closed syllables before geminates as well as heterorganic consonant clusters and single consonants.

## I.2 Motivation for Duration Experiments in Turkish

Since not many instrumental phonetic studies of Turkish have been performed, there is a lack of basic knowledge of durational facts relevant to devoicing in Turkish. The Korean data suggests that the stronger the glottal gesture the more devoicing is triggered. Thus, basic durational facts such as the duration of the accompanying aspiration of voiceless stop consonants need to be established for Turkish so that the impact of stops can be evaluated and predicted better. For example, Turkish contrasts voiced and voiceless stops (Kornfilt, 1986, 1987; Lees, 1961; Underhill, 1986) and thus, a potentially important factor is the duration of VOT (voice onset time) which is a measure of the lag or delay of voicing onset of the following voiced segment. In order to explain patterns of devoicing as a function of the preceding environment, we need to know more about the duration of aspiration accompanying the release of voiceless stops. The larger and longer the glottal opening gesture is, the greater is the potential for the consonantal glottal gesture to extend into the vowel's glottal gesture and delay or prevent it's onset of voicing. Thus, in experiment one, the duration of VOT of word initial voiceless stops [p t k] before non-high vowels (appendix A) was measured to establish voice onset time measures for Turkish in a fairly independent and unaffected context to vowel devoicing.

Secondly, durational information about vowels in closed and open syllables needs to be obtained. Generally, the assumption holds that the shorter the vowel, the greater the probability for the vowel to become devoiced. Maddieson (1985) summarized previous research on a variety of languages that have shorter vowels in closed versus open syllables, among these languages are English, Russian, Finnish, Korean, Chinese, and Thai. Maddieson (1985) also cites studies which established vowel duration to be shorter before geminate consonants (closed syllables) compared to singletons (open syllables), among these languages are Hausa, Italian, Norwegian, Finnish, Arabic, Bengali, Kannada and Amharic. These studies suggest that crosslinguistically, vowels are shorter in closed compared to open syllables. Lahiri and Hankamer (1988) measured vowel duration in Bengali and Turkish before geminates and single consonants and found an overall marginally significant effect for vowels to be shorter in closed syllables before geminates in Bengali, but did not find a significant effect for Turkish. In fact, in Turkish, mean duration of vowels was contrary to the predictions, slightly longer in closed syllables. Thus, according to their results, Turkish seems to be somewhat unusual in that vowel duration in open and closed syllables does not significantly differ but shows a small difference in the opposite direction. Han (1994) however reports vowels to be significantly longer (11% ~ 11ms) before geminate stop consonants in Japanese when running a 'simple binomial probability test ($p < 0.001$)', but not when doing a 'one-tailed difference *t* test ($t(9)=.054$, $p>0.05$)'. Maddieson (1985) refutes Japanese to be an apparent counterexample to what he calls *closed syllable vowel shortening*. He argues that in Japanese which is assumed to be organized temporally on the bases of the mora (Bloch, 1950; Han, 1994) the first part of the geminate does not close the preceding syllable but constitutes a mora by itself, leaving the preceding vowel in an open syllable.

A significant effect for vowel duration differences for open versus closed syllables with longer vowels in closed syllables in a language without a moraic temporal organization could show that vowel duration in closed syllables is either language specific (and thus, not a universal feature of language) or possibly an effect of the following consonantal environment. Three conceivable options to test are: vowels are generally longer before a) single consonants (CV.C) in open syllables versus b) geminates, one consonant belonging to the coda of the first syllable and the second consonant belonging to the onset of the second syllable ($CVC_i.C_i$) versus c) non-geminate heterorganic consonant clusters ($CVC_i.C_j$). In experiment two, duration of non-high vowels in open and closed syllables in disyllabic words (appendix A) was measured. Non-high vowels were chosen since they are more resistant (due to their intrinsically longer duration) of the dependent test variable (vowel devoicing). The corpus contained 'minimal pairs' that contrasted between a singleton ($VC_1$) and a consonant cluster containing that singleton immediately following the vowel ($VC_1.C_2$) in question. If longer vowel durations in closed syllables before heterorganic consonant clusters will be found then an explanation assuming vowel elongation only before geminates must be rejected and longer vowel durations in closed syllables are just another language specific factor a language chooses. As a control, words contrasting in having single consonants (and thus open syllables) CV•C and geminates $CVC_i{\bullet}C_i$ (closed syllables) used by Lahiri and Hankamer (1988) (see appendix B) were also recorded and vowel duration was measured.

In a third experiment, words with high vowels occurring in various prosodic and segmental contexts were elicited in three different speech rates. Based on categorization criteria similar to those used in Jun & Beckman (1994), the experimenter judged whether the vowel of interest was voiced, voiceless or partially voiced. The judgments were considered in a variable rule (VARBRUL) statistical analysis and conditioning factors were ranked according to the magnitude of their contribution to the process of devoicing.

Jun & Beckman (1993, 1994) suggest to interpret vowel devoicing in Japanese and Korean in terms of a gradual gestural overlap model (Browman & Goldstein, 1990) where the laryngeal gestures of the adjacent consonants to the left and right of the vowel overlap with the high vowel's glottal gesture.This suggestion will be taken up again and applied to the Turkish data. Generally, the gestural overlap model predicts that factors shortening the duration of vowels (i.e., a faster overall rate of speech) should increase the probability for vowel devoicing and factors that lengthen syllable duration (i.e., stress) should decrease the likelihood of devoicing.

## II. Methods

### 1. The Duration Experiments

Two sets of data were used for the duration measurements: one set of 18 'minimal-' and 'near-' pairs (36 words) previously used by Lahiri and Hankamer (1988), contrasting intervocalic single consonants and geminates, displayed in appendix B, and a second set of 28 (56 words) 'minimal' and 'near-minimal' pairs selected for this study, displayed in appendix A.

Lahiri and Hankamer (1988) elicited these 18 pairs of words (illustrating the difference between single consonants and geminates), in citation in a normal rate of speech. They measured VOT of word medial singleton stops and geminates as well as vowel duration in open and closed syllables in Turkish. Mean VOT was 34 ms in closed and 45 ms in open syllables and significantly different. None of the 18 pairs of words contained a [p], thus, the mean VOT-values reported in Lahiri and Hankamer's study do not encompass all three voiceless stops of Turkish. Mean vowel duration in closed syllables (116 ms) was insignificantly longer than in open syllables (112 ms).

The 28 pairs of disyllabic words (fifteen disyllabic minimal and thirteen disyllabic near minimal pairs) selected for this study, contained non-high vowels and with a contrast in syllable type (open versus closed). The syllable type was confirmed by two native speakers. The non-high vowels [a e o] were preceded by all three types of voiceless stop consonants [p t k]. One pair contained a geminate, the rest contrasted intervocalic clusters and singletons.

Each word was presented on an index card and elicited in citation form in a normal rate of speech (three repetitions in different randomized orders) as well as in three different self selected speech rates (slow, normal, fast) absolute utterance initially in carrier phrases. All three carrier-phrases were presented on a single index card. (Although in this experiment we are only marginally interested in the effect of rate on the non-high vowel's duration or the VOT, durations for all three speech rates were measured). The words were embedded in the following carrier phrases:

1. "____" kelimesini yavaş şekilde söyle. The word "___" I say in a slow mode.
   word slow mode say

2. "____" kelimesini normal hizla söyle. The word "___" I say in a normal speed.
   *word normal speed say.*

3. "____" kelimesini hizli şekilde söyle. The word "___" I say in a fast mode.
   *word fast mode say.*

## 1.1 VOT (Voice Onset Time)

Five educated male native speakers of Istanbul Turkish read the words in the above described conditions. VOT for Turkish syllable initial voiceless stops [p t k] before unstressed non-high vowels in the initial syllable (appendix A) was measured from the release burst of the stop to the onset of voicing of the following vowel visible as a voice bar on the spectrogram. In the repetition of Lahiri and Hankamer's 1988 experiment (appendix B), VOT for the word medial voiceless stops [t k] was measured from waveforms from the release of the burst of the voiceless stop closure of the geminate or single consonant to the onset of voicing of the following vowel. Since Turkish words are generally stressed on the final syllable (Lees, 1961; Underhill, 1986; van der Hulst & van de Weijer, 1991), VOT in these cases was measured before stressed vowels. Figure 2 shows spectrograms and waveforms for two Turkish words, exemplifying the VOT measurement criteria.

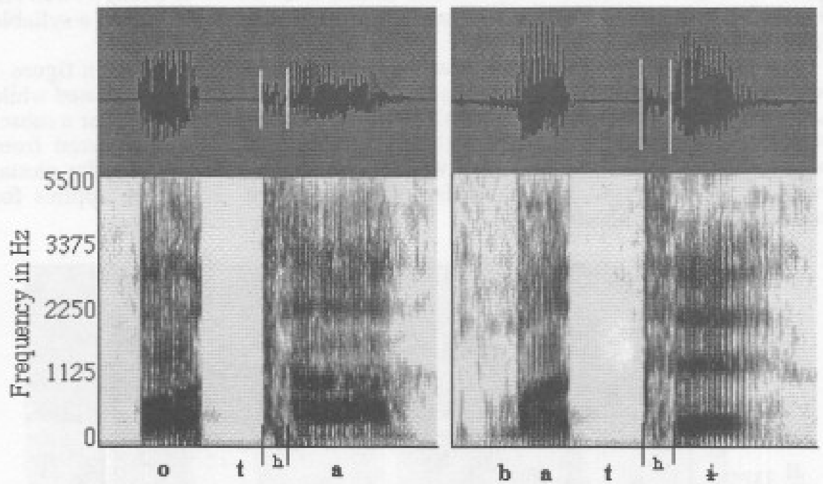



Figure 2. Spectrograms, waveforms and rms amplitude traces of the words [ota] 'grass' and [batɨ] 'west, western', demonstrating VOT measurement criteria.

For Lahiri and Hankamer's data, the total number of tokens was 180 (36 words x 5 speakers) tokens per speech rate in the phrasal condition and 540 (36 words x 5 speakers x 3 speech rates) in citation for the words where medial VOT was measured. For the words selected for this study, for the initial VOT measurements, 840 measurements were taken in the citation condition (56 words x 5 speakers x 3 speech rates) and 280 words in the phrasal condition (56 words x 5 speakers). In both corpora, several tokens had to be discarded from the study mainly because of incomplete stop closures.

## 1.2 Vowel-duration

The same 28 minimal and near minimal pairs (appendix A) contrasting open and closed initial syllables were used for the vowel duration measurements. Measurements were taken for the duration of the four non-high vowels [a e o] from the onset to the end of the vowel's formant structure. Measurements were made using wideband spectrograms generated on a KAY 5500-DSP real time sound spectrograph. 28 pairs of words contrasting open and closed syllable type generated 140 tokens (28 words x 5 speakers) that were elicited in the phrasal condition per rate per syllable type. In the citation condition, 420 tokens (28 words x 5 speakers x 3 rates) were elicited and analyzed.

For the replication of Lahiri & Hankamer's 1988 experiments (appendix B), the total number of tokens is 75 (15 words x 5 speakers) per speech rate in the phrasal condition and 225 (15 words x 5 speakers x 3 speech rates) in citation form. Three pairs of words were discarded from this set of data because of difficulties in applying consistent measurement criteria: for two pairs ([jata] - [jatta] 'Yacht' DAT and LOC; [jatɨ] 'Yacht ACC' and [jattɨ] 'lie down' PAST) no consistent segmentation landmarks could be found between the palatal glide and the low vowel; and the third pair because (contrary to Lahiri and Hankamer's assumptions) [saate] and [saatte] ('clock' DAT and LOC), are trisyllabic, with a syllable break in the middle of what Lahiri and Hankamer took as a long vowel. All five native speakers of Turkish analyzed the vowel sequence as having a syllable break in the middle.

To test the consistency of measurement criteria (demonstrated in figure 4 below) across measurement techniques (in this study spectrograms were used while Lahiri and Hankamer used waveforms), vowel duration measurements for a subset of five pairs of words (marked with a * in appendix B) was repeated from waveforms. The onset of the vowel was measured from the first regular glottal pulse to the last regular glottal pulse on a waveform. The same applies for measurements from spectrograms.

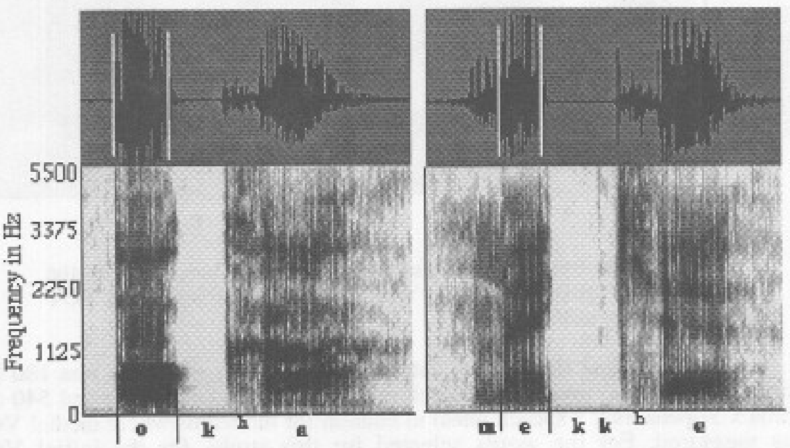



Figure 3. Spectrograms and waveforms of the words [oka] 'arrow' and [mek:e] 'Mecca', demonstrating vowel duration measurement criteria from waveforms and spectrograms.

## 2. Devoicing Experiment

Nine naive educated native speakers of Turkish (2 female, 7 male; 3 from Ankara and 6 from Istanbul) read 135 words positioned utterance initially positioned in carrier-phrases at three self-selected rates (slow, normal, fast). None of the words were monosyllables or contained [h] since the phonetic classification as fricative or approximant is not clear.

### 2.1 Devoicing Analysis

Each of the 3645 tokens (9 speakers x 135 words x 3 rates) was rated by the experimenter as containing either a voiced (clear voicebar with several glottal pulses), partially voiced (one or two faint glottal pulses) or completely devoiced vowel (no glottal pulses visible on spectrogram). Note that the dependent variable *voicing status* is continuous but by arbitrary criteria categorized into three discrete levels. The criteria for this categorization on a voicing continuum are similar to the ones used by Jun and Beckman (1994) for Korean.

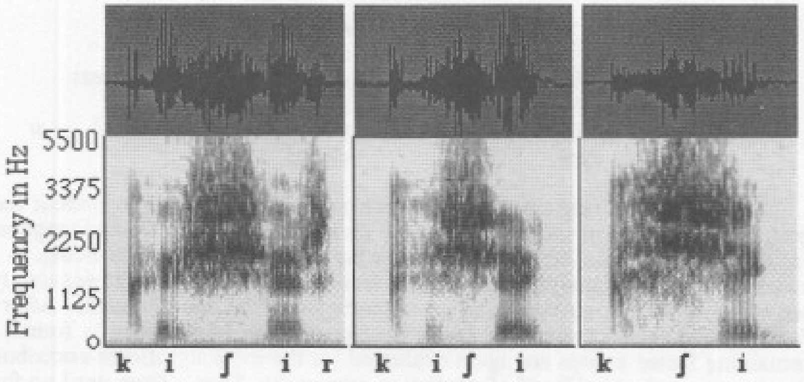



Figure 4. Spectrograms and waveforms of voiced, partially devoiced and completely devoiced vowel tokens in the word [**kiʃir**] 'crust, bark'.

### 2.2 Statistical Analysis

Varbrul (Variable Rule) analysis (Sankoff, 1988; Rand & Sankoff, 1990) was used to evaluate the relative importance to the distribution of devoicing of the different predictors *rate, stress, preceding-* and *following environment, vowel-* and *syllable type.* This analysis method uses step-wise multiple regressions on a logistic transform of the proportions of tokens which undergo a "rule" (in this case, vowel devoicing) for each combination of factors, with a maximum likelihood estimation criterion to accommodate imbalances (e.g., fewer token of words containing [y] than [i]; more unstressed syllables than stressed ones, etc.) of number of tokens within the various cells. In a logistic regression analysis, the sum of the factor effects does not equal the predicted percentage of a given choice but some quantity related to this percentage by the following formula:

(2) $\log\left(\frac{p}{1-p}\right)$ = "input"[1] + sum of factor effects

Because results are expressed as proportions, only binary oppositions for dependent factor groupings (i.e., *devoiced* vs. *voiced*) can be compared. Therefore,

[1]The "input" is the sum of all the averages that were subtracted from the different factor groups, also called *corrected mean.*

the factor *partially voiced* of the dependent variable was grouped with the *voiced token* and then compared to the *completely devoiced* ones. Since grouping the fully devoiced with the partially devoiced token generated exactly the same results in step-up and step-down VARBRUL analyses, the more conservative binary distinction between *voiced* (including *partially voiced*) and *completely devoiced* will be used to explain the ranking and the influence of prosodic and segmental factors on devoicing in Turkish.

Dependent Variable:

| | | | |
|---|---|---|---|
| 1.Voicing Status | :voiced | partially devoiced | voiceless |

Independent Variables:

| | | | | |
|---|---|---|---|---|
| 1. Rate: | slow | normal | fast | |
| 2. Stress: | yes | no | | |
| 3. preceding Env.: | fricative | stop | zero-context | |
| 4. following Env.: | ficative | stop | | |
| 5. Vowel Quality: | i | y | ɨ | u |
| 6. Syllable Type: | open | closed | | |

In binomial step-up and step-down analyses, the six independent factor groups (with a total of sixteen factors) are ranked according to which independent factor group (variable) contributes most to the dependent effect (*devoicing*, in this case). In an step-up analysis, initially, all factor groups are evaluated separately to see whether their contribution to the outcome could be due to chance. After the factor group that accounts for the largest proportion of variance is found, the remaining factor groups are again evaluated for the most significant contribution that increases the likelihood of prediction maximally. This is done until no factor groups remain or until no group significantly contributes to the results anymore. The significance level of = .05 was adjusted by adding up the number of levels of applications (6+5+4+3+2+1 = 21) and dividing the original = .05 by the 21 levels of application. The calculation generates an adjusted significance level of $p < .002$ per factor group. This calculation is performed to adjust for the number of times the factor levels are compared with one another and to adjust the level of significance a factor needs to reach in these multiple comparisons in order to significantly contribute to the outcome. In the step-down analysis, the program starts out with all factor groups and eliminates those that contribute least to the outcome. Ideally, the same factors are discarded in the step-up and the step- down analysis.

## III. Results

### 1. Proportions of Tokens within the Data

The following tables give an overview of the distribution of the various considered segmental and prosodic factors considered for the third experiment.

| | slow | normal | fast | Total |
|---|---|---|---|---|
| item | 45 | 45 | 45 | 135 |
| token | 1215 | 1215 | 1215 | 3645 |
| % | 33 | 33 | 33 | ~100 |

Table 1: Proportions of tokens in slow, normal and fast speech.

| | stress | no stress | Total |
|---|---|---|---|
| item | 128 | 7 | 135 |
| token | 3456 | 189 | 3645 |
| % | 95 | 5 | 100 |

Table 2: Proportions of stressed and unstressed tokens.

| | zero | fricatives | stops | affricate | Total |
|---|---|---|---|---|---|
| item | 40 | 37 | 46 | 12 | 135 |
| token | 1080 | 999 | 1242 | 324 | 3645 |
| % | 30 | 27 | 34 | 9 | 100 |

Table 3: Proportions of tokens with preceding zero-environment, fircatives, and stops and affricates.

| | [zero] | [p] | [t] | [k] | [f] | [s] | [ʃ] | [tʃ] | Total |
|---|---|---|---|---|---|---|---|---|---|
| item | 40 | 15 | 12 | 19 | 13 | 16 | 8 | 12 | 135 |
| token | 1080 | 405 | 324 | 513 | 351 | 432 | 216 | 324 | 3645 |
| % | 30 | 11 | 9 | 14 | 10 | 12 | 6 | 9 | 100 |

Table 4: Proportions of tokens with preceding [*zero*], [p], [t], [k], [f], [s], [ʃ], [tʃ] environment.

| | fricatives | stops | affricate | Total |
|---|---|---|---|---|
| item | 67 | 62 | 6 | 135 |
| token | 1809 | 1674 | 162 | 3645 |
| % | 50 | 46 | 4 | 100 |

Table 5: Proportions of tokens with following fricative- and stop, and affricate environment.

| | [p] | [t] | [k] | [f] | [s] | [ʃ] | [tʃ] | Total |
|---|---|---|---|---|---|---|---|---|
| item | 16 | 23 | 23 | 17 | 32 | 18 | 6 | 135 |
| token | 432 | 621 | 621 | 459 | 864 | 486 | 162 | 3645 |
| % | 12 | 17 | 17 | 13 | 24 | 13 | 4 | 100 |

Table 6: Proportions of tokens with following [p], [t], [k], [f], [s], [ʃ], [t] environment.

| | Unround | | Round | | Total |
|---|---|---|---|---|---|
| | i | ɨ | y | u | |
| item | 51 | 35 | 21 | 28 | 135 |
| token | 1377 | 945 | 567 | 756 | 3645 |
| % | 38 | 26 | 16 | 21 | 100 |

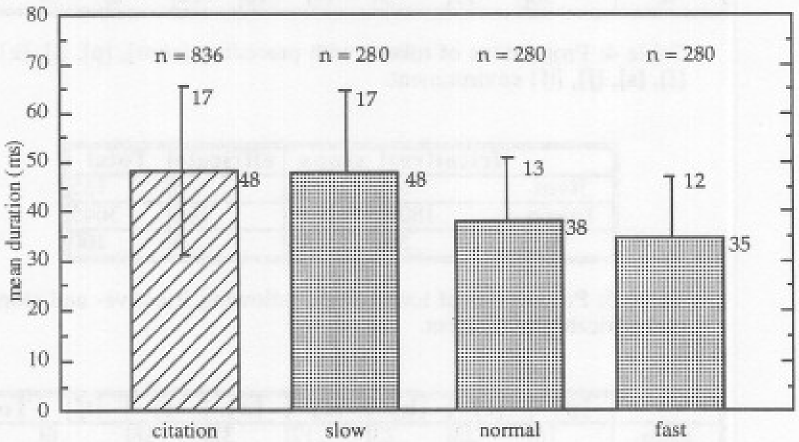
Table 7: Proportions of tokens containing the four different high vowels.

| | open | closed | Total |
|---|---|---|---|
| item | 69 | 66 | 135 |
| token | 1863 | 1782 | 3645 |
| % | 51 | 49 | 100 |

Table 8: Proportions of tokens containing open and closed syllables.

## II. VOT Duration

The graphs (with standard deviation bars) in figure 5 display obtained VOT values in citation and in three different rates in carrier phrases for both sets of data. Note that the upper graph does not encompass data for VOT values for [p].



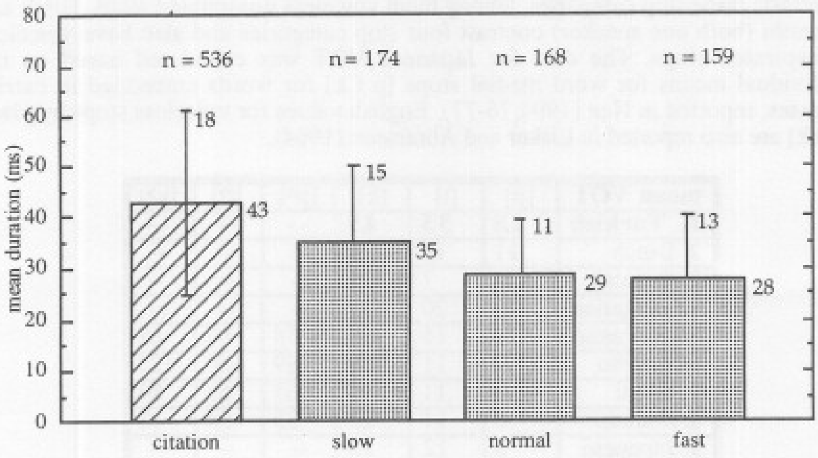



Figure 5.: Upper graph: mean duration of word initial VOT (in ms) of words in citation form and positioned utterance initially in carrier phrases in slow, normal, and fast speech collapsed over all three Turkish voiceless stops (appendix A). Lower graph: mean duration of word medial VOT (in ms) of words in citation form and positioned utterance initially in slow, normal, and fast speech (appendix B).

The mean VOT duration is indicated by the number next to the bar, standard deviation is displayed next to the standard deviation bar. The total number of tokens is given above the individual bars. In both sets of data in the citation condition (total: 840 tokens in the upper graph and 540 tokens in the lower graph), 4 tokens could not be measured due to incomplete stop closures. In the lower graph, in the phrasal condition (180 tokens total), 6 tokens had to be discarded due to incomplete closures during the stop production. In the normal and fastest rates 12 and 21 tokens were not measured because of incomplete stop closures or because the vowel following the stop was devoiced so that the data would have been confounded with these measurements.

Homma (1981:276) reports the mean VOT for initial voiceless unaspirated stops [p t k] in Japanese to be 37ms and for medial stops to be 16ms. The comparable results for Turkish in a phrasal condition in a normal rate of speech show a mean VOT of 38ms (collapsed over all three places of articulation) in initial position and 29ms in medial position. According to these results, Turkish initial voiceless stops have about the same amount of accompanying aspiration as the Japanese ones. The Turkish word medial values for VOT are slightly longer compared to the ones stated by Homma. Thus, we might expect a slightly different pattern for the preceding stops and fricatives in comparison to Japanese.

The results by place show that Turkish VOT durations are longer than the comparable values[2] for wordinitial voiceless unaspirated stops in sentence initial position (Lisker & Abramson, 1964) in Dutch (one speaker), Puerto Rican Spanish (two speakers), Hungarian (two speakers), or Cantonese (one speaker), languages that all contrast voiced and voiceless unaspirated stops. Korean (one speaker)

[2]The data from Lisker and Abramson reported in the table are VOT values for word initial stops in sentence initial position.

contrasts three stop categories, among them voiceless unaspirated stops. Hindi and Marathi (both one speaker) contrast four stop categories and also have voiceless unaspirated stops. The data for Japanese VOT was calculated based on the individual means for word medial stops [p t k] for words embedded in carrier phrases, reported in Han (1994:76-77). English values for voiceless stops by place [p t k] are also reported in Lisker and Abramson (1964).

| mean VOT | [p] | [t] | [k] | [pʰ] | [tʰ] | [kʰ] |
|---|---|---|---|---|---|---|
| **1. Turkish** | **28** | **35** | **49** | -- | -- | -- |
| 2. Dutch | 11 | 16 | 34 | -- | -- | -- |
| 3. Spanish | 4 | 7 | 25 | -- | -- | -- |
| 4.Hungarian | 0 | 20 | 28 | -- | -- | -- |
| 5.Cantonese | 11 | 15 | 34 | 58 | 62 | 68 |
| 6. Korean | 7 | 11 | 20 | 89 | 100 | 125 |
| 7. Hindi | 12 | 11 | 16 | 63 | 63 | 84 |
| 8. Marathi | 0 | 11 | 21 | 35 | 54 | 73 |
| 9. Japanese | 8 | 12 | 18 | -- | -- | -- |
| 10 English | -- | -- | -- | 28 | 39 | 43 |

Table 9: Mean VOT values for voiceless unaspirated and voiceless aspirated stops for ten languages. (Data in 2 through 8 and 10 from Lisker and Abramson, 1964; values for Japanese calculated from Han's (1994) data).

The cross-language comparison of voice onset time values for voiceless unaspirated stops shows that Turkish VOT duration for the three voiceless stops in utterance initial position falls in between the values established for languages with comparable two way contrasts between voiceless unaspirated and voiceless aspirated stops. Note however, that Turkish and English show very similar values for VOT in sentence initial position[3]. For a wider comparison, values for voiceless unaspirated and voiceless aspirated stops in Korean, Hindi, and Marathi are given as well. These values confirm that the accompanying aspiration of Turkish voiceless stops is slightly longer compared to other language's voiceless unaspirated stops, but shorter than values for aspirated stops. Throughout all three different speech rates, VOT in Turkish was longest for [k], and shortest for [p], thus decreasing with distance from the glottis. Figure 5 shows VOT values of word initial voiceless stops before low vowels utterance initially positioned in carrier phrases in three different speech rates, displayed by place of articulation.

[3]The results of the VOT-experiments might reflect that four out of five consultants had been living and studying in the United States for at least a year prior to recording and that all of them had learned English in school prior to their arrival. Thus, the prolonged durations of the VOT values for Turkish voiceless stops might be due to interference from English.

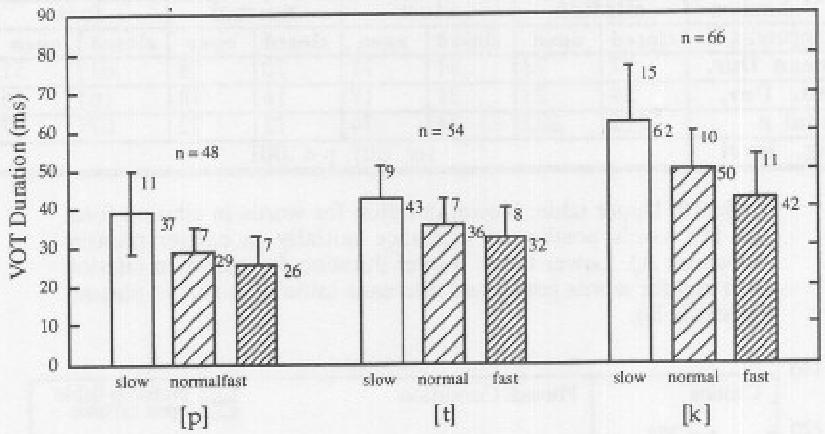



Figure 6: Mean duration of VOT (in ms) of word initial voiceless stops in slow, normal, and fast speech for all three places or articulation.

It is relevant to note that VOT in Turkish is slightly longer than in Japanese, thus we might expect a slightly different devoicing pattern for preceding stops compared to preceding fricatives in Turkish, in comparison to Japanese, Montreal French or Korean.

## III. Vowel Duration

According to duration measurements of non-high vowels in 28 open and closed syllable "minimal pairs" (see Appendix A), vowels are, contrary to findings for many other languages (Maddieson, 1985) significantly longer in closed syllables than in open syllables. This result was confirmed in a replication of an experiment (15 [minimal]-pairs of words contrasting open syllables with syllables closed by geminates) described in Lahiri & Hankamer (1988) who found non-significantly longer vowels in closed syllables. The following tables display mean vowel durations, standard deviation, and total number of tokens and the significance level at which vowel duration is different between open and closed syllables in citation form and when uttered phrase initially. The following tables show the mean vowel duration and standard deviation (in ms), the total number of tokens and the significance level for the data in appendix 1 (upper table) and appendix 2 (lower table) in citation form and embedded in phrases in three different rates of speech for open and closed syllable types.

| V-duration | **citation** | | **slow** | | **normal** | | **fast** | |
|---|---|---|---|---|---|---|---|---|
| appendix A | **closed** | **open** | **closed** | **open** | **closed** | **open** | **closed** | **open** |
| **mean Dur.** | 77 | 66 | 69 | 60 | 57 | 48 | 51 | 44 |
| **std. Dev.** | 17 | 19 | 13 | 13 | 12 | 10 | 13 | 13 |
| **total n** | 419 | 419 | 140 | 140 | 140 | 140 | 138 | 138 |
| **sig. level** | sig. diff. p < .001 | | | | | | | |

| V-duration | citation | | slow | | normal | | fast | |
|---|---|---|---|---|---|---|---|---|
| appendix B | closed | open | closed | open | closed | open | closed | open |
| **mean Dur.** | 87 | 78 | 81 | 71 | 70 | 58 | 61 | 51 |
| **std. Dev.** | 24 | 23 | 21 | 17 | 16 | 14 | 16 | 15 |
| **total n** | 222 | 222 | 74 | 74 | 72 | 72 | 67 | 67 |
| **sig. level** | sig. diff. $p < .001$ | | | | | | | |

Table 10: Upper table: Vowel duration for words in citation form and for words positioned utterance initially in carrier phrases (appendix A). Lower table: Vowel duration for words in citation form and for words positioned utterance initially in carrier phrases (appendix B).

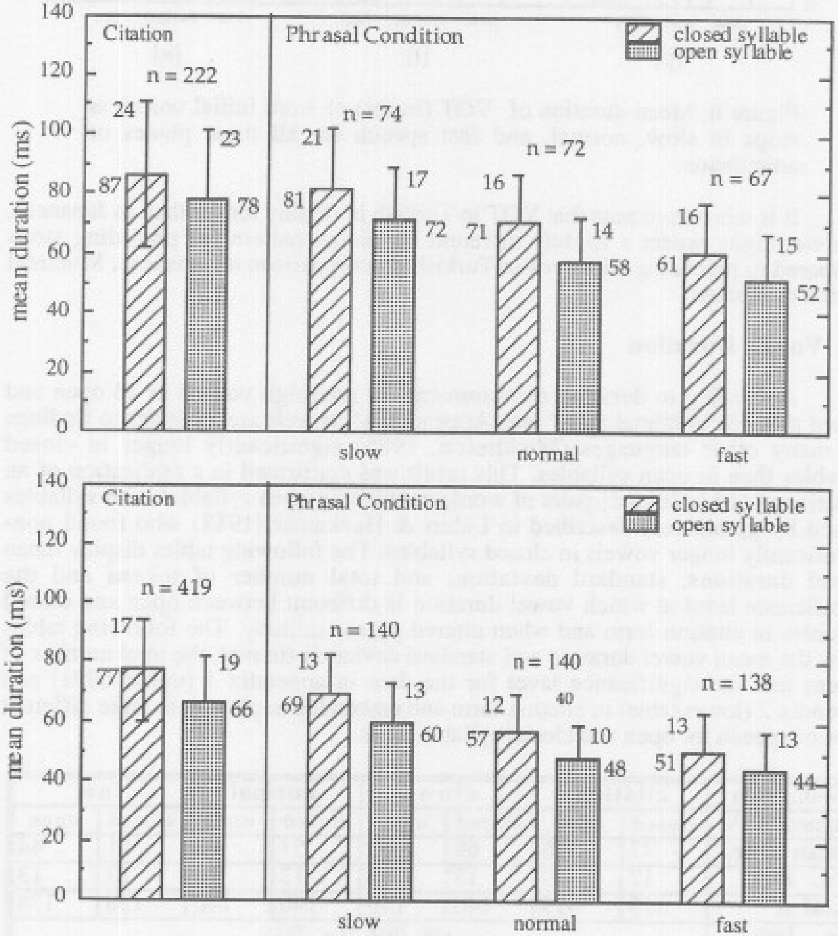



Figure 7. Upper graph: mean duration of non-high vowels in closed and open syllables in citation (left), and phrase initially in slow, normal, and fast speech (right) for 28 syllable type minimal-pairs.

Lower graph: mean duration of (mainly non-high) vowels in closed and open syllables in citation (left), and phrase initially in slow, normal, and fast speech (right) for 15 syllable type minimal-pairs.

Since Lahiri and Hankamer originally measured vowel duration from waveforms and not spectrograms, measurements for five pairs of their 15 words were repeated from waveforms and correlated with the measurements for the identical token obtained from spectrograms. The correlation was r = .82 for open and r = .85 for closed syllables. A paired t-test showed that even with a relatively small total *n* of 75, mean differences in vowel duration in open and closed syllables were highly significant regardless of measurement tool (spectrogram: t = -8.84, p < .001; waveform: t = -6.86; p < .001). Significantly longer vowels in closed syllables in Turkish are a robust effect.

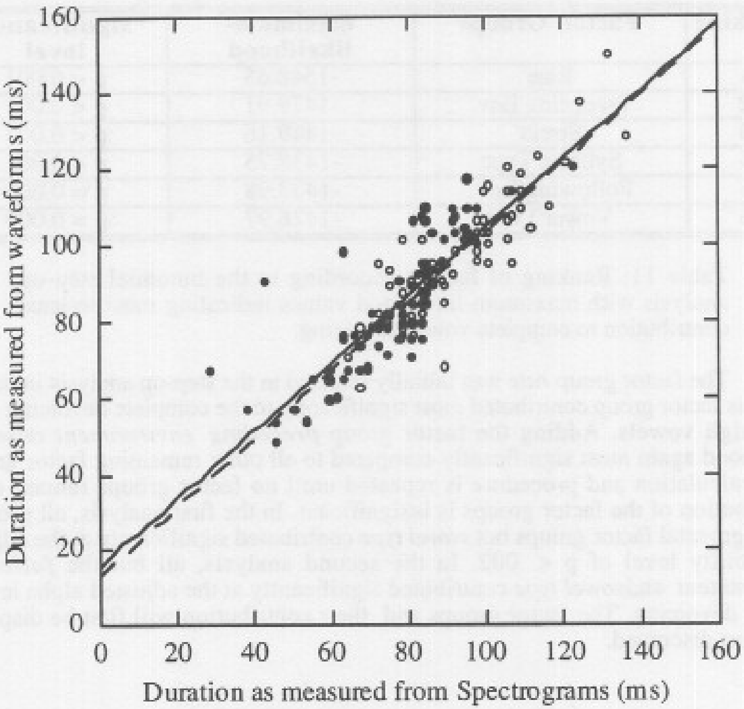



Figure 8.: Left: correlation of vowel duration measurements obtained from spectrograms (x-axes) and waveforms (y-axes) in open and closed syllables.

## 4. Devoicing

Grouping the partially devoiced with the fully voiced vowels generates the overall rankings of the factor groups as displayed in the following tables. The first table shows the slightly different ranking of the factors when considering each preceding and following consonant type separately. The second table shows the ranking of the factors when collapsing the preceding and the following environment by consonant manner (stops vs. fricatives vs. affricates vs. no onset).The value for the maximum-likelihood is an estimator of the effectiveness of the factor in accounting for the pattern of vowel devoicing. The significance levels reflect the probability that this factor is selected by chance.

| Ranking | Factor Groups | maximum-likelihood | significance level |
|---|---|---|---|
| 1 | Rate | -1548.65 | $p < 0.001$ |
| 2 | Preceding Env. | -1475.95 | $p < 0.001$ |
| 3 | Stress | -1442.69 | $p < 0.001$ |
| 4 | Following Env. | -1419.02 | $p < 0.001$ |
| 5 | Syllable Type | -1406.47 | $p < 0.001$ |
| 6 | Vowel Type | -1399.82 | $p = 0.006$ |

| Ranking | Factor Groups | maximum-likelihood | significance level |
|---|---|---|---|
| 1 | Rate | -1548.65 | $p < 0.001$ |
| 2 | Preceding Env. | -1479.91 | $p < 0.001$ |
| 3 | Stress | -1449.16 | $p < 0.001$ |
| 4 | Syllable Type | -1438.75 | $p < 0.001$ |
| 5 | Following Env. | -1432.98 | $p = 0.005$ |
| 6 | Vowel Type | -1426.97 | $p = 0.009$ |

Table 11: Ranking of factors according to the binomial step-up analysis with maximum-likelihood values indicating most to least contribution to complete vowel devoicing.

The factor group *rate* was initially selected in the step-up analysis indicating that this factor group contributed most significantly to the complete devoicing of the four high vowels. Adding the factor group *preceding environment* raises the likelihood again most significantly compared to all other remaining factor groups. This calculation and procedure is repeated until no factor groups remain or the contribution of the factor groups is insignificant. In the first analysis, all prosodic and segmental factor groups but *vowel type* contributed significantly at the adjusted probability level of $p < .002$. In the second analysis, all but the *following environment* and*vowel type* contributed significantly at the adjusted alpha level to vowel devoicing. The factor groups and their contribution will first be displayed and then discussed.

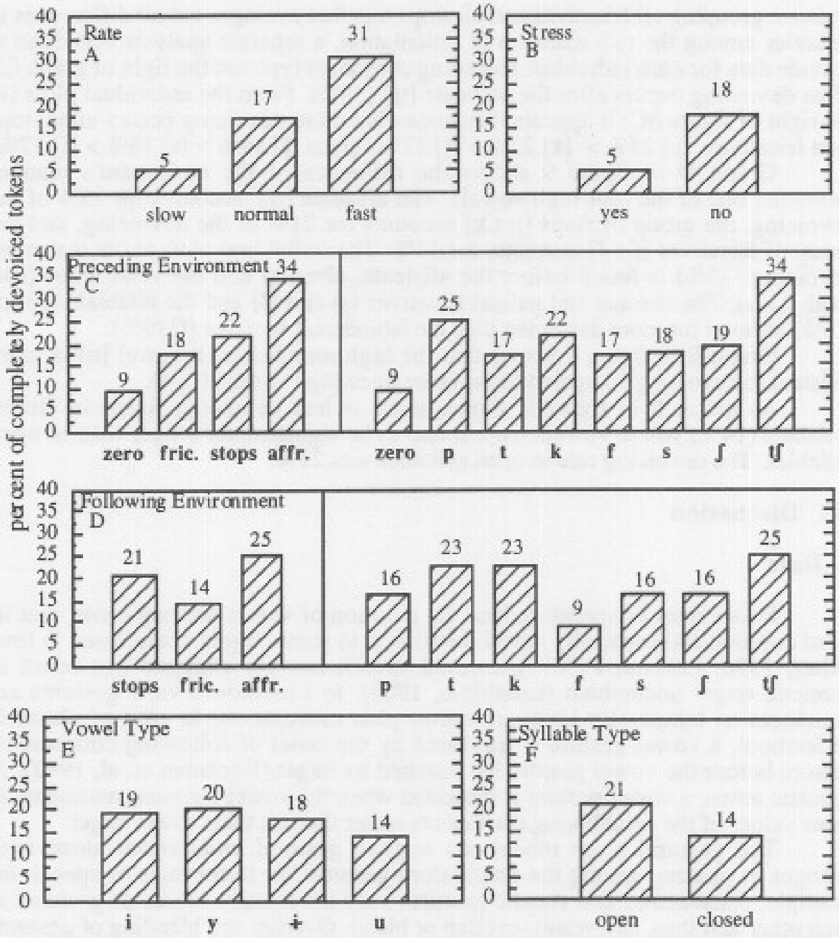



Figure 9.: Influence of rate, stress, segmental environment, vowel- and syllable type on the process of high vowel devoicing in Turkish.

In Turkish, high vowels were more frequently devoiced at faster speaking rates. Graph A in figure 12 shows the contribution of different rates of speech to the process of vowel devoicing by showing the percent of completely devoiced tokens averaged over all nine speakers and all 3645 tokens. The number of tokens that are devoiced as an effect of speech rate increases from slow (5%) to normal (17%) to fast (31%).

Only 5% of all stressed syllable token underwent vowel devoicing, whereas 18% of the vowels devoiced when the syllable was unstressed. Graph B in figure 12 shows a plot of the percent of token that were fully devoiced in stressed and unstressed position.

Stops, fricatives and affricates in the preceding environment were more closely associated with the devoicing of the vowel than no consonant in the onset. Graph C in figure 12 shows the effect of a *null-context* (9%) and the contribution of *stops* (22%) and *fricatives* (18%) and affricates (34%) on high vowel devoicing.

Because grouping all fricatives and all stops together averages out all differences in behavior among the two manners of articulation, a separate analysis was done to provide data for each individual preceding consonant type (on the right of graph C). Most devoicing occurs after the affricate [tʃ] (34%). From the individual plots (in the right of figure 8C) it appears that somewhat more devoicing occurs after stops than fricatives: [p] 25% > [k] 22% > [t] 17% versus [ʃ] 19% > [s] 18% > [f] 17%.

Graph D in figure 9 shows the influence of the consonant's manner following one of the four high vowels. The affricate [tʃ] accounts fpr 25% of the devoicing, the group of stops [p t k] accounts for 21% of the devoicing, and the group of fricatives [f s ʃ] accounts for 14%. The individual plots show that most devoicing (25%) is found before the affricate, alveolar and the velar stops [t k] (both 23%). The coronal and palatal fricatives [s] and [ʃ] and the bilabial stop [p] (16%) account for more devoicing than the labiodental fricative [f] (9%).

Graph E in figure 9 shows that the high rounded back vowel [u] is more resistant to devoicing compared to the other three high vowels [i y ɨ].

As graph F in figure 9 shows, there is less devoicing found in closed syllables (14%) where vowels were found to be significantly longer than in open syllables. The devoicing rate in open syllables was 21%.

## IV. Discussion

### 1. Rate

Faster rates of speech reduce the duration of words and segments, that is, word duration and especially vowel duration is to some degree compressed in time (Klatt, 1976; Lehiste, 1970). Increased speech rate for example can result in phonetic target undershoot (Lindblom, 1963). In Lindblom's view gestures are sequences of temporally invariant motor-plan movements. In case of phonetic undershoot, a vowel gesture is truncated by the onset of following consonantal gesture before the vowel gesture has reached its target (Beckman et. al, 1992). In acoustic terms, a vowel gesture is truncated when the vowel formants assimilate to locus values of the neighboring consonants rather than hit their vowel target.

The gestural score model can explain gestural *undershoot* alone with changes in phasing among the articulatory gestures. In faster rates of speech for example, consonantal and vocalic gestures vary the relative onsets of gestures to each other and thus, as a result, overlap or blend. Overlap and blending of gestures can result in hiding a gesture so that acoustically no output is generated (Beckman, et al., 1992; Munhall & Löfqvist, 1992). Vowel devoicing can nicely be explained by this model, too. It is predicted to occur with greater frequency at faster rates of speech because the glottal gestures of neighboring consonants overlap the vowel's laryngeal gesture to greater or lesser extend. In a $C_1VC_2$ sequence for example, a vowel's laryngeal gesture can be partially or completely overlapped by adjacent voiceless consonantal gestures only by modifying the phasing between the gestures: $C_1$'s glottal gesture extends into the vowels gesture and $C_2$'s glottal gesture sets on earlier than in normal or slower rates of speech. The figure below illustrates this gestural reorganization resulting in undershoot of the voicing gesture.

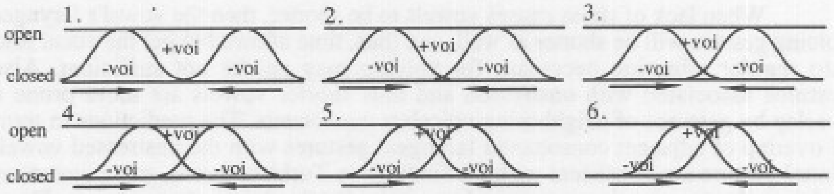



Figure 10: Hypothetical changes in gestural phasing on the glottal tier due to rate. Two voiceless gestures gradually overlap an intervening voiced gesture.

A gestural overlap interpretation of the results of this study regarding the impact of speech rate on glottal gestures is supported much by a study conducted by Munhall and Löfqvist (1992). The authors elicited multiple renditions of the phrase *Kiss Ted* in different speech rates ranging from slow to fast. In the slowest renditions they found two distinct glottal opening and closing movements at the word boundary between the [s] in *Kiss* [kɪs] and the aspirated [t] in *Ted* **[tʰɛd]**. With increasing rate of speech (in intermediate tempi) the two glottal gestures blended and the gesture for [s] became a shoulder of the gesture for the aspirated [tʰ]. In the fastest rates, Munhall and Löfqvist interpret the two gestures as completely overlapped and blended into one glottal opening, whereby the [t] acoustically lost its aspiration due to gestural reorganization resulting in a change in timing of the glottal gesture in relation to the oral gesture for the [t]. These results show that faster tempi trigger overlap of two adjacent glottal gestures, and thus, by analogy, more devoicing of high vowels should be found in faster rates where the preceding and the following voiceless consonant's glottal gesture are predicted to overlap with the vowel's glottal gesture.

## 2. Stress

Since the 1940s it was debated in the literature (Benzing, 1941; Collinder, 1939; Duda, 1940; Grønbech, 1940) if there is *stress* in Turkish and how it is distributed. Newer literature and phonological descriptions report Turkish as regularly having stress on the final syllable (Lees, 1961; Underhill, 1986; Van der Hulst & Van der Weijer, 1991), whether the word is derived or not.[4] This appears to be the most widely held position. A perception study testing the bias to perceive stress at a particular location in synthesized non-sense words with constant f0, amplitude, duration and target formant values for the syllable nucleus (Konrot, 1987) does not show consistent results. Except for one study by Boyce (1978) the question of what the phonetic correlates of *stress* in Turkish are is practically unaddressed. In contrast to English, where stressed vowels are longer compared to unstressed vowels, Boyce found the durational differences between stressed and unstressed syllables in Turkish to be less striking. Nevertheless, we need to explain why stressed vowels in Turkish are more resistant to devoicing than unstressed ones. Although the influence of stress on vowel duration in English and Turkish is incomparable, data from Montreal French (Cedergren and Simeneau, 1986; Cedergren, 1985) suggests that vowels in rhythm group final syllables are particularly resistant to vowel syncope due to the stress placement on the final syllable.

[4]There are irregular word stress rules which will not be discussed here. See Van der Hulst & Van der Weijer (1991) for a discussion of *stress* in Turkish.

When lack of stress causes vowels to be shorter, then the vowel's laryngeal voicing gesture will be shorter as well, and thus, time allowed to get the vocal folds into regular vibration necessary for voicing may not be not sufficient. Also, gestures associated with unstressed and thus shorter vowels are more prone to overlap by gestures of neighboring voiceless consonants. The predictions in terms of overlap of adjacent consonantal laryngeal gestures with the unstressed vowel's glottal gesture are consistent with the findings in Turkish, that is, more unstressed and thus shorter vowels were more frequently devoiced than stressed ones. Figure 14 shows hypothetical gestural phasings within stressed and unstressed CVC syllables in which both flanking consonants are assumed to be voiceless.

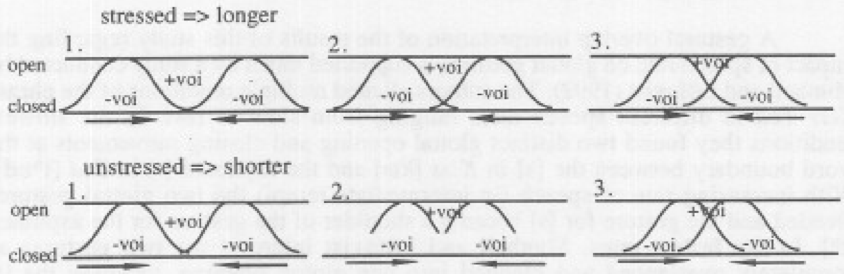



Figure 11: Hypothetical changes in gestural phasing in stressed and unstressed environment.

## 3. No preceding consonant (Zero Onset)

Turkish does not readily allow word initial consonant clusters (Kornfilt, 1987) since they are a violation of the phonotactic constraints of that language (van der Hulst & van der Weijer, 1991). Some borrowings into Turkish, mainly words of western origin (Özen, 1985; Kornfilt, 1987) contain syllable initial consonant clusters. To break up these disfavored consonant clusters, the languages uses two different mechanisms: one is vowel epenthesis (Lees, 1961; Clements and Sezer, 1982; van der Hulst & van der Weijer, 1991) between two cluster consonants, and the other is prothesis (van der Hulst & van der Weijer, 1991) by which a vowel becomes inserted before word initial consonant clusters such as *sp*-, *st*-, and *sk*-:[5]

| | |
|---|---|
| spanak | 'spinach' |
| statistik | 'statistics' |
| iskelet | 'skeleton' |

The prothesized initial vowel causes resyllabification resulting in the resolution of the violation of the constraint against onset clusters. Although these word initial consonant clusters are undesirable in Turkish, we find cases where this phonological rule of prothesis is revoked by the phonetics, that is, the vowel is completely devoiced and (in effect perceptually) deleted with no formant structure in the fricative spectrum so that a word initial consonant cluster resurfaces[6].

[5] These examples are given by Van der Hulst & Van der Weijer (1991:14).

[6] The language- and alphabet reform, propagated by Atatürk in 1928 prescribed the usage of Turkish words over foreign words (i.e., *istanbul* for *Constantinople*) as well as the conversion of the Arabic writing system to the Roman alphabet (Brendemoen, 1990). Grønbech (1940) writes that one of his colleagues retuned from *Stambul* where he did field work on Turkish. This

Vowel devoicing in word- or utterance initial position occurred in 9% of the vowel initial cases. This is less easy to interpret in terms of the gestural score model, because only the following laryngeal gesture (the one to the right of the vowel) can overlap with the vocalic gesture to cause it to be devoiced. A simple overlap explanation predicts that devoicing will be less frequent when no consonantal onset precedes the vowel. The glottal gesture of the consonant following the vowel must be phased with the vowel's laryngeal gesture in such a way that the following consonantal gesture must completely overlap with the preceding vowel gesture so that the vowel's target cannot be realized. This is represented in Figure 15 below.

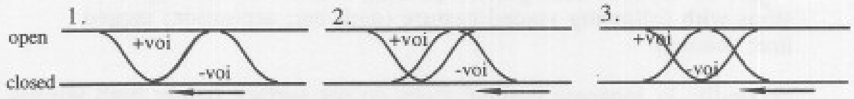



Figure 12: Hypothetical gestural phasing of vowel gesture and following consonantal gesture.

It is noteworthy that disobeying the phonological constraint, still 9% of the vowels with zero-onset devoiced. There is evidence, that at least in some cases, the utterance initial vowel is preceded by a glottal stop onset. But even sequences like $ʔ\text{-}V\text{-}C_1\bullet C_2$ cannot be readily explained since the devoicing of the vowel between the glottal stop and $C_1$ should cause a $ʔ\text{-}C$ sequence to appear which also is also a violation of the constraint against syllable initial consonant clusters in Turkish.

## 4. Preceding Consonant Type

Two observations are to be made with regard to the preceding consonantal environment: generally, in Turkish, preceding stops appear to account for more devoicing than fricatives, contrary to Jun and Beckman's (1994) findings for Korean, Cedergren and Simeneau's (1985) counts for Montreal French, and a report by Nagano-Madsen (1994:120) citing studies on Japanese unavailable in English. Fricatives in Korean and Japanese have a longer peak glottal opening than stops (Kayaga, 1974; Yoshioka et. al., 1986) but Japanese data provided by Sawashima & Hirose (1983) shows no delay in voicing onset after either voiceless fricative or unaspirated stop.

A higher devoicing rate in the presence of prevocalic stops versus fricatives in Turkish can maybe be explained by the slightly longer aspiration phases of Turkish voiceless stops compared to those of Japanese (and Montreal French, although no VOT data is available here) which have roughly three times less accompanying aspiration at the release of the unaspirated voiceless stops than the Turkish stops. Interestingly, most devoicing is found after the least aspirated stop [p] which suggests that there is some other overlooked contributing factor. How this can be explained remains unclear at this point since no data on glottal opening is available for Turkish. The hypothetical differences in phasing for unaspirated stops with more and less aspiration is shown in the next figure.

annecdotal evidence of Benzing's orthographic representation of this city name hints at how (at least Bechman) perceived the name of Turkey's capitol city.

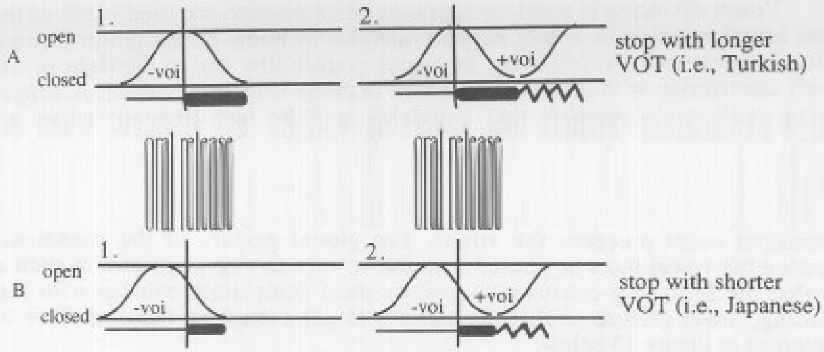



Figure 13: Hypothetical gestural phasing of more and less aspirated stops with following voiced gesture (dark bar: aspiration; jagged line: voicing).

Secondly, in Japanese coronal stops do not surface before high vowels. That is, /ti/ surfaces as [tʃi] and /tu/ surfaces as [tsu]. Underlying coronal stops before high vowels surface as affricates and not as stops. In Japanese, manner (stop versus fricative) is confounded with place, thus, we might expect coronal consonants with their faster oral gestures to show glottal overlap more easily independently of oral overlap. Affricates preceding vowels should behave more like fricatives. Observing the overall pattern of how affricates pattern in comparison to stops and fricatives, they appear to act more like stops when preceding vowels since slightly more devoicing is found after stops than fricatives. However, in Turkish, preceding stops and fricatives account for very similar amounts of devoicing for following high vowels. One could speculate that duration of peak glottal opening during consonant articulation and even possibly the size of the glottal opening might be language specific. Dixit (1989:228) states that unaspirated voiceless plosives in Dutch, French, Japanese and other languages, show "a considerable variability in the degree of glottal opening and the positioning of glottal peak during the initial unvoiced unaspirated plosives across languages; [...]". Also, it is conceivable that affricates like [t] are single phonological entities with their own intergestural timing properties. Even though acoustically, one might think of affricates as being a combination of a stop and a fricative, that is, a stop with a fricative release, it is not outrageous to assume that in production, the closure and release phases of affricates have different phase relations than a stop closure followed by a fricative. In other words, a single segment affricate might not just simply be a combination of two gestures with different manners.

## 5. Following Consonantal Environment

As for the impact of the following environment on vowel devoicing, results for Turkish are consistent with findings of Jun & Beckman (1994) for Korean where more devoicing was found before stops than fricatives. As data by Sawashima & Hirose (1983) show for Japanese, vocal fold vibration ceases abruptly in *V-stop* sequences compared to *V-fricative* sequences where vocal fold vibration ceases gradually: in order to sustain frication at a constriction in the upper vocal tract, airflow needs to be maintained, whereas a following stop requires a complete blockage of the airstream (Ohala, 1983). This is schematized in the following figure. Thus, more devoicing is expected for vowels followed by stops than by fricatives. Cedergren and Simoneau (1985) report generally less devoicing before voiceless fricatives than before voiceless affricates and voiceless stops in Montreal French. Also, in terms of the gestural overlap model and as shown previously by Munhall and Löfqvist (1992), gestural overlap can just as effectively occur across syllable boundaries as within syllables.

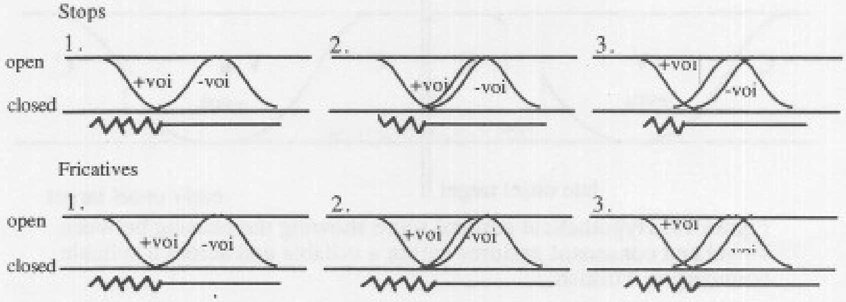



Figure 14: Hypothetical gestural phasing of following stop- and fricative environments with preceding vowel.

The exception among the Turkish stops in the following environment seems to be the bilabial voiceless stop [p] that shows somewhat less devoicing than [k] and [t]. Both labial sounds [f] and [p] show the least amount of impact on the devoicing process, possibly because in the articulation of labials the jaw is involved as an articulator. The affricate, which should behave like a stop following a vowel, is ranked second highest right after the voiceless non-labial stops.

## 6. Syllable Type

Unlike in most other languages, vowel duration is significantly shorter in open syllables ($p < .001$) than in closed ones. Durational differences of vowels in open and closed syllables have also been noted by Boyce (1978). The current study showed that greater vowel duration in closed syllables is a robust effect true for vowels before geminates as well as heterorganic consonant clusters. Whether the statistically significant difference in duration of vowels in production is perceptually salient, is currently under investigation. Since more devoicing is expected for shorter vowels, the outcome is just as we might expect: more vowels are devoiced in open than in closed syllables.

The occurrence of longer vowels in closed syllables through different experimental conditions (read in different speech rates, within a carrier phrase, in isolation, measurements from spectrogram and waveform) is interesting in itself. One might hypothesize phasing relationships of articulatory gestures to be language specific to explain this finding: possibly, in Turkish, consonantal gestures following a vowel within a syllable have a later onset phase target with regard to the preceding vowel in comparison to when a syllable boundary is intervening between the vowel and the consonant.

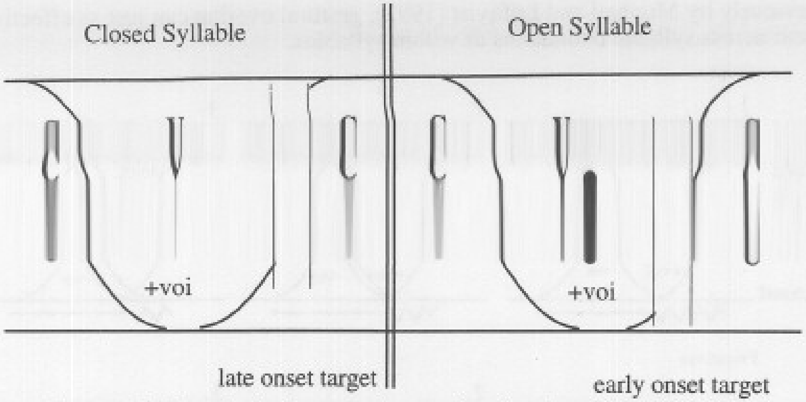



Figure 15: Hypothetical gestural score showing the phasing between vowel and consonant gestures within a syllable and across a syllable boundary in Turkish.

Phonotactic constraints on the possible shape of a syllable appear to be important as well: the complete devoicing of a syllable initial vowel in words like *spanak* 'spinach' or *Istanbul* 'Istanbul (city)' generates undesired and less frequent syllable onset clusters (attested in some western loan words), and thus devoicing or complete deletion in this position is possible but fairly rare. Vowels might be longer in closed syllables with a $C_1VC_2{\bullet}C_3$ structure so that consonant clusters or consonant sequences like $C_1C_2{\bullet}C_3$ or $C_1{\bullet}C_2C_3$ are prevented after devoicing or deletion and resyllabification.

Although Maddieson (1985) reports languages with longer vowels in closed syllables to be fairly unusual, Han (1994) provides data on Japanese, showing that in minimal pairs, differing in the openness and closedness of the first syllable, less devoicing occurs in the closed syllables (vowels before geminates) than in open syllables (vowel before single consonant) because the vowel in the closed syllable is generally longer. If indeed this finding can be explained with the moraic structure in Japanese, Turkish might be temporally organized on the basis of the mora too. However, there is no evidence for this assumption as of now.

### 7. Vowel Type

The impact of vowel type is statitically insignificant. All four Turkish high vowels can become devoiced, however [u] is slightly more resistant to devoicing than [i y ɨ]. This resistancy of [u] is difficult to explain, without more knowledge about the articulation of the sounds.

## V. Conclusions

The presented data showed that various prosodic and segmental factors influence the process of high vowel devoicing in Turkish. As proposed by Jun & Beckman (1993, 1994), these findings can be explained in terms of gestural overlap where the laryngeal gestures of consonants overlap or blend with the glottal adduction gesture of the preceding or following vowel. The data also suggest language specific timing relations between glottal gestures. When comparing devoicing patterns in i.e., Japanese and Turkish, we find that Turkish VOT durations of voiceless stops are roughly three times as long in comparison to Japanese. Thus, language specific differences (like VOT-duration or vowel duration differences in open and closed syllables) play a role in explaining the overall pattern of devoicing in languages.

## VI. References


Beckman, M. E. & Shoji, A. (1984) *Spectral and Perceptual Evidence for CV Coarticulation in Devoiced /si/ and /syu/ in Japanese*. Phonetica, 41: 61-71.

Beckman, M. E., De Jong, K., Jun, S.-A.; Lee, S.-H. (1992) *The Interaction of Coarticulation and Prosody in Sound Change*. Language and Speech, 35 (1,2), pp. 45-58.

Benzing, J. (1941) *Noch einmal die Frage der Betonung im Türkischen*. Zeitschrift der Deutschen Morgenländischen Gesellschaft, Bd. 94 (19), pp. 300-301.

Bloch, B. (1950) *Studies in Collocquial Japanese: IV phonemics*. Language, 26, 86-125.

Boyce, S. (1978) *Accent or Accident:? The Acoustic Correlates of Word Level Prominence in Turkish*. A.B. Thesis, Harvard University (Linguistics Deparment), Cambridge, MA.

Brendemoen, Bernt (1990) *The Turkish Language Reform and Language Policy in Turkey*. Hazai, G. (ed.) Handbuch der Türkischen Sprachwissenschaft, Teil I, pp. 454-493.Otto Harrassowitz, Wiesbaden.

Browman, C. & Goldstein, L. (1990) *Tiers in articulatory phonology, with some implications for casual speech*. Kingston, J. & Beckman, M. E. (eds.) Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, pp. 341-376. Cambridge: Cambridge University Press.

Cedergren, H. J. & Simeneau, L. (1985) *La chute des voyelles hautes en français de Montreal <<As-tu entendu la belle syncope?>>*. Lemieux, M. & Cedergren, H. J. (eds.) Les tendences dynamiques du français parlé a Montréal, Tome 1, 57-144.

Cedergren, H. J. (1986) *Metrical Structure and Vowel Deletion in Montreal French*. Sankoff, D. (ed.) Diversity and Diachrony, 293-300.

Clements & Sezer, (1982) *Vowel and Consonant Disharmony in Turkish*. Van der Hulst, H. & Smith, N. (eds.) The Structure of Phonological Representations, Part II, pp. 213-255.

Collinder, B. (1939) *Gibt es im Türkischen keinen Akzent?*. Zeitschrift der Deutschen Morgenländischen Gesellschaft, Bd. 94(20), pp. 305-310.

Dixit, R. P. (1989) *Glottal gestures in Hindi Plosives*. Journal of Phonetics, 17, 213-237.

Duda, H. W. (1940) *Zu Björn Collinder's osmanisch-türkischen Lautstudien*. Zeitschrift der Deutschen Morgenländischen Gesellschaft, Bd. 94(20), pp. 86-105.

Griffen, T. D. (1983) *The Svabian voiceless vowel*. Word, 92, Vol. 34, pp. 145-173.

Grønbech, K. (1940) *Der Akzent im Türkischen und Mongolischen*. Zeitschrift der Deutschen Morgenländischen Gesellschaft, Bd. 94(19), pp. 375-390.

Han, M. S. (1994) *Acoustic manifestations of mora timing in Japanese*. Journal of the Acoustical Society of America, Vol. 96 (1), pp. 73-82.

Homma, Y. (1981) *Durational relationship between Japanese stops and vowels*. Journal of Phonetics, Vol. 9, pp. 273-281.

Jun, S.-A. & Beckman, M. E. (1993) *A gestural-overlap analysis of vowel devoicing in Japanese and Korean*. Paper presented at the 1993 Annual Meeting of the Linguistic Society of America, 7-10 January 1993, Los Angeles, CA, USA.

Jun, S.-A. & Beckman, M. E. (1994) *Distribution of Devoiced High Vowels in Korean*. Proceedings of the International Conference on Spoken Language Processing, Yokohama, Japan, Vol. 2, 479-482.

Kayaga, R. (1974) *A Fiberscopic and Acoustic Study of the Korean Stops, Affricates and Fricatives*. Journal of Phonetics, Vol.2, pp. 161-180.

Klatt, D. H. (1976) *Linguistic Uses of segmental duration in English: Acoustic and Perceptual Evidence*. Journal of the Acoustical Society of America, Vol. 59 (5), pp. 1208-1221.

Konrot, A. (1987) *Stress in Türkish*. In Boeschoten, H. E. & Verhoeven, L. Th. (eds.) Proceedings of the third conference on Turkish Linguistics, 3-12. TilburgUniversity Press.

Kornfilt, J. (1986) *Stem-Penultimate Empty Cs, Compensatory Lengthening, and Vowel Epenthesis in Turkish*. Wetzels, L. & Sezer, E. (eds.) Studies in Compensatory Lengthening, ch. 4, pp. 79-96. Foris Publications, Dordrecht - Holland.

Kornfilt, J. (1987) *Turkish and the Turkic Languages*. Comrie, B. (ed.) The World's Major Languages, ch. 30, pp 619-644.

Lahiri, A. & Hankamer, J. (1988) *The Timing of Geminate Consonants*. Journal of Phonetics, 16, pp. 327-338.

Lees, R. B. (1961) *The Phonology of Modern Standard Turkish*. Indiana Iniversity Publications, Uralic and Altaic Series, Vol 6. Indiana University: Bloomington.

Lehiste, Ilse (1970) *Suprasegmentals*. MIT Press, Cambridge, MA.

Lindblom, B. (1963) *Spectrographic Study of Vowel Reduction*. Journal of the Acoustical Society of America, Vol. 35 (11) , pp. 1773-1781.

Lisker, L. & Abramson, A. S. (1964) *A Cross-Language Study of Voicing in Initial Stops: Acoustic Measurements*. Word 20, pp. 384-422.

McCawley, J. D. (1968) *The phonological component of a grammar of Japanese*. Mouton, The Hague, Paris.

Maddieson, I. (1985) *Phonetic Cues to Syllabification*. Fromkin, V. A. (ed.) Phonetic Linguistics. Essays in Honor of Peter Ladefoged, ch. 14, pp 203-221.

Munhall, K. & Löfqvist, A. (1992) *Gestural Aggregation in Speech: laryngeal gestures*. Journal of Phonetics, 20, pp. 111-126.

Nagano-Madsen, Y. (1994) *Vowel Devoicing Rates in Japanese from a Sentence Corpus*. Department of Linguistics and Phonetics, Lund, Sweden, Working Papers 42, pp. 117-127.

Ohala, J. (1983) *The Origin of Sound Patterns in Vocal Tract Constraints*. MacNeilage, P. F. (ed) The Production of Speech, ch. 9, pp. 189-216.

Özen, E. (1985) *Untersuchungen zu einer kontrastiven Phonetik. Türkisch-Deutsch*. Wodarz, H. W. et al. (eds.) Forum Phoneticum, Bd. 30. Helmut Buske Verlag, Hamburg.

Rand, D. & Sankoff, D. (1990) *Gold Varb, Version 2, A Variable Rule Application for the Macintosh*. Centre de recherches mathematiques Universite de Montreal.

Sankoff, D. (1988) *Variable Rules*. Ammon, U., Dittmar, N., Mattheier K. J. (eds) Sociolinguistics: An International Handbook of Science and Language, Vol. 2, ch. 112, pp 984-997.

Sawashima, M. & Hirose, H. (1983) *Laryngeal Gestures in Speech Production*. MacNeilage, P. F. (ed.) The Production of Speech, ch. 2, pp. 11-38. Springer Verlag, New York, Heidelberg, Berlin.

Underhill, R. (1986) *Turkish*. Slobin, D. I. & Zimmer, K. (eds.) Typological Studies in Language: Studies in Turkish Linguistics, Vol. 8, pp. 7-21. John Benjamins Publishing Company.

Van der Hulst, H.; Van de Weijer, J. (1991) *Topics in Turkish Phonology*. Boeschoten, H. & Verhoeven, L. (eds.) Turkish Linguistics Today, pp. 11-59. E. J. Brill, Leiden, The Netherlands.

Yoshioka, A.; Löfqvist, A.; Hirose, H.; Collier, R. (1986) *How voiceless sound sequences are organized in terms of glottal opening gestures*. Annual Bulletin Research Institute of Logopedics and Phoniatrics, 20. University of Tokyo.

**Appendix A:**

| | | | |
|---|---|---|---|
| 1. pata | *'wave hand in greeting'* | patla | *'burst, torn open'* |
| 2. peki | *'very good'* | peklik | *'firmness'* |
| 3. potuk | *'puckered'* | potluk | *'bagginess in garment'* |
| 4. taka | *'small sailing boat'* | takla | *'somersault'* |
| 5. tekil | *'singular'* | teklil | *'crowning'* |
| 6. tepe | *'hill, summit'* | tepke | *'reflex'* |
| 7. tetik | *'quick, sharp'* | tetkik | *'close examination'* |
| 8. toka | *'shaking hands'* | tokta | *'to fix, settle, establish'* |
| 9. topu | *'in all'* | toplu | *'tidy (place); having a round head'* |
| 10. kak**i**r | *'dry, rustling'* | kakt**i**r | *'of'* |
| 11. kapan | *'to shut, close'* | kaptan | *'captain'* |
| 12. katan | *'loin, lumber'* | katlan | *'to bear, tolerate'* |
| 13. kekik | *'thyme'* | keklik | *'partrige'* |
| 14. köke | *'obsolete kind of ship'* | kökle | *'to tune'* |
| 15. kota | *'quota'* | kotra | *'cutter'* |
| 16. paket | *'package'* | pakla | *'to clean'* |
| 17. pasak | *'dirty untidy clothes'* | paskal | *'comic of old. Turk. theatre'* |
| 18. pesek | *'tartar (of teeth)'* | peste | *'pistachio nut'* |
| 19. peşiz | *'very small coin'* | peşkir | *'table napkin'* |
| 20. petek | *'honeycomb'* | petgir | *'hair sieve'* |
| 21. tapu | *'written survey of province'* | tapkur | *'row, line'* |
| 22. tatar | *'courier'* | tatla | *'to sweeten'* |
| 23. topak | *'roundish lump'* | toplat | *'of'* |
| 24. kepez | *'rock, cliff, hill'* | kepçe | *'laddle, scope'* |
| 25. ketal | *'starched'* | kettan | *'flax'* |
| 26. köpür | *'to froth, foam'* | köprü | *'(dent.) bridge'* |
| 27. kokart | *'cockade'* | koklat | *'cause to smell'* |
| 28. kopar | *'to pluck'* | kopsar | *'ring of iron'* |

**Appendix B:**

| | | | |
|---|---|---|---|
| 1. yat**i** | *'yacht (ACC)* | yatt**i** | *'lie down' (PAST)* |
| 2. bat**i*** | *'west'* | batt**i*** | *'sink' (PAST)* |
| 3. at**i*** | *'horse' (ACC)* | att**i*** | *'throw' (PAST)* |
| 4. ak**i*** | *'white' (ACC)* | akk**i*** | *'right' (ACC)* |
| 5. raket | *'raquet'* | takke | *'skull-cap'* |
| 6. sakal* | *'beard'* | bakkal* | *'grocer'* |
| 7. oka* | *'arrow'* | okka* | *'measure of weight'* |
| 8. leke | *'spot'* | Mekke | *'Mecca'* |
| 9. eti | *'meat' (ACC)* | etti | *'do' (PAST)* |
| 10. ete | *'meat' (DAT)* | ette | *'meat (LOCATIVE)* |
| 11. ata | *'horse' (DAT)* | atta | *'horse'* |
| 12. saate | *'clock' (DAT)* | saatte | *'clock' (LOCATIVE)* |

| | | | |
|---|---|---|---|
| 13. demete* | *'bunch' (DAT)* | demette* | *'bunch' (LOCATIVE)* |
| 14. ota | *'grass' (DAT)* | otta | *'grass' (LOCATIVE)* |
| 15. batar | *'sink'* | battaniye | *'blanket'* |
| 16. yata | *'yacht' (DAT)* | yatta | *'yacht' (LOCATIVE)* |
| 17. catal | *'fork'* | hatta | *'line' (LOCATIVE)* |
| 18. diken | *'thorn'* | sikke | *'Dervishe's cap'* |

*OSU Working Papers in Linguistics* 45 (85-97)

# Intervocalic consonant sequences in Korean*

**Keith Johnson, Mira Oh****
kjohnson@ling.ohio-state.edu

**Abstract:** This paper reports the results of an instrumental phonetic study of intervocalic consonant sequences in Korean. The study explored a putative positional neutralization produced at the phonetics/phonology interface. It was designed to determine whether Korean intervocalic laryngeal consonants are phonetically distinct from geminates, plain consonants, or laryngeal consonants in consonant clusters. The results showed that the contrast between intervocalic tensed singletons and geminates was neutralized, and that both of these patterned with heterorganic consonant sequences rather than plain singletons. Moreover, we found that this neutralization persisted across (limited) variation in speaking rate, although intervocalic tense consonants were more compressible in faster speech than were post-consonantal tense consonants.

## 1. Introduction

Informal listening tests, and some preliminary acoustic studies (Han, 1992), have suggested that the contrast between bare tense consonants and geminate tense consonants in Korean is neutralized intervocalically. For instance, [ik'i] 'moss' is neutralized with [ikk'i] 'being ripe' (which is composed of the morphemes /ik/ and /ki/). It has also been suggested (Iverson & Kim-Renaud, 1994) that in Korean there are two processes associated with speaking style which conspire to maintain this neutralization. In careful or expressive speech emphatic gemination gives [itt'a] from /it'a/ 'later', while geminate reduction is active in casual speech to produce [it'a] from /itt'a/ 'there is'.

In this study we explored these issues in an acoustic/phonetic analysis of Korean intervocalic consonants and consonant sequences, focusing on variation in speaking style and on the cross-speaker reliability of typical acoustic patterns.

The evidence shows that intervocalic tense consonants in Korean are phonologically geminates (at the output of the phonology) and that in fast speech geminates are more compressible than singletons - leading to the impression that there is a categorical process of geminate reduction. We conclude that 'geminate reduction' is a result of phonetic realization and not a categorical rule.

---

* An earlier version of this paper was presented at the meeting of the Linguistic Society of America, January 6, 1995. We gratefully acknowledge the input of Stuart Davis, Ken deJong, Sun-Ah Jun, Hyeon-Seok Kang, Joyce McDonough, and Bob Port.

** English Department, Yeo Joo Technical College, Kyunggi-do, Korea.

**2. Methods**

We measured vowel and consonant durations associated with intervocalic consonants and consonant sequences in Korean words produced by six native speakers of the Seoul dialect.

*2.1 Subjects.* Three female speakers (HO, MO, SI) and three male speakers (OJ, JC, MH) participated in the experiment. One subject was in his late twenties and the others were in their late thirties. The speakers reported no history of speech or hearing impairment.

*2.2 Materials.* We recorded productions of the words shown in Table 1 which illustrate intervocalic contrasts among lax and tense stops, fricatives, and affricates in Korean. These words are written in a broad phonetic transcription and do not reflect certain properties of the putative underlying representations of the morphemes.

| | Plain (C) | Tense (C') | Geminate (CC') |
|---|---|---|---|
| 1 | **sapuni** '4 minutes (nom)' | **sap'**uni 'lightly' | **sapp'**uni 'only shovel' |
| 2 | **ita** 'be' | **it'a** 'later' | **itt'a** 'there is' |
| 3 | tə**ita** 'be more' | tə**it'a** 'more later' | tə**itt'a** 'there is more' |
| 4 | cokim**ita** 'is a little' | cokim**it'a** 'a little later' | cokim**itt'a** 'there is a little' |
| 5 | **osak** nonword | **os'ak** 'a shiver' | **oss'ak** 'tailor's fee' |
| 6 | **iki** 'selfishness' | **ik'i** 'moss' | **ikk'i** 'being ripe' |
| 7 | **kaca** 'let's go' | **kac'a** 'fake' | **kacc'a** 'let's have' |

**Table 1** (continued).

| | Sonorant (RC') | NonSonorant (XC') |
|---|---|---|
| 1 | **samp'**uni 'only three' | **sakp'**uni 'fee only' |
| 2 | **ilt'**a 'to read' | **ipt'**a 'to wear' |
| 3 | tə**ilt'**a 'to read more' | tə**ipt'**a 'to wear more' |
| 4 | cokim**ilt'**a 'to read a little' | cokim**ipt'**a 'to wear a little' |
| 5 | **oms'**ak 'flinch' | **oks'**ang 'roof' |
| 6 | i**ŋk'**i 'popularity' | **ipk'**i 'wearing' |
| 7 | **kamc'**a 'let's wind' | **kapc'**a 'let's pay back' |

**Table 1**. Words examined in the study. On the left side of each cell the word is written in broad phonetic transcriptions (C' is used to transcribe the tense consonants) and on the right side of each cell is the English gloss.

We took duration measurements of the segments printed in bold face in the table. The words in the first column have plain, lax consonants, which in

...ic position are produced with voicing throughout the consonant closure. The words in the second column contain a bare tense consonant. The words in the third column contain a geminate tensed consonant. In these words the gemination occurs across a morpheme boundary. The words in the fourth column contain a sonorant/obstruent sequence and the words in the fifth column contain a sequence of heterorganic obstruents. In the last two columns the second consonant in the sequence is tense.

Each speaker read the words five times (in random order) for a total of 1050 tape-recorded tokens. The recordings were made at the Linguistics Laboratory at The Ohio State University. In each production the speaker read aloud a disambiguating meaningful utterance containing the target word and then the target word in isolation. We took measurements from the isolated word reading.

*2.3 Measurements.* Figure 1 illustrates the duration measurements that we took in this study. This figure shows a spectrogram and time-aligned acoustic waveform of the word [sap'uni] 'lightly'. The vertical cursors mark the consonant closure interval in [p']. Using such time-aligned waveform and spectrogram displays, we measured the duration of the vowel preceding the consonant of interest, the release phase of the consonant and, when possible, the closure interval of the sonorant or obstruent in the intervocalic clusters. Note that it was not possible to distinguish the closure intervals in nonhomorganic stop clusters.

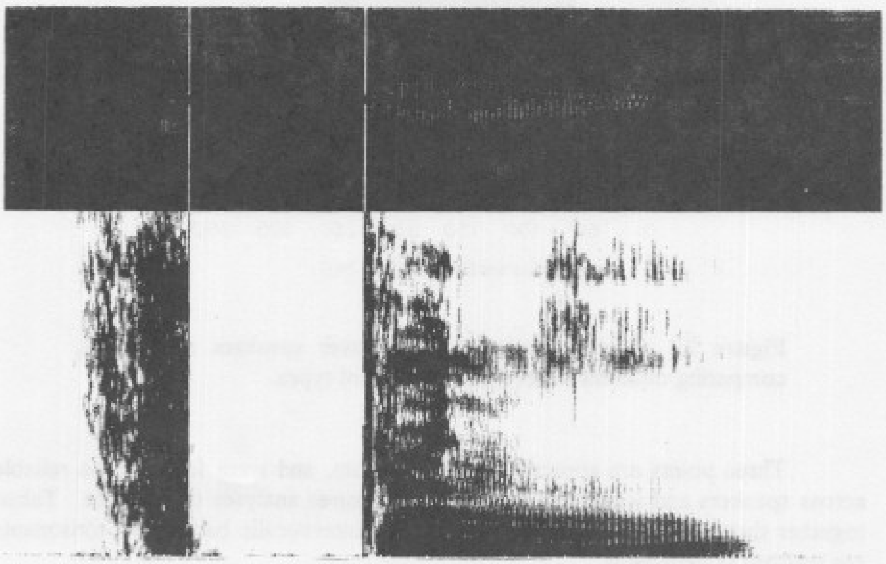


**Figure 1.** An example spectrogram and waveform display illustrating the consonant closure duration in [sap'uni] 'lightly'.

### 3. Intervocalic Bare Tense Consonants

Figure 2 shows results averaged over speakers and words. The horizont axis shows a time line that plots cumulative duration during the course of th word. The vowel portion of each word (the unfilled portion of each bar) starts a 0. Then, for words that had a non-identical sequence of sounds, the X or R interval is shown with light-hatch fill. The dark-hatched portion of each bar shows the interval of the consonant closure of the lax or tense consonant, and the filled portion of each bar shows the release interval. The horizontal bars show the different types of intervocalic consonants. Starting from the top, C stands for the plain lax consonants, C' stands for the bare tense consonants, CC' stands for the geminate tense consonants, RC' stands for the sonorant-obstruent sequence, and XC' stands for the heterorganic obstruent sequence.

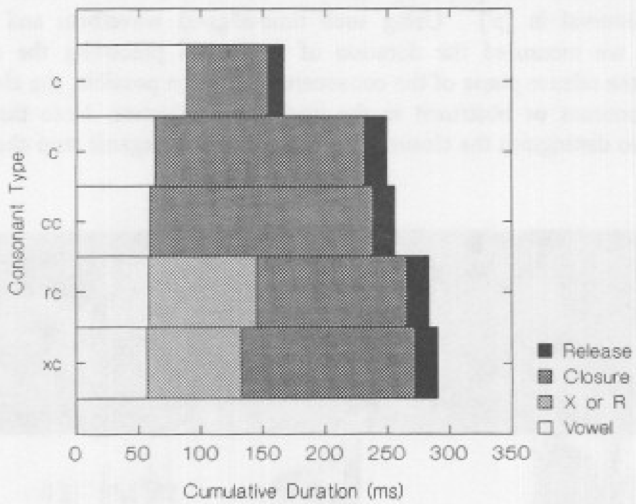



**Figure 2**. Overall results averaged over speakers and words, comparing different intervocalic consonant types.

Three points are apparent from these data, and were found to be reliable across speakers and word-sets in repeated measures analyses of variance. Taken together these three observations suggest that intervocalic bare tense consonants are realized as geminates.

*First*, vowels preceding lax consonants were longer than vowels preceding any of the other consonant types. (There was a main effect of consonant type on vowel duration [$F(4,20)=70.752$, $p<0.01$] and a post-hoc comparison of means ound that vowels before lax consonants were longer than the other vowels which lid not differ from each other. This is illustrated by a list of the different consonant types where underlining indicates the consonant types that had comparable vowel durations. (See Table 2. In this section we are discussing the 'all speakers' row of the table. We will return to speakers OJ and SI in the next section.) In particular, we find it interesting that vowels before bare tense consonants (C') patterned with vowels before consonant sequences. It might be argued that vowels before lax consonants are longer because the lax consonants are voiced; and thus follow a well-known cross-linguistic tendency for vowels to be longer before voiced consonants than before voiceless ones. However, the fact that vowels before sonorant/obstruent sequences are short suggests that voicing is not the relevant factor. The relevant generalization seems to be that vowels are short before consonant sequences, provided we consider the bare tense consonants to be sequences.

| | vowel duration | total C closure | consonant closure |
|---|---|---|---|
| all speakers | $\underline{\text{rc' xc' cc' c'}}$ c | c $\underline{\text{c' cc'}}$ $\underline{\text{rc' xc'}}$ | c rc' $\underline{\text{c' cc'}}$ |
| speaker SI | $\underline{\text{rc' xc' cc' c'}}$ c | c $\underline{\text{c' cc'}}$ $\underline{\text{rc' xc'}}$ | c rc' $\underline{\text{c' cc'}}$ |
| speaker OJ | $\underline{\text{rc' xc' cc' c'}}$ c | c $\underline{\text{c' cc'}}$ $\underline{\text{rc' xc'}}$ | c rc' $\underline{\text{c' cc'}}$ |

**Table 2**. Results of Bonferroni post-hoc comparisons of means (ordered from shortest to longest). Labels for the consonant types are as given in Table 1. Consonant types that are connected by a line were not reliably different on a given measure. The first row shows results of repeated-measures analyses of variance of the data pooled across speakers, while the second and third rows show results for two selected speakers.

*Second*, there was a two-way split in total consonant sequence closure duration, which in the RC' and XC' sequences is the combination of both the light- and dark-hatched portions of the bars. Plain lax consonants have short closure durations while the other consonant types have long total closure durations. (There was a main effect of consonant type on total closure duration [$F(4,20)=89.563$, $p<0.01$] and a Bonferroni post-hoc comparison of means gave the results shown in Table 2, top row, second column.) One point of interest here is that there was no reliable difference in the durations of the geminate tense consonants and the heterosegmental sequences. That is, total closure duration in CC' is not statistically different from total closure duration in RC' or XC'. As with the vowel duration data, the bare tense consonants patterned with the consonant sequences and not with the plain lax consonant. Total closure duration in the C' words was not reliably different from the total closure duration in the CC' words.

*Third*, in addition to the two-way split in total consonant sequence closure duration just discussed, there is a three-way split in the test-consonant closure duration (the portion of the bars marked with dark-hatching). In this analysis we found that the closure duration of the plain lax consonant was shorter than t closure duration in the sonorant/obstruent sequence, which in turn was short than the closure duration in the geminate and bare tense consonants. Note tha because measurement of the closure interval was only rarely possible with th heterorganic obstruent sequences we did not include the XC' words in this analysis. (There was a main effect for consonant type [$F(3,15)=59.33$, $p<0.01$] and a Bonferroni post-hoc comparison of means gave the results shown in Table 2, third column, first row.) The tense consonants in sonorant/obstruent sequences are by all accounts singletons. Therefore, this comparison suggests that closure duration in tense consonants are inherently longer than in lax consonants: a phonetic fact about the realization of tense consonants. The comparison also suggests that closure duration in intervocalic bare tense consonants is longer than in singleton tense consonants (the C' in RC'). We take this to reflect the (surface) phonological representation, namely that tense consonants are geminates in intervocalic position.

**4. Emphatic Gemination and Geminate Reduction**

These data show that intervocalic bare tense consonants and geminate tense consonants did not differ phonetically. However, several authors have suggested that one or both of these consonant types may be realized as geminates in careful speech or as singletons in casual speech. For instance, Iverson & Kim-Renaud (1994) adopt an analysis in which bare tense consonants and geminate tense consonants are neutralized, but may be realized either as geminates or as singletons depending on speaking style. In their analysis, a process of geminate reduction (1) affects geminate tense consonants in casual speech, and a process of emphatic gemination (2) affects bare tense consonants in careful speech, yielding variable, but always neutralized, realizations as in (3).

```
(1) Geminate Reduction          (2) Emphatic Gemination
    X X  =>  X                      X   =>   X X
    \ /      |                      |        \ /
     c       c                      c         c
```

(3) Careful speech . . . . . . Casual speech

akk'i . . . . . . . . . . . . . . . . ak'i    /ak'i/ 'to hold dear'

akk'i . . . . . . . . . . . . . . . . ak'i    /ak-k'i/ 'instrument'

We were able to provide a preliminary test of this analysis because our speakers adopted different speaking styles.

Figure 3 shows average segment durations indicating rate-of-speech differences among our speakers. Speaker OJ read the words more quickly while speaker SI adopted a slower, more careful rate. Assuming processes of geminate reduction and emphatic gemination we predict that speaker OJ is more likely to have produced the intervocalic tense consonants as singletons while speaker SI is more likely to have produced them as geminates.

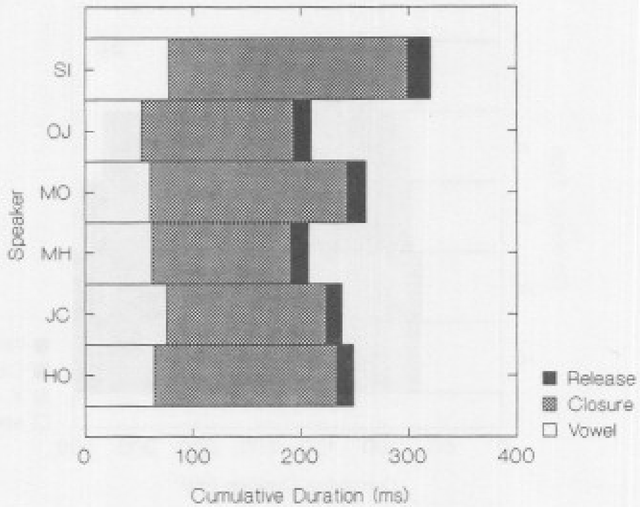



**Figure 3.** Vowel, consonant closure, and release durations by speaker averaged over words and consonant types. These data indicate differences among the speakers in rate of speech.

Figure 4 shows duration data for speaker SI, and Figure 5 shows the results for speaker OJ. Both speakers show about the same pattern of durations that we found in the overall data (as one would predict given the results of our statistical analyses which tested for the consistency of the patterns across speakers). In both fast and slow speech, the intervocalic tense consonants behaved like geminates. They were preceded by short vowels, had closure intervals which were comparable to the interval occupied by a two consonant sequence, and had closure intervals that were longer than those found in post-consonantal tense consonants (see the results of post-hoc tests shown in the second and third rows of Table 2). There is one difference between the speakers to which we will return below. Speaker OJ produced the C' and CC' words with shorter total closure durations than in the XC' and RC' sequences.

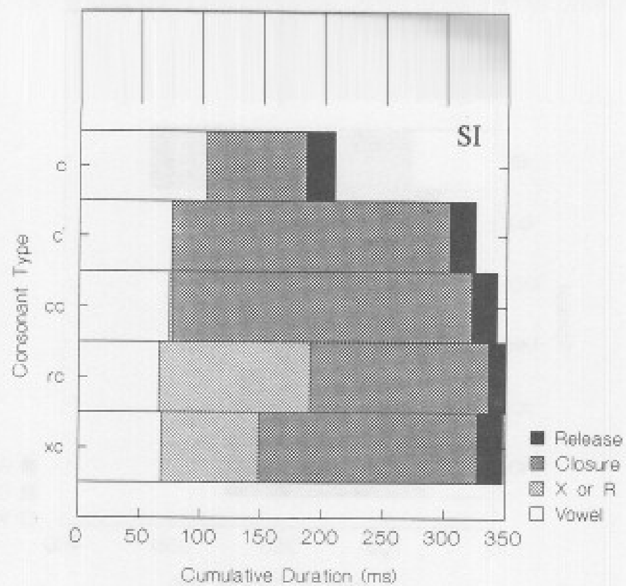



**Figure 4**. Duration results (as in Figure 3) for speaker SI.

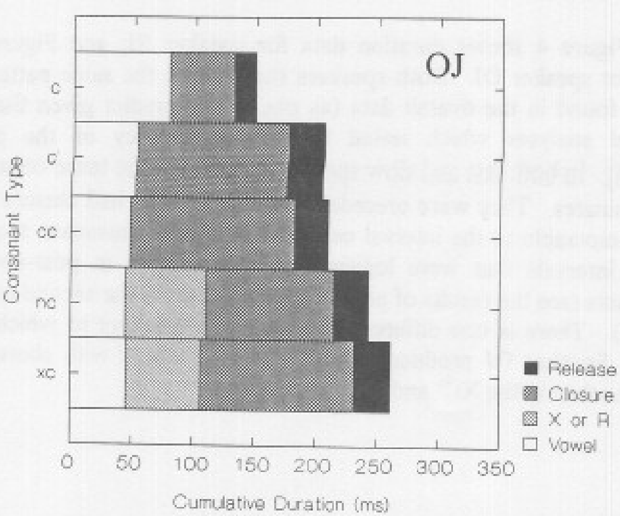



**Figure 5**. Duration results (as in Figure 3) for speaker OJ.

Another prediction of the emphatic gemination/geminate reduction analysis f speaking style variation is that the relative durations of tense consonants will fall ı a bimodal distribution. That is: durations will tend to be either long or short vith no intermediate values, because in any set of data within which there is some variation of speaking style we expect to find examples of both geminate and nongeminate tense consonants. To test this prediction we computed histograms of the relative closure durations of bare and geminate tense consonants. To control for speaking rate, we defined relative duration as the ratio the closure duration to the total duration of VC sequence.

Figure 6 is an illustration of a bimodal distribution of the consonant closure duration data from the lax consonants (C) and the geminate tense stops (CC'). Relative duration is shown on the horizontal axis and the bars represent the number of tokens that had a particular relative closure duration. The distribution has two peaks, one for the lax consonants and one for the geminate tense consonants. In a somewhat literalistic interpretation of the durational values of timing slots we could say that we have a group of tokens with one slot on the timing tier (the lax consonants) and another group of tokens with two slots on the timing tier (the geminate tense consonants).

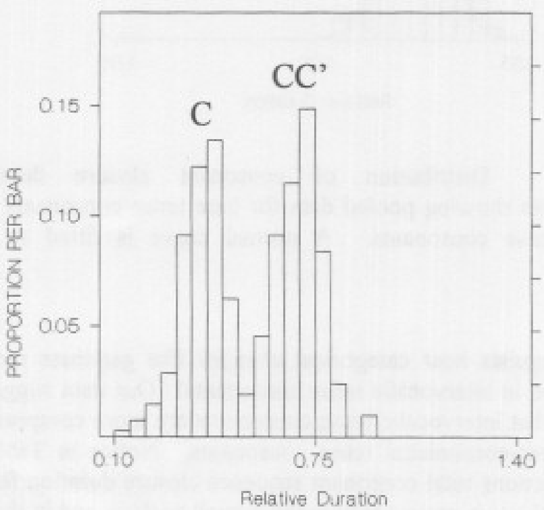



**Figure 6**. Distribution of consonant closure duration relative to total duration of the VC sequence for pooled data from the lax consonant and geminate tense consonants. These data clearly fall in two groups; one for C and one for CC'.

Figure 7 shows a similar plot of the relative closure durations of bare ten consonants and geminate tense consonants. This plot shows that there was n tendency for a bimodal distribution for these consonant types. Therefore we hav no evidence in favor of analyzing speaking style variation in Korean usin categorical rules like emphatic gemination and geminate reduction.

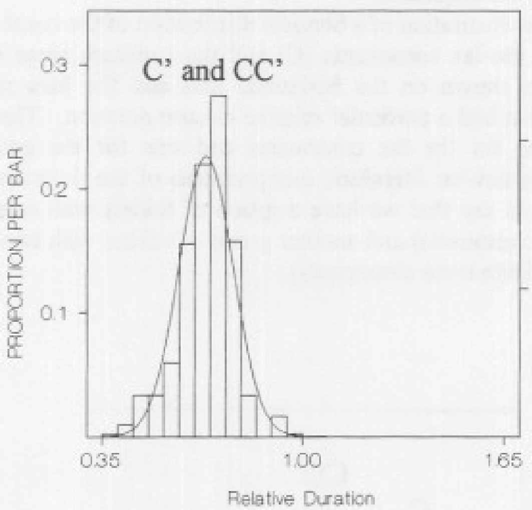



**Figure** 7. Distribution of consonant closure duration measurements showing pooled data for bare tense consonants and geminate tense consonants. A normal curve is fitted to the distribution.

Why do linguists hear categorical changes like geminate reduction and emphatic gemination in intervocalic tense consonants? Our data suggest that one possible answer is that intervocalic tense consonants are more compressible in fast speech than are post-consonantal tense consonants. Notice in Table 2 that in speaker OJ's productions total consonant sequence closure duration fell into three groups rather than the two groups seen in the overall analysis and in the analysis of speaker SI's productions. Comparing Figures 4 and 5 we see that C' and CC' total closure durations for speaker OJ were on average about 50 ms shorter than were the total closure durations in the RC' and XC' sequences. SI did not show this distinction between the consonant types. Apparently, although speaker OJ maintained a contrast between the consonant closure duration for C'/CC' and the consonant closure in RC' sequences (third column of Table 2) - which, along with ne vowel duration data, is evidence that the intervocalic tense consonants emained geminates - at his faster rate-of-speech the intervocalic tense consonants were more compressible than were the intervocalic consonant sequences RC' and XC'.

We investigated this compressibility explanation further by comparing consonant closure durations of intervocalic tense consonants and post-consonantal tense consonants. Figure 8 shows the difference between the average consonant closure duration in the C' and the average duration of post-consonantal tense consonants (the C' of RC'), for each speaker. The speakers are ordered from slowest (SI) to fastest (OJ). For speaker SI, closure duration in C' was about 80 ms longer than C' closure duration in the RC' sequence, while for speaker OJ the difference was only 20 ms, but still reliably different. We see in this figure a good correlation between speaking rate and the difference between closure durations in the bare C' and C' in the RC' sequences. As speaking rate increased the difference decreased. The results for the closure durations in geminate tense consonants were very similar (Figure 9). This pattern of results indicates that as speaking rate increased geminate tense consonants (taken here to include both C' and CC') shrank more quickly (were more compressible) than post-consonantal tense consonants.

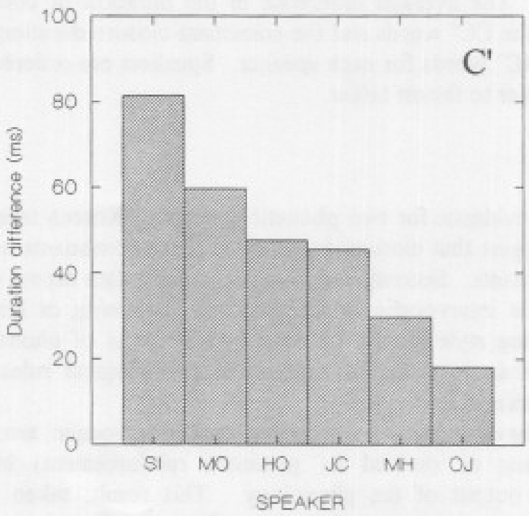



**Figure 8**. The average difference in the duration of consonant closure in the C' words and the consonant closure duration in the C' of the RC' words for each speaker. Speakers are ordered from slowest talker to fastest talker.

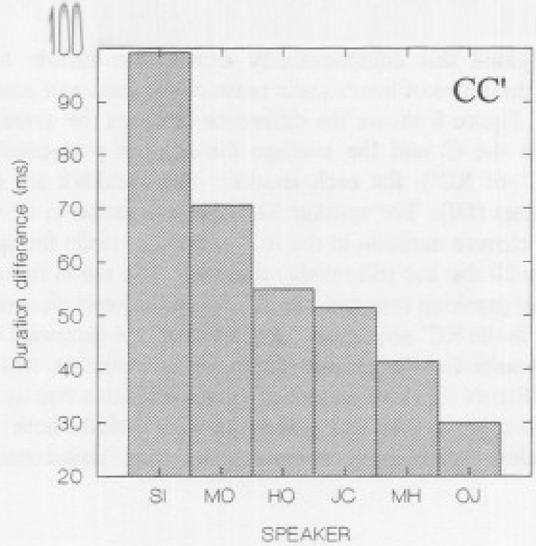



**Figure 9**. The average difference in the duration of consonant closure in the CC' words and the consonant closure duration in the C' of the RC' words for each speaker. Speakers are ordered from slowest talker to fastest talker.

**5. Conclusions**

We found evidence for two phonetic aspects of Korean tense consonants. First, our data suggest that closure durations in tense consonants are longer than those in lax consonants. Second, we have preliminary data across speaking rates which suggests that intervocalic tense consonant shortening or lengthening as a function of speaking style should be described in terms of phonetic realization processes rather than in terms of categorical phonological rules of Emphatic Gemination or Geminate Reduction.

We also have evidence suggesting that intervocalic tense consonants (whether underlying or derived by geminate reinforcement) in Korean are geminates at the output of the phonology. This result, taken together with previous research, suggests that the inventory of intervocalic consonants in Korean includes lax (C), and geminated tense (CC') consonants but no tense singletons (C'), while in initial position the inventory includes lax (C) and tense (C') consonants but no tense geminates (CC'). Putative phonological processes such as geminate reinforcement (CC => CC') and tense consonant gemination (C' => CC') conspire to limit the number possible realizations of intervocalic consonants to a set of easily perceived contrasts, at the cost of the resulting homophony of certain forms such as [ikk'i] 'moss' and [ikk'i] 'being ripe'.

## References


Han, J. (1992) On the Korean tensed consonants and tensification. *Chicago Linguistic Society*, **28**, 206-223.

Iverson, K. and Y. Kim-Renaud (1994) Phonological incorporation of the Korean glottal approximant. paper presented at the Ninth International Conference of Korean Linguistics, London.

*OSU Working Papers in Linguistics* 45 (98-130)

# Acoustic and Intonational Correlates of Informational Status of Referring Expressions in Standard Korean

**Hyeon-Seok Kang**
hkang@ling.ohio-state.edu

**Abstract**: This paper examines the phonetic correlates of informational status of referring expressions in Standard Korean. Two experiments are conducted. The results show that speakers of Standard Korean indicate the informational status of referring expressions not only with acoustic cues such as amplitude and duration but also with intonational phrasing. The results also suggest that speakers of Standard Korean rather clearly distinguish "current" and "displaced" entities. The term 'intonational attenuation' is proposed to describe the use of intonational phrasing by some languages to indicate the given status of referring expressions.

## 1. Introduction

As previous studies (e.g. Halliday 1967, Chafe 1972 & 1976, Brown 1983, Terken 1984, Horne 1991, Fowler & Housum 1987) have suggested, new and given referring expressions are likely to be produced by speakers in a different manner both syntactically and phonetically. Syntactically it was shown for such languages as English and Dutch that given entities[1] are more likely to be referred to by definite expressions — such as *the definite article + noun* form or a pronominal form — than "new" entities. It was also suggested that there are differences between new and given linguistic entities in their phonetic realizations. Previous studies have suggested f0 movement, f0 value, amplitude and duration as possible phonetic cues for distinguishing new and given linguistic entities. For instance, Halliday (1967) claimed that given and new information are distinguished primarily by pitch prominence. He made a suggestion that in English, new information is given pitch accent while old information is not. A similar suggestion was made by Terken (1984) for Dutch. He (1984) found in his Dutch data that new linguistic entities are generally more often pitch-accented than given linguistic entities. Brown (1983) makes a rather different suggestion. She suggests that new linguistic entities are generally produced with a higher f0 than given linguistic entities.

On the other hand, Chafe (1972, 1976) suggested that informational status is cued both by f0 value and amplitude in English, stating that " ... given information is pronounced with lower pitch and weaker stress than new (information)..."(1976:31). However, the claims of these researchers are either based on subjective judgments or on the measurements of just one phonetic cue — f0 value. The study that investigated the various possible phonetic cues of "given" vs. "new" status by instrumental measurements and attempted to determine which of them are significant cues was Fowler & Housum's (1987) study. They

[1] Prince (1981:235) defines an "entity" as a discourse-model object which may represent an individual (existent or not in the real world), a class of individuals, an exemplar, a substance, or a concept. I adopt her definition in this paper. I will also use the term "linguistic entities" in this paper to indicate those expressions that refer to "entities".

compared duration, amplitude and f0 (peak value of the stressed syllable) between new and given referring expressions. They report that only duration is a "reliable" cue for distinguishing new and given referring expressions in English.

Fowler & Housum (1987) used natural (or close-to natural) speech data (a monologue from a radio program titled "Prairie Home Companion" and interview dialogues from "the MacNeil-Lehrer News") for their study. The use of natural speech data has both merits and defects in this type of study. The merits are that the investigator is able to analyze the data produced by speakers when they actually try to communicate something to listeners and that the experimenter doesn't have to worry about the possible distortions of the results for the experiment caused by the subjects' guess on the experiment's purpose. The defects of the use of natural speech data are also significant.

These defects follow from the fact that the investigator cannot exercise any control over the talkers' use of new and given information. The investigator's inability to exercise control over the talkers' speech production results in two major problems. The first is that the speaker can use phonetic features for purposes other than cuing informational status of linguistic entities. In other words, speakers can use voice pitch, amplitude and duration for other attention-getting purposes — for focal or contrastive purposes, or for holding the floor or for directing the flow of information (Lehman 1977).

The second defect seems to be more serious. It is known that there is a declination phenomenon in f0 toward the end of an intonational phrase in English (Pierrehumbert 1979, Maeda 1976), in Japanese (Fujisaki et al. 1979, Pierrehumbert & Beckman 1988), in Dutch (Cohen & 't Hart 1967), and in Korean (Koo 1986, Ko 1988) — possibly universally (Vassiere 1983). The amplitude early in a sentence is known to be greater than that at the end in English (Ohala 1977, Pierrehumbert 1979) and French (Vassiere 1983) and also possibly universally. Further, it is well known that in languages such as English, French and Korean, vowel duration is significantly lengthened at the end of an intonational phrase (cf. Oller 1973, Delattre 1966, Koo 1986). This suggests that it is very problematic to compare the values of duration, f0 and amplitude between linguistic entities that occur at different sentential positions as Fowler & Housum (1987) did.

Another problem in Fowler & Housum's analysis is that they disregard the fact that in English, which is a pitch accent language, not just higher pitch but extended pitch movement can also contribute to perceptual prominence of linguistic entities. This is a same type of error Brown (1983) makes in her study. They further oversee the fact that in English there are various types of pitch accents. As Beckman & Pierrehumbert (1986:256) suggest, there are 6 different types of pitch accents in English — not only H*, L+H*, H*+L accents but L*, L*+H and H+L* accents[2], all of which can make the linguistic entity perceptually pitch-prominent. This wide variety of pitch accents in English suggests that it is possible to compare the f0 values of the stressed syllables of new and given linguistic entities only when they are accented with identical pitch accents, i.e. only when the stressed syllables have the same type of pitch accent. Further it follows that a simple measurement of the f0 value of the stressed syllable cannot reveal whether that syllable is pitch accented (or pitch-prominent) or not. This is another serious defect in the method adopted in Fowler & Housum's study.

Because of these methodological problems mentioned above, their claim that duration is the sole "reliable" cue distinguishing new and given entities leaves not a little room for doubt. With these problems in mind, in this paper, I will examine how linguistic entities of different informational status are phonetically cued by

[2] The * indicates that the tone associated to the star is linked to and realized at the stressed syllable.

speakers of Standard Korean. An attempt will be made, using necessary contro to identify which phonetic cues are exploited by speakers of Standard Korean distinguish new and given linguistic entities.

This paper has four specific sub-goals. The primary goal is, as mention above, to identify the phonetic cue(s) used by Standard Korean speakers t distinguish entities of new and given status. Since Standard Korean does not hav a pitch accent system, I will consider f0 value, amplitude, duration and intonation phrasing as four possible cues and examine which of them are significant. Th motivation to include intonational phrasing as one of the potential cues is provide by Jun's (1993) suggestion that given information and new information are likely to be produced with different intonational phrasing.

As suggested earlier and also by previous studies (e.g. Terken 1984, Brown 1983), speakers of various languages tend to use both phonetic and syntactic cues to distinguish given and new linguistic entities. The second sub-goal of this paper is the examination of the interaction between these two types of cues. I will examine whether phonetic cues are given by speakers of Standard Korean when a syntactic cue is provided for the identification of informational status of linguistic entities. This sub-goal is pursued in Study 2.

Brown & Yule (1983) further divide given entities into "current evoked entities" and "displaced evoked entities" based on how recently a given entity is introduced into the discourse. The third sub-goal is to examine whether there is any acoustic or prosodic basis in Brown & Yule's (1983) division of given entities into these two categories. This is examined in Study 3. The results of this investigation will be compared to Brown's (1983) study, which suggested that English speakers do not distinguish current and displaced evoked entities by pitch, though the distinction between the two may be marked by syntactic cues.

Jun (1993) claims that informational status interacts with other factors in affecting how Korean speakers intonationally phrase linguistic entities. She suggests that phonological weight of linguistic entities and the speaker's speech rate also play an important role in intonational phrasing. The examination of the interaction among these three factors is another sub-goal of this paper.

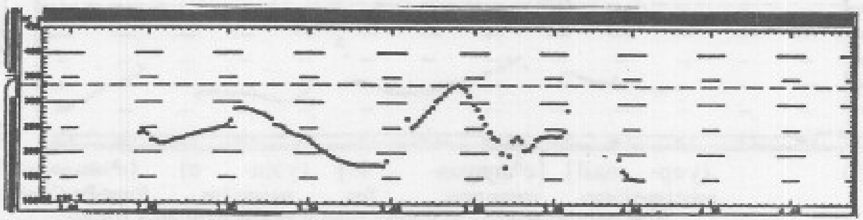
From the judgment that the use of natural speech data has much more significant problems than advantages for the purposes of this study and that the defects of a controlled experiment can be minimized with proper methods, three data sets were constructed and two controlled experiments were conducted. These experiments will be described in Section 3

## 2. A Brief Sketch of the Intonational Structure of Standard Korean

Before the detailed description of the two experiments in Section 3, a brief explanation of the intonational structure of Standard Korean seems to be in order because knowledge on the intonational structure of Standard Korean will be essential in understanding the method and results of the two experiments.

Intonation of Standard Korean is hierarchically organized like in other languages such as English (Pierrehumbert 1980), French (Hirst & di Cristo 1984, di Cristo 1978) or Japanese (Beckman & Pierrehumbert 1986). That is, a Standard Korean sentence consists of one or more intonational phrases; an intonational phrase usually consists of more than one accentual phrase; and an accentual phrase can have more than one word. Thus, the f0 contour of the intonational phrase is realized with the tonal patterns of one or more accentual phrases and the intonational phrase boundary tone.

The underlying tonal pattern of the accentual phrase in Standard Korean is LHLH[3] (Jun 1993). The first accentual phrase of the sentence in Figure 1 shows the typical tonal patterns of Standard Korean (The brackets in the Korean sentences in the figures below indicate that the word or phrase in the brackets forms one accentual phrase).

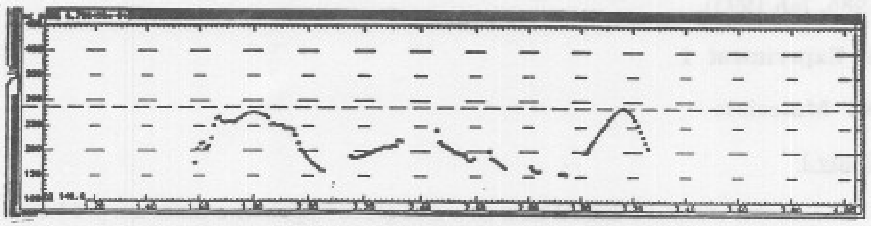


(Hz) {pukmi- san yəu- ka} {na-n coh -a}
North American-produced fox-Subj I-Top like-Decl
'I like North American foxes.'

L H L H L H L%

Figure 1: F0 contour of a sentence produced in two accentual phrases

However, the first H tone and the second L tone often undershoot when the accentual phrase is short (Lee 1989, Jun 1993) — i.e. in most cases when the accentual phrase consists of one to three syllables (sometimes four) as illustrated in Figure 2, where undershoot is observed in the first two accentual phrases of the sentence.



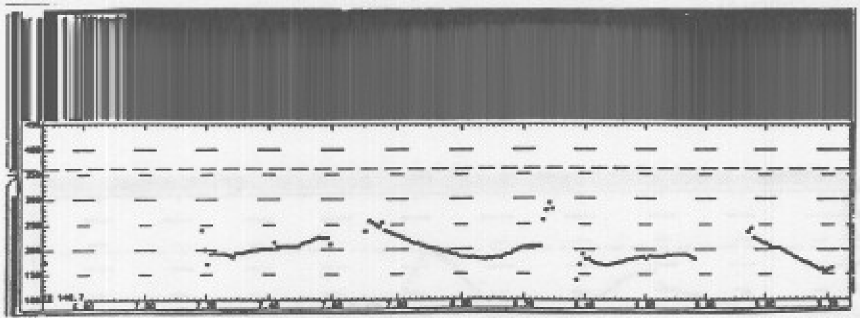
(Hz) {atamha-n}{pyəlcaŋ-il} {ci-ə poci kɨr-æ}
little-Adj P villa-Acc build why don't you
'Why don't you build a little villa?'

L H L H LH L HL%

Figure 2: Fo contour of a sentence whose first two accentual phrases have an undershoot of the first H tone

[3]For a somewhat different view, see de Jong (1989) and Lee (1989).

Further when an aspirated or glottalized consonant begins an accentual phrase, the first L tone is replaced by a H tone (Jun 1993) as shown in Figure 3.



(Hz) {yəp- mail} {cʰəŋnyən- ɨn} {yəin- e} {cʰinku-yə-ss-ko}
neighbor-village- young-man- Top woman-Pos friend-Pst-Cop-Pst-and
'The neighbor-village young man was the woman's friend'

L H H L H% L H H L%

Figure 3: F0 contour of a sentence where H tones begin a new accentual phrases

Standard Korean has at least 6 intonational phrase boundary tones including L, H, LH, HL, LHL and HLH (Koo 1986, Jun 1993). When the accentual phrase boundary coincides with an intonational phrase boundary, the final H tone of the accentual phrase is preempted by an intonational phrase boundary tone (Jun 1993), as shown in Figure 2, where the accentual phrase final H tone is replaced by the intonational phrase boundary tone HL%; and in Figure 3, where the H tone of the second accentual phrase was preempted by boundary tone H% and the H tone of the last accentual phrase was replaced by boundary tone L%. The syllable where the intonational phrase boundary tone is realized becomes noticeably lengthened as in languages like English or French, and optionally followed by a pause (Koo 1986, Jun 1993).

## 3. Experiment 1

### 3.1 Materials

Study 1

In this study, I examine how new and given linguistic entities are produced acoustically and prosodically in Standard Korean. The new linguistic entity is, here, defined as a referring expression that introduces an entity into the discourse; and the given linguistic entity, as an expression that refers to the entity later in the discourse. The purpose of Study 1 is to identify which phonetic cue(s) Korean speakers use to distinguish new and given entities.

In Korean, a noun phrase usually appears without any determiner. Phrases like *cohɨn saɹam* "good man" and *cəlmin yəca* "young girl" are perfectly grammatical in Korean. Sometimes, however, the determiner *kɨ* "the" can precede the noun phrase and show that the following noun phrase is a given expression. That is, *kɨ* can function as a syntactic cue indicating that the following noun phrase is given information. As mentioned earlier, the scope of the present research is limited to the examination of referring expressions. The referring expressions in the corpus are all composed of one adjectival word and a noun. This is to examine Jun's (1993:199) suggestion that new information is apt to be produced in two separate accentual phrases when they are composed of two words, while given information tends to form one accentual phrase.

This experiment, in addition, attempts to examine the interaction of the informational status of entities with those factors that Jun (1993) suggested affect accentual phrasing of linguistic entities — i.e. speech rate and phonological weight. As observed earlier, Jun (1993) claims that speech rate affects accentual phrasing. This claim is in line with the claim by Selkirk (1984) and Nespor & Vogel (1986) that speech rate affects the phrasing of intonational phrases and also with Vassiere's (1983) suggestion that speech rate influences the number of prosodic words in the utterance.

Jun also claims that a phrase tends to be more often produced in two accentual phrases as the phonological weight of the phrase becomes heavier, and that phrases whose phonological weight is heavier than 6 syllables are generally produced in two accentual phrases. In order to examine how the informational status of entities and the phonological weight factor affects accentual phrasing of the referring expressions, 9 four-turn dialogues were constructed where 4- 5- and 6-syllable referring expressions appear in the discourse both in new and given status attached by a grammatical particle. The 4-syllable referring expressions consist of a 2-syllable adjectival word + a 2-syllable noun, while the 5-syllable expressions and 6-syllable expressions were composed of a 3-syllable adjectival word + a 2-syllable noun and a 3-syllable adjectival word + a 3-syllable noun, respectively. These referring expressions are listed in Table 1.

| No. of syllables | Phrase | Gloss |
|---|---|---|
| 4 | *nəlpin maɨm* | 'broad mind' |
| 4 | *yep'ɨ-n inhyəŋ* | 'pretty doll' |
| 4 | *koun noɹæ* | 'pretty song' |
| 5 | *pukmisan yəu* | 'North-American fox' |
| 5 | *wantosan iŋə* | 'Wanto-produced carps' |
| 5 | *atamhan pyəlcaŋ* | 'little villa' |
| 6 | *siwənha-n əlɨmmul* | 'cool icewater' |
| 6 | *kapyəun taliki* | 'light running' |
| 6 | *səlikɨn panana* | 'unripe banana' |

Table 1. The list of the noun phrases used for study 1

For a reliable analysis of f0, each referring expression was constructed so that a majority of the segments could be sonorants. Since each referring expression forms a bigger phrase combined with a particle, the phrases have 5, 6 or 7 syllables as their phonological weight. A sample dialogue (where a 6-syllable phrase — 5-syllable referring expression + particle — appears) is given below.

A. ipən kail-e-nɨn muənka-lɨl hæ po-ko sipʰ-inte
this fall-Loc-Top something-Obj do try-Adv P want to-Decl
"I want to do something this fall."

B. atamha-n pyəlcaŋ-ɨl ci-ə po-ci kiɹæ?
little-Adj P villa-Obj build-Adv P why don't you
"Why don't you build a little villa?"

A. kɨkə kwenchanh-ɨn sæŋkak-i-nte.
that good-Adj P idea-Cop-Int
"That sounds like a good idea."

B. atamha-n pyalcaŋ-i iss-imyən ne maim-to pʰukinha-lkəya.
little-Adj P villa-Subj present-if your heart-also warm-Decl
"If you have a little villa, your heart will also become warm (and cosy)."
(P: particle)

An attempt was made to distinguish focus from new information. This i because new information is not identical to focus in its narrow sense. Dialogue were constructed so that the new linguistic entities do not appear in the discourse a an answer to a wh-question. Care was also taken so that given linguistic entities dc not receive any contrastive focus in the dialogue. The sentential position where new and given entities appear was controlled for subject position only, which would enable us to compare f0 value, amplitude and duration of a new entity and its given counterpart. The dialogues constructed for Study 1 are attached as Appendix 1.

Study 2

In this study, I investigate how the same referent is produced acoustically and prosodically in its occurrence as a new and a given entity when a syntactic cue is already given by the speaker. The referent is introduced into the discourse in the form *han* 'a' + noun and referred to later with determiner *ki* 'the' + noun. Only the 2-syllable nouns were chosen for this study. The 6 nouns used for Study 2 are listed in Table 2. For this study, 6 pairs of three-turn dialogue were constructed. Example dialogues are given below:

as a new entity

B. na yocim noɹæ hana-l mantil-ko iss-ə.
I nowadays song one-Obj make-Adv P Aux-Decl
"I am writing a song these days."

A. ət'ə-n noɹæ-nte?
what kind of song-Int
"What kind of song is it?"

B. han inə-ka waŋca-eke saɹaŋ-il pʰyohyənha-nin noɹæ-ya.
a mermaid-Subj prince-Dat love-Obj express-Adj P song-Decl
"A mermaid expresses her love for a prince in that song."

as a given entity

B. na yocim han inə-e-tæhan noɹæ-l mantil-ko iss-e.
I nowadays a mermaid-about song-Obj make-Adv P Aux-Decl
"I am writing a song about a mermaid these days."

A. ət'ə-n næyoŋ-i-nte?
what content-Cop-Int
"What is the content of the song like?"

B. ki inə-ka waŋca-eke saɹaŋ-il pʰyohyənha-nin noɹæ-ya.
the mermaid-Subj prince-Dat love-Obj express-Adj P song-Decl
"The mermaid expresses her love for a prince in that song."

As in the first corpus, the new and the given expression appear only at the subject position of the sentence. The dialogues constructed for Study 2 are attached as Appendix 2.

| word | gloss |
| --- | --- |
| *inə* | 'mermaid' |
| *mæŋin* | 'blind man' |
| *yəŋuŋ* | 'hero' |
| *kunin* | 'soldier' |
| *yəu* | 'fox' |
| *manyə* | 'witch' |

Table 2. The list of nouns used for study 2

### 3.2 Methods

Subjects

Three male and three female native speakers of Seoul Korean participated in the experiment. The subjects were all in their late twenties or early thirties. The three male speakers were OSU graduate students majoring in social or biological sciences. One female speaker was a linguistics graduate student and the other two female speakers had master degrees in other areas but were not students at the time of the experiment. All the speakers had fairly close friendships of at least a year's standing with the author. All were naive to the purpose of the experiment.

Procedures

Recordings were made in a sound-treated booth in the Linguistics Laboratory at the Linguistics Department of OSU. The constructed dialogues for Study 1 and Study 2 were 9 and 12 respectively. 18 foil dialogues were also constructed. Each of these dialogues was written on a card. The cards were randomized. There was one reading session before the actual recording. The subject was asked to read the B sentences trying to understand the content of the dialogue while the author read the A sentences (cf. Appendix 1, 2). This session was devised to make the subject understand the contents of the dialogues and make the recording session approximate a conversational situation. Then the following two reading sessions were recorded. The subjects were asked to read as if they were engaged in a real conversation. When the subjects made an error in their reading, they were not asked to reread. This was because speakers are apt to emphasize what they produced wrongly in their second production. For the same reason, self-corrected items were not included either as data for analysis.

Among the recorded sentences, those of interest were digitized. Measurements were made using the Waves program (Version 5.0) developed by Entropics Inc. First of all, the intonational pattern of the sentence in which a linguistic entity of interest appears was analyzed. The general intonational and accentual phrasing was observed. And then the accentual phrasing of the linguistic entity was examined and also the existence of a boundary tone at the end of the entity was checked. Duration was measured using wide band spectrograms. When the expression begins with a stop, its duration was measured from the release of the closure. For Study 1, the duration of the noun phrase excluding the particle (i.e. adjectival + noun) was measured. For Study 2, durations of the nouns were measured and compared. This was because there is an inherent difference in duration among the particles and also between two determiners *han* and *ki*.

As observed earlier, previous work has suggested that in languages such as English (Klatt 1976) and French (Benguerel 1970), the last stressed vowel of the intonational phrase is significantly lengthened. Jun (1993:38) suggests that Korean too the last syllable of the intonational phrase is lengthened. However, t current data contain instances where not only the last syllable but the penultima syllable in the intonational phrase is lengthened. What was observed in the data that some intonational phrase boundaries are prosodically marked more clearly th

others. There was much variation in the peak f0 value of the H or HL boundar tone. In other words, Some H or HL boundary tones were realized with higher f than others (see Koo 1986 for the variation in the f0 value of H boundary tones i Korean).

Likewise the lengthening of the last syllable of the intonational phrases wa not always comparable. That is, the lengthening of the last syllable of the intonational phrase was variable except that the last syllable was substantially longer than the other syllables in the intonational phrase. The impression was that sometimes the last vowel of the intonational phrase was lengthened more than two times its normal duration and sometimes the lengthening occurred only about one half times the normal duration of the vowel. The data suggest that the lengthening of the penultimate syllable in the intonational phrase occurs when the speaker marks the intonational phrase boundary 'very' clearly. This variation of the final lengthening at the end of an intonational phrase in Standard Korean is in line with Vassiere's (1983:61) observation that "phrase-final lengthening" can be realized in different manners in different languages. Accordingly, I decided not to compare those pairs of expressions which were produced with an intonational phrase boundary either in new or given status. This decision was made to prevent the lengthening influence of the intonational phrase boundary from distorting the results of the analysis.

It was observed earlier that the underlying tonal pattern of the accentual phrase in Standard Korean is LHLH (Jun 1993:42), but that the first H tone and the second L tone often undershoot when the accentual phrase is short — i.e. when the accentual phrase consists of one to three syllables (Jun 1993:43). It was also observed that the final H tone of the accentual phrase is preempted by an intonational phrase boundary tone when the accentual phrase boundary coincides with the intonational phrase boundary. Since the subjects occasionally made an intonational phrase break and marked the phrase of interest with a boundary tone, the comparison of the f0 values of the final H tone of the phrase was not possible. This is because the f0 value of the H boundary tone is significantly higher than that of the final H tone of an accentual phrase and also because there was much variation (as suggested earlier) across the peak f0 values across H or HL boundary tones.

Accordingly, when both phrases (given phrase and new phrase) were produced in one accentual phrase, the f0 values of the first H tones of the accentual phrases were compared (cf. Figure 4a). When the phrases were produced in two accentual phrases, the f0 values of the phrase-final H tones of the first accentual phrases were compared (cf. Figure 4b: the underlying first H tone was not realized in the data because the phrase was not long enough — i.e. 1 to 3 syllables). This latter comparison was possible because the adjectival word (in Study 1) and the determiner (in Study 2) were never marked with a boundary tone in the data. Needless to say, the f0 values were compared only when accentual phrasing was identical between the new entity and its given counterpart. The decision as to whether the phrase was produced in one accentual phrase is based on my perception as a native speaker of Seoul Korean and also on the f0 contour. The most important cue in the f0 contour is the slope of the f0 fall from the first H tone. The slope is much steeper when the phrase is produced in two accentual phrases than when produced in one accentual phrase (cf. Jun 1993).

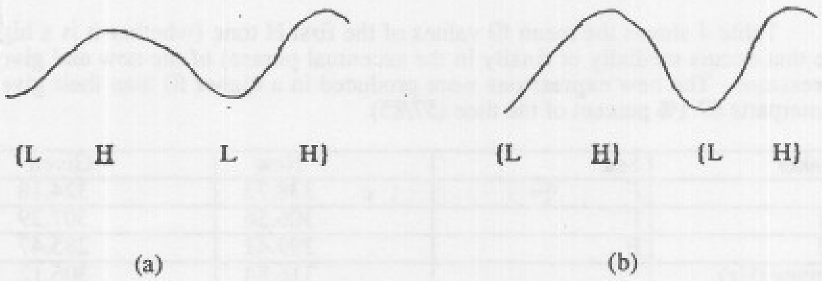



Figure 4: The H tones measured (the phrase-initial H tone, when phrases were produced in one accentual phrase (4a); the phrase-final H tone, when phrases were produced in two accentual phrases (4b))

Average intensity (i.e. average intensity with which the whole referring expression was spoken) was used for intensity measures. Average intensity was calculated using a program written by Mary Beckman. It was found through the examination of the data that the existence of a boundary tone exerted little — if any — influence on the intensity the linguistic entity was produced with. Accordingly, all the tokens were available for the comparison, regardless of whether a boundary tone followed the token or not.

For the calculation of the approximate speech rate of each speaker, 20 sentences were chosen. The chosen sentences were 10 from the data for Study 1 and 10 from the data for Study 2. Identical sentences were selected across speakers. The speech rate was calculated by dividing the duration of the sentence by the number of the syllables in the sentence.

### 3.3 Results of Study 1

The number of the comparable pairs of new and given entities was 84 in duration, 85 in f0 and 104 in amplitude. Table 3 shows the mean duration values of the new and given expressions according to the number of syllables the expressions consist of. The new expression was longer than its given counterpart 75 percent of the time (63/84). As the number of syllables increases, the difference between new and given expressions becomes larger — probably because there is more room for shortening for longer expressions as Fowler & Housum (1987) suggested.

| | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|
| Speaker | New | Given | New | Given | New | Given |
| CK | 660 | 597 | 767 | 717 | 860 | 798 |
| YEJ | 542 | 540 | 703 | 657 | 772 | 717 |
| PYJ | -- | -- | 761 | 723 | 787 | 787 |
| LHS | 517 | 482 | 650 | 638 | 738 | 687 |
| LSH | 513 | 478 | 654 | 630 | 712 | 670 |
| CDS | 633 | 687 | 863 | 802 | 953 | 865 |
| Average (ms) | 551 | 540 | 732 | 694 | 796 | 743 |

Table 3. Duration of new and given expressions with different numbers of syllables

Table 4 shows the mean f0 values of the first H tone (whether it is a hig tone that occurs medially or finally in the accentual phrase) of the new and give expressions. The new expressions were produced in a higher f0 than their give counterparts 67.1% percent of the time (57/85).

| Speaker | Sex | New | Given |
|---|---|---|---|
| CK | F | 338.73 | 324.18 |
| YEJ | F | 306.56 | 307.29 |
| PYJ | F | 299.41 | 285.47 |
| Average (Hz) | | 316.84 | 306.12 |
| LHS | M | 168.36 | 151.05 |
| LSH | M | 138.82 | 136.86 |
| CDS | M | 186.51 | 170.92 |
| Average (Hz) | | 163.24 | 153.02 |

Table 4. Measurements of f0 of new and given expressions

The average amplitudes with which the phrases were spoken are given in Table 5. The new expressions were produced with more average intensity than their given counterparts 64.4% percent of the time (67/104).

| Speaker | New | Given |
|---|---|---|
| CK | 663.45 | 673.27 |
| YEJ | 662.94 | 637.61 |
| PYJ | 537.04 | 482.58 |
| LHS | 582.24 | 559.76 |
| LSH | 705.24 | 666.12 |
| CDS | 725.53 | 693.94 |
| Average (rms unit) | 649.95 | 622.81 |

Table 5. Measurements of average amplitude of new and given expressions

New and given expressions also showed quite different patterns of accentual phrasing. The new expressions were produced in two accentual phrases 46.1% of the time (47/102), while the given expressions were produced likewise 31.7% of the time (33/104). A $\chi^2$ test was performed to examine how significant the difference between the two types of expressions is. The difference was significant at .05 level ($\chi^2=3.981$, df=1, $p<.05$).

A paired t-test was performed to see whether there is a significant difference in the mean duration, f0 and average amplitude between the new expressions and their counterparts. Duration, f0 and average amplitude were all found to be significantly different (duration: ($T=5.74$, df=83, $p<0.0005$), f0: ($T= 3.84$, df=84, $p<0.0005$), average amplitude: ($T=2.62$, df=103, $p=.01$)). A paired t-test also found the difference in accentual phrasing between new phrases and their given counterparts to be significantly different ($T=4.04$, df=103, $p<0.0005$).

Jun's (1993:180) claim that phonological weight (number of syllables in a phrase) is one of the important factors on how speakers accentually phrase expressions was generally supported by the results. While 7-syllable phrases (including the particle) and 6-syllable phrases were produced in two accentual phrases 66.7 (48/72) and 44.1 (30/68) percent of the time respectively, 5 word phrases were produced in two accentual phrases only 2.9 (2/68) percent of the time.

| Speaker | Percentage |
|---|---|
| CK | 32.4% (11/34) |
| YEJ | 55.6% (20/36) |
| PYJ | 47.1% (16/34) |
| LHS | 47.1% (16/34) |
| LSH | 19.4% (7/36) |
| CDS | 29.4% (10/34) |

Table 6. Speakers' frequencies of producing the expressions in two accentual phrases

There was a major difference in the patterns of accentual phrasing across speakers, as Table 6 suggests. Significant cross-speaker variation was observed. While speaker YEJ produced the phrases in two accentual phrases 55.6% of the time, speaker LSH produced the phrases in two accentual phrases only 19.4% of the time.

### 3.4. Results of Study 2

As mentioned earlier, when consonants such as aspirated or glottalized obstruents or a glottal fricative /h/ begins an accentual phrase, the accentual phrase begins in a H tone (Jun 1993:42). This means that while the new phrase (i.e. *han* phrase) begins in a H tone, the given phrase (*kɨ* phrase) starts in a L tone. Together with the fact that different segments make up the two different determiners (*han* and *kɨ*) — most importantly, there is an inherent difference in f0 value between high vowels and low vowels (Lehiste & Peterson 1961) — this suggests that the comparison of the f0 value of the first H tone is not possible. Accordingly only the duration of the noun, average amplitude and accentual phrasing were measured and examined for Study 2. The number of comparable pairs was 61 in duration, 63 in average amplitude and 68 in accentual phrasing.

The duration of the noun showed no difference as a function of informational status. The new 'noun' was longer than the given noun 50.8% of the time (31/61) but the average duration of the new nouns was not longer than that of the given nouns (Table 7). The paired t-test found the difference to be minimal (T=.189, df=60, p=.855).

| Speaker | New | Given |
|---|---|---|
| CK | 358 | 340 |
| YEJ | 304 | 323 |
| PYJ | 342 | 338 |
| LHS | 366 | 359 |
| LSH | 241 | 249 |
| CDS | 343 | 338 |
| Average (ms) | 324 (N=61) | 325 (N=61) |

Table 7. Measurements of duration of new and given nouns

The speakers didn't produce the new noun with greater average intensity either. The new noun was spoken with greater average intensity only 42.9% of the time (27/63). As shown in Table 8, the average intensity with which new nouns were produced was less than their counterparts. The difference was not significant either (T=.533, df=62, p=.600).

| Speaker | New | Given |
|---|---|---|
| CK | 699 | 719 |
| YEJ | 657 | 640 |
| PYJ | 581 | 568 |
| LHS | 564 | 658 |
| LSH | 719 | 735 |
| CDS | 777 | 766 |
| Average (rms unit) | 667 (N=63) | 681 (N=63) |

Table 8. Measurements of average amplitude of new and given nouns

However, a significant difference was observed in accentual phrasing between new and given phrases. While given phrases (*ki* phrases) were produced in two accentual phrases 33.8% of the time (23/68), new phrases (*han* phrases) were produced likewise 72.1% of the time (49/68). Though the phrases were all composed of 4 syllables (including the particle), they were produced in two accentual phrases 52.9% of the time (72/136). A clear difference across speakers in accentual phrasing was again observed (Table 9).

| Speaker | Percentage |
|---|---|
| CK | 80% (16/20) |
| YEJ | 65% (13/20) |
| PYJ | 58% (14/24) |
| LHS | 29% (7/24) |
| LSH | 29% (7/24) |
| CDS | 63% (15/24) |

Table 9. Speakers' frequencies of producing the phrases in two accentual phrases

### 3.5 Discussion

Unlike Fowler & Housum's (1987) study, Study 1 finds a significant mean difference not only in duration but also in f0 value and average amplitude. This result suggests that duration, f0 value and amplitude are all phonetic correlates of informational status in Standard Korean and contribute to how speakers of Standard Korean distinguish linguistic entities of these two different informational statuses. Differences in results between the present research and Fowler & Housum's (1987) study might come from the factor that the latter researchers didn't control sentential positions and intonational patterns of linguistic entities when they compared f0 and average amplitude of new and old expressions. Or this difference might possibly come from cross-linguistic differences. Only future studies on English that compare new and given linguistic entities with necessary controls will be able to give the answer.

Another important thing to note is the role of accentual phrasing in distinguishing new and given information in Standard Korean. In Study 1, the paired t-test found a significant difference in accentual phrasing between the two informational types of expressions. In Study 2, accentual phrasing was the only phonetic cue that distinguished new information from given information.

The results of Study 2 suggest that duration and amplitude do not play a role in distinguishing a new word from a given word when a syntactic cue is given for this purpose. It is very interesting that Korean speakers use not duration or amplitude but accentual phrasing as a complement to a syntactic cue. This result supports Jun's (1993:199) claim that given information in the discourse is more likely to be produced in one accentual phrase.

Phonological weight showed a significant influence on accentual phrasing in Study 1, which supports Jun's (1993) suggestion that the length of a phrase is an important factor affecting whether that phrase is spoken in one accentual phrase or in two accentual phrases. However, Jun's (1993:180) claim that a potential accentual phrase that consists of more than 5 syllables tends to be phrased as two was not supported by the results for Study 1. The 6-syllable sequences were produced in two accentual phrases only 44.1 percent of the time. This suggests that at least in speech close to conversational speech, 6-syllable phrases are produced in one accentual phrase a majority of the time.

However, it doesn't seem to be the case that phonological weight always plays a crucial role. Though the new and given linguistic entities were all 4-syllable phrases in Study 2, they were spoken in two accentual phrases 52.9% of the time (72/136). This frequency is even higher than the frequency with which the 5- and 6- syllable phrases were produced in two accentual phrases in Study 1 (2.9%, 30% respectively). This peculiar behavior of *han* and *kɨ* phrases seems to come from the fact that these two determiners perform a function of cuing informational status of the following noun — or the referent. In some sense, we might say that the semantic weights (see Bolinger 1972, Jun 1993) of these determiners are heavy. The patterns of accentual phrasing of the *han* and *kɨ* phrases in Study 2 suggest that the phonological weight factor can be overridden when accentual phrasing is the only phonetic cue that shows the informational status of phrases.

The speech rate of each speaker showed a relatively high correlation to how often the speaker produces the phrases in two accentual phrases (r=.729), as Table 10 suggests. This correlation was marginally significant (df=4, p=.10). However, the slowest speaker (CDS), a male speaker, was only the fourth following the three female speakers in the order of frequency of producing the phrases in two accentual phrases. Further among the three female speakers the speech rate was inversely proportional to the rate of producing the phrases in two accentual phrases. This result does not allow us to reach any conclusion at the present time.

The results shown in Table 10 make another suggestion. Though our samples are not representative samples of the two gender populations, one possible interpretation of the result is that the female speakers may be more apt to produce the phrases in two accentual phrase than the male speakers. This, in turn, might suggest that female speakers of Standard Korean speak producing more accentual phrases than male speakers of Standard Korean, i.e. with more intonational contours than male speakers. This interpretation agrees to the results of various studies that suggested women's speech involve more intonational contours and rapid pitch shifts (McConnel-Ginnet 1983, Fichtelius, Hohansson & Nordin 1980, Terango 1966) However, a much more extensive sociolinguistic study would be needed to support this interpretation.

| Speaker | Sex | Speech Rate (ms/syl) | Percentage |
|---|---|---|---|
| LSH | M | 117 | 23.3% (14/60) |
| LHS | M | 138 | 39.7% (23/58) |
| YEJ | F | 154 | 58.9% (33/56) |
| PYJ | F | 162.6 | 51.7% (30/58) |
| CK | F | 163.2 | 50.0% (27/54) |
| CDS | M | 175 | 43.1% (25/58) |

Table 10. Speech rate and frequency of producing the phrases in two accentual phrases

## 4. Experiment 2

## Study 3

Brown (1983), Yule (1980, 1981) and Brown & Yule (1983) classify the textually given entities — i.e. entities that were already introduced in the previous discourse and reappear — into current evoked entities and displaced evoked entities. As Chafe (1972:50) observes, a given entity is "foregrounded" at one point in a discourse but later it slips out of the foreground of the discourse participants' consciousness as other entities are introduced and discussed. Thus the "psychological" status of an entity that has just been introduced in the discourse cannot be identical to that of the one that was introduced earlier in the discourse. Brown and Yule's division of given entities into current and displaced evoked entities can be understood as an attempt to distinguish these two different psychological statuses.

Then the question is whether and how the difference between these two types of given entities is reflected in speakers' production. One research issue is whether the two types of given entities are produced acoustically and prosodically in a different manner. Brown (1983) suggested in her study on English that the current entity and the displaced entity are not produced differently in terms of pitch, though they are apt to be produced differently syntactically. The purpose of Experiment 2 is to examine how new, current and displaced entities are produced phonetically by Standard Korean speakers. Since we have already observed how speakers of Standard Korean distinguish new vs. given entities, the focus of the study will be to investigate to what degree the phonetic realizations of current and displaced entities are different. In addition, the role of informational status, phonological weight and speech rate on the speakers' pattern of accentual phrasing is reexamined here. Brown (1983:75) defines a current evoked entity as "an item which has just been introduced into the discourse and which is currently the entity to which new information is being related" and a displaced evoked entity as an item "which has been introduced into the discourse at a point previous to the currently evoked item". I will adopt her definitions in this study.

### 4.1 Materials

As the material for Experiment 2, I decided to use a narrative rather than a dialogue. The main reason for this decision was that a narrative is a genre of discourse where the same referring expressions can appear in the text very naturally. Also it was observed that the investigator is able to control the speaker's use of f0, amplitude and duration for other purposes noted earlier (e.g. for focal or contrastive purposes, or for holding the floor or for directing the flow of information) better in narrative speech. This is primarily due to the fact that narrative speech is unidirectional speech (i.e. from the speaker to the listener) that involves less interaction with the listener, while dialogue speech is bi-directional involving much more interaction. Another reason for the choice of a narrative was to elicit more formal, slow speech data from the speakers in this experiment. It was expected that a more careful, slow style of speech can be elicited by having the subject read the narrative. The narrative constructed for this experiment and its English translation are given in Appendix 3, where four triplets of a new entity, a current evoked entity and a displaced evoked entity appear.

In the narrative, as soon as the sentence which is initiated by a new 'phrase' is finished, the current evoked phrase starts the next sentence. After three sentences of relatively comparable lengths the displaced evoked phrase appears in the discourse. Efforts were made to minimize the effects of discourse structure, which is known to affect especially the pitch of speakers' production at the beginning of a discourse segment (Grosz & Hirschberg 1992, Hirschberg & Pierrehumbert 1986, Yule 1980). The narrative was constructed so that new expressions do not appear at the beginning of a discourse segment. Efforts were also made that the expressions do not receive any special emphasis (e.g. focal or contrastive emphasis) in the discourse. The constructed expressions were two 4-syllable and two 5-syllable noun phrases. Accordingly, the phonological weights of the phrases were 5 and 6 syllables including the particle attached to the noun phrase. The constructed linguistic entities are listed in Table 11.

| No. of syllables | Entity | Gloss |
|---|---|---|
| 4 | yepɨn yəin | 'pretty woman' |
| 4 | kəmɨn iɹi | 'black wolf' |
| 5 | yəpmaɨl chəŋnyən | 'neighbor-village young-man' |
| 5 | nuɹən koyaɲi | 'yellow cat' |

Table 11. The list of the expressions used for study 3

## 4.2 Methods

The same 6 speakers participated in Experiment 2. The procedure for the recording was as follows. I first outlined the content of the narrative for the subject. Then there was one practice reading session. The subject was asked to read the narrative trying to understand the content of the narrative. The subjects were asked to read the narrative naturally and vividly as if they actually read the narrative to a listener. They were encouraged to think of the author as a listener. The next two readings were recorded and those sentences of interest were digitized. Measurements were made again using the Entropics Waves program (Version 5.0).

The identical method used in Study 1 and 2 was used in Study 3 for the comparison of duration of the expressions. That is, the duration of the noun phrase (adjectival + noun) excluding that of the particle was measured. Also duration was compared only among those triplets of entities that were produced without a boundary tone at the end in any informational status. The comparison of f0 was also conducted in the identical manner. The f0 values of the first H tone during the production of the expression were compared.

16 identical sentences — 8 from the first reading and 8 from the second reading — were selected for the calculation of speech rate of each speaker in this experiment. Speech rate was calculated in the same method as in Experiment 1.

## 4.3 Results

34 and 35 triplets were available for the comparison of duration and f0 values, respectively. 47 triplets were used for the comparison of average amplitude and accentual phrasing. Table 12, 13 and 14 show the mean values of f0, average amplitude and duration of expressions of three different informational statuses as produced by the subjects. Table 15 gives the frequency in which the expressions of different informational statuses were produced in two accentual phrases.

If new, current and displaced entities have different informational statuses, it is expected that new expressions be produced with higher f0, greater average amplitude and longer duration than displaced and current expressions, and that current expreessions be spoken with lower f0, less average amplitude and shorter duration than displaced expressions. The results were interpreted in this perspective.

| Speaker | Sex | New | Current | Displaced |
|---|---|---|---|---|
| CK | F | 283 | 245 | 262 |
| YEJ | F | 283 | 248 | 271 |
| PYJ | F | 232 | 220 | 243 |
| Average (Hz) | | 267 | 236 | 259 |
| LHS | M | 148 | 134 | 139 |
| LSH | M | 139 | 133 | 139 |
| CDS | M | 168 | 153 | 164 |
| Average (Hz) | | 152 | 139 | 147 |

Table 12. Measurements of f0 of the expressions in three different informational statuses

| Speaker | New | Current | Displaced |
|---|---|---|---|
| CK | 677 | 588 | 624 |
| YEJ | 699 | 596 | 658 |
| PYJ | 676 | 640 | 714 |
| LHS | 660 | 484 | 508 |
| LSH | 718 | 615 | 719 |
| CDS | 733 | 624 | 721 |
| Average (rms unit) | 695 | 593 | 656 |

Table 13. Average Amplitude of the expressions in three different informational statuses

| Speaker | New | Current | Displaced |
|---|---|---|---|
| CK | 711 | 672 | 648 |
| YEJ | 632 | 583 | 613 |
| PYJ | 743 | 723 | 683 |
| LHS | 646 | 590 | 606 |
| LSH | 577 | 543 | 580 |
| CDS | 735 | 620 | 623 |
| Average (ms) | 672 | 623 | 628 |

Table 14. Measurements of duration in the expressions of three different informational statuses

As expected, the speakers clearly distinguished new entities from current entities. Paired t-tests showed that the average values of three of the four phonetic cues — i.e. duration, f0 value and average amplitude — were significantly different between new and current entities — f0 (T=5.589, df=34, p<.0005), duration (T=6.49, df=33, p<.0005), average amplitude (T=3.638, df=46, p=.001). The expressions were produced in one accentual phrase more often in current status than in new status but not significantly (T=.618, df=46, p=.533)

| Speaker | New | Current | Displaced |
|---|---|---|---|
| CK | 3/8 | 3/8 | 2/8 |
| YEJ | 6/8 | 6/8 | 6/8 |
| PYJ | 4/7 | 4/7 | 4/7 |
| LHS | 4/8 | 2/8 | 4/8 |
| LSH | 1/8 | 1/8 | 3/8 |
| CDS | 2/8 | 2/8 | 2/8 |
| Average | 20/47: 43% | 18/47: 38% | 21/47: 45% |

Table 15. Frequency of producing the expressions in two accentual phrases

The phonetic distinctions between current and displaced entities were relatively less clear than those between new and current entities. F0 and average amplitude were found by paired t-tests to be significantly different. However, the average duration of displaced expressions was not significantly longer than that of current expressions (T=1.21, df=31, p=.237). The difference in accentual phrasing was also not significant (T=1.00, df=46, p=.323).

The distinctions between new and displaced entities were also not as clear as those between new and current entities. Paired t-tests found duration and f0 as significantly different. Duration was different at .0005 level (T=4.37, df=41) and f0 was different at .05 level (T=2.10, df=42). The difference in average amplitude between new and displaced entities was marginally significant (T=1.73, df=46, p=.091). Against the expectation, the displaced entities were produced in two accentual phrases one more time than the new entities (21/47 vs. 20/47). However, the difference was minimal. Table 16 shows the frequency in which these three different types of entities were produced in accordance with the aforementioned prediction.

| | Duration | F0 | Amplitude |
|---|---|---|---|
| new vs. current | 91.1% (31/34) | 88.6% (31/35) | 74.5% (35/47) |
| displaced vs. current | 53.1% (17/32) | 71.4% (25/35) | 72.3% (34/47) |
| new vs. displaced | 78.6% (33/42) | 60.5% (26/43) | 51.1% (24/47) |

Table 16. Frequency in which speakers produce the expressions of three different informational statuses as expected

Though accentual phrasing did not reflect different informational statuses very reliably, the speakers' pattern of accentual phrasing varied significantly according to the phonological weights of the expressions. While the 6-syllable phrases were spoken in two accentual phrases 63.9% of the time, the 5-syllable phrases were produced likewise only 18.9% of the time (Table 16). The $\chi^2$ test found the difference significant ($\chi^2$=29.38, df=1, p<.001). The speakers produced the expressions in two accentual phrases more frequently than in Study 1. This is clearly shown by Table 17, which compares the accentual phrasing of the 5- and 6-syllable phrases of Study 1 with that of Study 3.

| | Dialogue | Narrative |
|---|---|---|
| 5-syllable phrase | 2.9% (2/68) | 18.9% (13/69) |
| 6-syllable phrase | 44.1% (30/68) | 63.9% (46/72) |

Table 17. Accentual phrasing of phrases with two different phonological weights

## 4.4 Discussion

Though the difference in duration between the displaced entities and current entities missed significance (p=.237), the result that the displaced entities were produced with a significantly higher average f0 and a significantly greater average intensity than the current entities seems to support and justify Brown and Yule's division of given information into current vs. displaced evoked entities. The results show that speakers of Standard Korean distinguish phonetically not only new vs. given entities but current vs. displaced entities. This, in turn, supports the claim that these three different types of entities (new, current and displaced) have independent psychological status.

The results for Experiment 2 are significantly different from those for Brown's (1983) study, which did not find any significant f0 difference between current and displaced entities. Rather the results of the present study support Chafe's (1972:51) observation that a foregrounded entity is produced with lower pitch and less amplitude than the entity not foregrounded. This difference in results between Brown's study and the present study again could come from some different sources. First, like Fowler & Housum's (1987) study, Brown didn't control the sentential position and intonational phrasing of linguistic entities under study, which can be a serious confounding factor. Secondly, this difference might come from different intonational structures of English and Standard Korean. As observed earlier, the fact that English has L*, L*+H and H+L* pitch accents in addition to H*, L+H* and H*+L accents could have been a factor in Brown's finding. Another possible factor is that in her study, new and displaced entities had significantly different syntactic realizations. While the displaced entities appeared exclusively as either *the + adjectival + noun* or *the + noun*, the current entities occurred in these syntactic forms only 24% of the time. It seems possible that Brown's finding that the displaced entities were not produced with pitch prominence might come from the fact that the syntactic cue is already given to distinguish the two different types of given entities.

In this study too, the phonological weight of the phrase significantly influenced how speakers accentually phrase the expressions (Table 17). Jun's claim that speakers generally produce the phrases longer than 5 syllables in two accentual phrases was better supported by the results for Experiment 2 than those for Experiment 1.

As expected, the speakers' speech rate was generally slower than in Experiment 1 (Table 18). Four of the speakers read the narrative more slowly than the dialogues. But the speakers did not show a major difference in speech rate in these two types of data. The mean speech rate of the 6 speakers was 151.6 ms/syl in the dialogue data and 157.1 ms/syl in the narrative data. A paired t-test found the difference between the speech rates of each speaker in the two experiments not to be significant (T=.731, df=5, p=.497). Considering a significant difference in accentual phrasing between the two studies shown in Table 17, this result suggests that not only speech rate but the genre (or type) of discourse also affects how the speaker produces his/her speech intonationally. This interpretation seems plausible because it is often observed that a narrator recites the story with slightly exaggerated intonational contours.

Unlike in Experiment 1, speech rate showed no correlation with the accentual phrasing pattern (r= –.021). I interpret this result as implying that the speech rate factor in accentual phrasing applies intra-individually rather than inter-individually. In other words, the same speaker will produce more accentual phrases if s/he speaks faster, but it does not seem to be the case that a slow speaker necessarily produces more accentual phrases than a fast speaker. It is apparent that

| Speaker | Sex | Speech Rate (ms/syl) | Frequency |
|---|---|---|---|
| LSH | M | 138.6 | 20.8% (5/24) |
| YEJ | F | 141.8 | 75.0% (18/24) |
| CDS | M | 154.1 | 25.0% (6/24) |
| LHS | M | 162.9 | 41.7% (10/24) |
| PYJ | F | 169.8 | 57.1% (12/21) |
| CK | F | 175.6 | 33.3% (8/24) |

Table 18. Each speaker' speech rate and frequency of producing the expressions in two accentual phrases

some speakers have an idiolectal habit of producing sentences making more accentual phrases than others. In this experiment too, the female speakers produced more accentual phrases than the male speakers.

In this experiment, accentual phrasing was not a good indicator of informational status of the expressions. It didn't show much variation across the expressions of different informational statuses. Still the notable point is that new expressions were produced more frequently in two accentual phrases than the current expressions and that the displaced expressions were more often produced in two accentual phrases than the current expressions. That is, the results show the expected trend that the entities 'foregrounded' (current evoked entities) are more often produced in one accentual phrase than the entities 'in the background' (new and displaced entities).

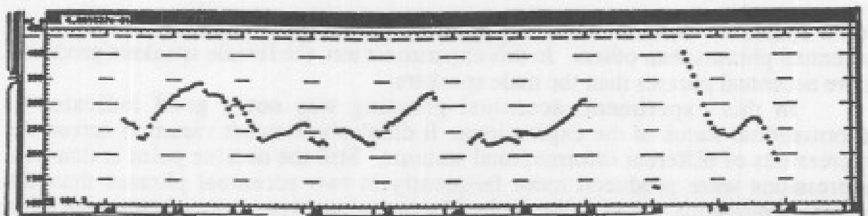
However, it cannot be claimed that the results for Study 3 fully support the findings in Study 1 and Study 2. Nevertheless, since Study 1 and 2 showed a strong indication that accentual phrasing is part of the cues that distinguish different informational statuses, I interpret the combined results for the three studies as suggesting that in Standard Korean, accentual phrasing is one of the cues that distinguish new and given information. It seems to be sensible to interpret the results for the three studies as follows. Intonational phrasing sometimes plays a crucial role as a prosodic cue to distinguish linguistic entities of different informational statuses. But it is rather doubtful whether it is always an important cue like the other phonetic cues. It seems to be that sometimes intonational phrasing plays only a complementary role to the other phonetic cues or a syntactic cue.

### 5. General Discussion — 'Intonational Attenuation'

Past studies have identified f0, duration and intensity as possible phonetic correlates of prominence. However, Vassiere (1983) and Beckman (1986) suggest cross-linguistic differences in these phonetic correlates of prominence. What they suggest is that different languages could use different strategies for prominence purposes. Some languages might use f0 exclusively; some other languages might use both f0 and duration; and still other languages might use all these features (f0, duration and amplitude) for prominence purposes. Vassiere (1983:65) states that while f0, duration and intensity show a close correlation in an accented syllable in English, this isn't necessarily the case in French. Behne (1989) also suggests that English and French do not share the identical phonetic correlates of focal emphasis. These cross-linguistic differences will be closely related to the nature of the prosodic system each language has — i.e. whether that language is a stress language, whether that language has a pitch accent system (and what type of pitch accents), and which type of prosodic and intonational structure the language has. The results for the two experiments of the present research suggest that all of the features mentioned above — f0, duration and amplitude — are used for prominence purposes in Standard Korean.

But the way Standard Korean takes advantage of f0 movement for a prominence purpose seems to be different from pitch accent languages like English or Dutch. While the latter languages mainly use an f0 movement on the accented syllable(s) of the linguistic entity for prominence, Standard Korean uses the f0 movement over the whole linguistic entity for a prominence purpose. This use of f0 movement for this purpose is directly connected to how a linguistic entity is intonationally phrased — accentually phrased, to be specific.

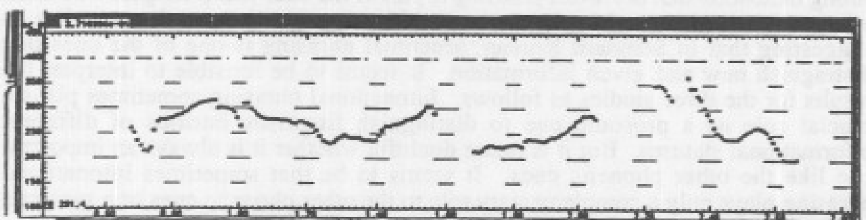
(a)



(Hz) {ka pyɔ -un} {tal li ki- ka} {kɔn kaŋ-e} {$c^h$we ko- ya}
light-Adj P running-Subj health-to best-Decl

L H L H L H L H H L HL%

(b)



(Hz) {ka pyɔ -un tal li ki- ka} {kɔn kaŋ-e} {$c^h$we ko- ya}
light-Adj P running-Subj health-to best-Decl

L H L H L H H L HL%

Figure 5: F0 contours of the sentence *kapyəu-n talikika kənkaŋe* $c^h$*wekoya* 'Light running is the best to health' when produced in two accentual phrases (5a) and in one accentual phrase (5b)

When an expression is produced in two accentual phrases in Standard Korean, it is perceptually more salient than when it is produced in one accentual phrase. It is because the production of the expression is realized with a more rapid and variable pitch movement. This is illustrated in Figure 5, where 5a and 5b, respectively, shows the f0 contour when the expression is produced in two accentual phrases and when it is produced in one accentual phrase.

The exploitation of accentual phrasing for a prominence purpose does not seem to be unique to Standard Korean. Vassiere (1983:60) suggests a similar phenomenon in such languages as French and Shanghai, a Chinese dialect. It is predicted that some other languages with similar intonational structures to that of Standard Korean use this strategy.

Standard Korean speakers seem to use higher f0, longer duration, greater intensity along with a more distinctive pitch movement for the emphasis of new information. But when the speaker assumes the information to be 'given' and readily available to the listener, the speaker seems to attenuate the expression that refers to the entity acoustically and prosodically. I suggest that producing the expression in one accentual phrase is part of 'phonetic attenuation' Chafe (1972) discusses, which I will refer to as 'intonational attenuation'. This rather unique form of attenuation is made by a simplification of f0 movement.

## 6. Conclusion

Chafe (1972) observes that given information is apt to be attenuated both phonetically and syntactically. The present research fully supports his claim on phonetic attenuation in that each of the four prosodic features (phonological features, in Chafe's terms) was found to be attenuated in the predicted direction.

The findings of the present research can be summarized as follows. This research found that f0 value, duration, amplitude are important correlates of informational status in Standard Korean. It was also found that intonational phrasing can play sometimes a crucial and sometimes complementary role to the above-mentioned prosodic features. This result is very different from Fowler & Housum's (1987) study, which identified duration as a sole correlate of informational status. A couple of dubious points in the research method of their study have been also pointed out.

This research also generally supports the distinction of the three different informational statuses — new, current and displaced — by providing evidence that linguistic entities of these three different informational statuses are produced prosodically differently by Standard Korean speakers. However, the present research was very different from Brown's (1983) study in that this study found a clear difference in f0 value between current entities and displaced entities.

The results for this research generally support Jun's (1993) claim that informational status interacts with the phonological weight of expressions in affecting how Korean speakers accentually phrase them. The present research also showed that there is a clear idiolectal variation in accentual phrasing. Some speakers have the tendency to produce utterances with more accentual phrasings than others. It was also suggested that the effect of speech rate on accentual phrasing should be understood intra-individually rather than inter-individually. Finally I made a cautious suggestion that speakers of Standard Korean may show different patterns of accentual phrasing across different genres of discourse and that there might be cross-gender differences in the frequency of accentual phrasing. These are two areas where more extensive research could produce interesting findings.

## 7. References

Beckman, M. 1984. *Stress and nonstress accent.* Netherlands Phonetics Archives. 7. Dordrecht. Foris.

Behne, D. M. 1989. *Acoustic effects of focus and sentence position on stress in English and French.* Ph.D. dissertation. Department of Linguistics. University of Wisconsin at Madison.

Benguerel, A.P. 1970. *Some physiological aspects of stress in French.* Ph.D. dissertation. University of Michigan.

Bolinger, D. 1972. Accent is predictable (if you're a mind reader). *Language*. 48. 633-644.

Brown, G. 1983. Prosodic structure and the given/new distinction. In D.R. Ladd and A. Cutler (eds.), *Prosody: models and measurements*. 67-78. Springer Verlag. Berlin.

Brown, G and G. Yule. 1983. *Discourse analysis*. Cambridge University Press. New York.

Chafe, W. 1972. Discourse structure and human knowledge. In J.B. Carroll and R.O. Freedle (eds.), *Language comprehension and the acquisition of knowledge*. Washington. Wiley.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (eds.), *Subject and topic*. 25-55. New York. Academic Press.

Cohen, A. and J. 't Hart. 1967. On the anatomy of intonation. *Lingua*. 19. 177-192.

de Jong, K. 1989. Initial tones and prominence in Seoul Korean. *Papers from the linguistics laboratory*. Department of Linguistics. Ohio State University. 43. 1-14.

Delattre, P. 1966. *Comparing the phonetic features of English, German, Spanish, and French*. Heidelberg. Julius Verlag.

di Cristo, A. 1978. De la microprosodie a l'intonosyntaxe. These de Doctorat d'Etat. Universite de Provence.

Fichtelius, A, I. Johansson and K. Nordin. 1980. Three investigations of sex-associated speech variation in day school. In C. Kramarae. (eds.), *The voices and words of women and men*. 219-26. Oxford. Pergamon Press.

Fowler, C.A. and J. Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*. 26. 489-504.

Fujisaki, H., H. Hirose, and K. Ohta. 1979. Acoustic features of the fundamental frequency contours of declarative sentences in Japanese. *Annual Bulletin* 13. 163-173. Research Institute of Logopedics and Phoniatrics. University of Tokyo.

Grosz, B. and J. Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English: part 2. *Journal of Linguistics*. 3. 109-244.

Hirschberg, J and J. Pierrehumbert. 1986. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting*. 163-171. Stanford University. Association for Computational Linguistics.

Hirst, D.J. & A. di Cristo. 1984. French intonation: a parametric approach. *Die Neueren Sprachen*. 83:5. 554-569.

Horne, M. 1991. Accentual phrasing in 'new' vs. 'given' subjects in English. *Working Papers* 36. Department of Linguistics. Lund University. Lund.

Jun, S.A. 1993. *The phonetics and phonology of Korean prosody*. Ph.D. dissertation. Department of Linguistics. Ohio State University.

Klatt, D.H. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *JASA*. 59. 1208-1221.

Ko, D.H. 1988. Declarative intonation in Korean: an acoustical study of f0 declination. Seoul. Hanshin Publishing Co.

Koo, H.S. 1986. *An experimental acoustic study of the phonetics of intonation in Standard Korean*. Ph.D. dissertation. University of Texas at Austin. Hanshin Publishing Co. Seoul.

Lee, S.H. 1989. Intonational domains of the Seoul Dialect of Korean. Paper presented at the 117th meeting of the Acoustical Society of America. Syracuse, N.Y.

Lehiste, I. and G.E. Patterson. 1961. Some basic consideration in the analysis of intonation. *JASA*. 33. 367-84.

Lehman, C. 1977. A re-analysis of givenness: stress in discourse. In *Papers from the 13th Annual Meeting*. 316-330. Chicago Linguistic Society.

Maeda, S. 1976. *A characterization of American English intonation*. Ph.D. dissertation. M.I.T.

McConnel-Ginnet, S. 1983. Intonation in a man's world. In T. Barrie, K. Cheris and H. Nancy (eds.), *Language, gender and society*. Rowley, MA. Newbury House.

Nespor, M and I. Vogel. 1986. *Prosodic phonology*. Dordrecht. Foris.

Ohala, J. 1977. Physiology of stress. In L.M. Hyman (eds.), *ILL studies in stress and accent*. Southern California Occasional Papers in Linguistics. 4. 145-168.

Oller, D.K. 1973. The duration of speech segments: the effect of position in utterance and word length. *JASA*. 54. 1235-73.

Pierrehumbert, J. 1979. The perception of fundamental frequency declination. *JASA*. 66. 363-369.

Pierrehumbert, J. & M. Beckman 1988. *Japanese tone structure*. Cambridge. MIT Press.

Prince, E.F. 1981. Toward a taxonomy of given - new information. In P. Cole (eds.), *Radical pragmatics*. New York. Academic Press.

Selkirk, E. 1984. *Phonology and syntax: the relation between sound and structure*, Cambridge. MIT press.

Terken, J. 1984. The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*. 27. 269-289.

Terango, L. 1966. Pitch and duration characteristics of theoral reading of males on a masculinity-feminity dimension. *Journal of Speech and Hearing Research*. 9. 590-95.

Vassiere, J. 1983. Language-independent prosodic features. In D.R. Ladd and A. Cutler (eds.), *Prosody: models and measurements*. 53-66. Springer Verlag. Berlin.

Yule, G. 1980. Speakers' topics and major paratones. *Lingua*. 52. 33-47.

Yule, G. 1981. New, current and displaced entity reference. *Lingua*. 55. 41-52.

* I thank Mary Beckman and Keith Johnson for their helpful comments. I also thank Sun-Ah Jun for her suggestions during the experiments and the analysis of the data.

## Appendix 1

**5-syllable phrases**

**5.1**

A. yocim na-n saɹam-til-etæhæ silmaŋ-il manhi hæ.<br>
nowadays I-Subj person-Pl-about disappointment-Obj much do.<br>
"Nowadays I am very disappointed about people."

B. nəlp-in maim-il saɹam-til-i kacy-ess-tamyən coh-il$^{h}$ente.<br>
broad mind-Obj person-Pl-Subj have-Pst-if good-would be<br>
"It would be good if people had a broad mind."

A. kiɹəke mal-ya.<br>
right words-Decl<br>
"That's right."

B. nəlp-in maim-i cəŋmallo cuŋyoha-n kəs kat$^{h}$-æ.<br>
broad-Adj P mind-Subj really important seem to-Decl<br>
"A broad mind seems to be really important."

**5.2**

A. nə-n acikto æki-kat$^{h}$-kuna.<br>
you-Top still baby-like-Excl<br>
"You are still like a baby."

B. kiɹæ mac-a. yep'i-n inhyaŋ-i na-n sesaŋ-esə ceil coh-a.<br>
yes right-Decl pretty-Adj P doll-Subj I-Top world-Loc most like-Decl<br>
"Yes, it's right. I like a pretty doll most in the world."

A. cəŋmal kiɹ-æ?
really so-Int
"Is it really so?"

B. yep'i-n inhyaŋ-i iss-imyən na-n əncena hæŋpokh-æ.
pretty-Adj P doll-Subj present-when I-Top always happy-Decl
"When I have a pretty girl, I am always happy."
(Excl: exclamatory)

**5.3**

A. ol kail-e-nin uɹi hamkke muənka-lil hæ poca.
this fall-Loc-Top we together something-Obj do try-Hort
"Let's do something together in this fall."

B. kiɹæ. kou-n noɹæ-lil yep'i-ke hana mantil-ə po-ca.
yes pretty-Adj P song-Obj pretty-Adv P one make-Adv P try-Hort
"Yes. Let us make a pretty song *prettily*"

A. coh-in sæŋkak-i-ya.
good-Adj P idea-Cop-Decl
"That's a good idea."

B. kou-n noɹæ-ka iss-imyən uɹi maim-to kop-ke twel-kəya.
pretty-Adj P song-Subj present-if our heart-also pretty-Adv P become-Decl
"If we have a pretty song, our heart will also become pretty."
(Hort: hortatory)

**6-syllable phrases**

**6.1**

A. nə toŋmul-il coh-a ha-ni?
you animal-Obj like-Adv P like-Int
"Do you like animals?"

B. pukmisan yəu-ka na-n coh-a.
North-American fox-Subj I-Top like-Int
"I like North-American foxes."

A. na-n yəu-ka silh-inte.
I-Top fox-Subj dislike-Decl
"I do not like foxes."

B. pukmi-san yəu-lil han pən po-a. cəŋmal məsiss-ə.
North American fox-Obj once look at-Int really charming-Decl
"Take a look at North American foxes once. They are really charming."

**6.2**

A. nə sæŋsən-il coha-ntamyə?
you fish-Obj like-I heard
"I heard that you like fish, is that true?"

B. wanto-san iŋə-il na-n cohah-æ.
Wanto-produced carp-Obj I-Top like-Decl
"I like Wanto-produced carps."

A. com s'i-ci anh-a?
slightly spicy-Adv P not-Int
"Isn't it slightly spicy?"

B. anya. <u>wanto-san iŋa</u>-lil mək-ɨmyən na-n acu kənkaŋhæ ci-
No Wanto-produced carp-Obj eat-when I-Top very healthy become-

nɨn kipun-iya.
Adj P feeling-Decl

"No. When I eat Wanto-produced carps, I feel as if I were becoming healthy."

6.3

A. ipən kail-e-nɨn muənka-lɨl hæ po-ko sipʰ-inte
this fall-Loc-Top something-Obj do try-Adv P want to-Decl
"I want to do something this fall."

B. <u>atamha-n pyəlcaŋ</u>-il ci-ə po-ci kiɹæ?
little-Adj P villa-Obj build-Adv P why don't you
"Why don't you build a little villa?"

A. kikə kwenchanh-ɨn sæŋkak-inte.
that good-Adj P idea-Int
"That is a good idea."

B. <u>atamha-n pyəlcaŋ</u>-i iss-ɨmyən ne maim-to pʰukinha-Ikəya.
little-Adj P villa-Subj present-if your heart-also warm-Decl
"If you have a little villa, your heart will also be warm."

**7-syllable phrases**

7.1

A. nals'i-ka we iɹəh-ke təp-ci?
weather-Subj why like this-Adv P hot-Int
" Why is the weather this hot?"

B. kiɹə-ke mal-ya. <u>siwənha-n alim-mul</u>-i iss-ɨmyən coh-kess-ta.
so-Adv P word-Decl cool-Adj P ice-water-Subj present-if good-will-Decl
"Very right. It would be good if there were ice-water."

A. we. næŋsuyok-il haɹyə-ko?
why cold bath-Obj do-Int
"Why? Do you want to have a cold bath?"

B. aniya. <u>siwənha-n alim-mul</u>-il manhi com masi-ɹyə-ko.
no cool-Adj P ice-water-Obj much drink-Vol-Decl
"No. I want to drink much ice water."
(Vol: volition)

7.2

A. na yocim untoŋ-i pʰilyoha-n kəss kat-æ.
I nowadays exercise-Subj necessary-Adj P seem to-Decl
"I seem to need exercise these days."

B. kapyau-n taliki-ka kənkaŋ-e cʰweko-ya.
light-Adj P running-Subj health-to best-Decl
"It is best to do a light running."

A. kike kiɹəh-ke coh-a?
it so-Adv P good-Int
"Is it that good?"

B. kapyau-n taliki-lil nalmata h-æ po-a. kənkaŋ-e kiman-iya.
light-Adj P running-Obj everyday do-Adv P try-Imp health-to very good-Decl
"Try to do a light running everyday. It is very good for your health."

7.3

A. kwail-iɹako ta masiss-in kess-in ani-ci?
fruit-Subj every tasty-Adj P not-Int
"Not every fruit is tasty. Isn't that right?"

B. ki-ɹæ. səlik-in panana-ka na-nin silh-ə.
right-Decl unripe-Adj P banana-Subj I-Top hate-Decl
"Right. I dislike unripe banana."

A. tikpyəl-han iyu-ka iss-ni?
special-Adj P reason-Subj exist-Int
"Is there a special reason for that?"

B. səlik-in panana-il mək-imyən na-n cal cʰeh-æ.
unripe-Adj P banana-Obj eat-if I-Subj easily have a stomachache-Decl
"When I eat much unripe banana, I easily have a stomachache."

## Appendix 2

Han vs. Ki (6 pairs)

1:han

B. na yocim noɹæ hana-l mantil-ko iss-ə.
I nowadays song one-Obj make-Adv P Aux-Decl
"I am making a song nowadays."

A. ət'ə-n noɹæ-nte?
what kind of song-Int
"What kind of song is it?"

B. han inə-ka waŋca-eke saɹaŋ-il pʰyohyənha-nin noɹæ-ya.
a mermaid-Subj prince-Dat love-Obj express-Adj P song-Decl
"A mermaid expresses her love for a prince in that song."

1:ki

B. na yocim han inə-etæhan noɹæ-l mantil-ko iss-e.
I nowadays a mermaid-about song-Obj make-Adv P Aux-Decl
"I am making a song about a mermaid nowadays."

A. ət'ə-n næyoŋ-inte?
what content-int
"What is the content of the song like?"

B. kɨ inə-ka waŋca-eke saɹaŋ-ɨl pʰyohyənha-nɨn noɹæ-ya.
the mermaid-Subj prince-Dat love-Obj express-Adj P song-Decl
"The mermaid expresses her love for a prince in that song."

2:han

B. næ tu pən-c'æ hɨikok-i kəcin wansəŋtwe-ss-ə
my second comedy almost is finished-Pst-Decl
"I almost finished my second comedy."

A. ət'ə-n culkəɹi-ci?
what content-Int
"What is its story like?"

B. han mæŋin-i nun-ɨl t'i-ke twe-nɨn iyaki-ya.
a blind man-Subj eye-Obj open-Adv P come to-Adj P story-Decl
"A blind man opens his eyes in the story."

2:kɨ

B. næ tu pən-c'æ hɨikok-in han mæŋin-etæhan iyaki-ya.
my second comedy-Top a blind man-about story-Decl
"My second comedy is a story about a blind man."

A. ət'əhke yæki-ka cinhæŋtwe-ci?
how story-Subj proceed-Int
"How does the story go?"

B. kɨ mæŋin-i nun-ɨl t'i-ke twe-nɨn iyaki-ya.
the blind man-Subj eye-Obj open-Adv P come to-Adj P story-Decl
"The blind man opens his eyes in the story."

3:han

B. cinu-ka sosəl-ɨl hana s'ə-ss-te.
Cinu-Subj novel-Obj one write-Pst-I heard
" Cinu wrote a novel, I heard"

A. ət'ə-n næyoŋ-ici?
what content-int
"What is the content of the novel like?"

B. han yəŋuŋ-i nacuŋe cewaŋ-i twe-nɨn iyaki-ɹæ.
a hero-Subj later king-Comp become-Adj P story-I heard
"A hero becomes a king in the novel, I heard."

3:kɨ

B. cinu-ka han yəŋuŋ-etæhan sosəl-ɨl s'ə-ss-tæ.
Cinu-ka a hero-about novel-Obj write-Pst-I heard
"Cinu wrote a novel about a hero, I heard."

A. ət'ə-n næyoŋ-ici?
what content-int
"What is the content of the novel like?"

B. kɨ yəŋuŋ-i nacuŋe cewaŋ-i twe-nɨn iyaki-ɹæ.
the hero-Subj later king-Comp become-Adj P story-I heard
"The hero becomes a king in the novel, I heard."

4:han

B. minsu-ka ipən-e yəŋhwa-lɨl hana mantɨl-ess-tæ.
Minsu-Subj this time movie-Obj one make-Pst-I heard
"Minsu made a movie this time, I heard."

A. ət'ə-n næyoŋ-i-nci a-ni?
what content-Cop-Adv P know-Int
"Do you know what is the story of that movie?"

B. han kunin-e pikɨkcək sæŋæ-lɨl kɨɹi-n cakp$^h$um-iɹæ.
a soldier-Pos tragic life-Obj describe-Adj P work-I heard
"It is a work which describes the tragic life of a soldier, I heard."

4:kɨ

B. minsu-ka ipən-e han kunin-etæhan yəŋhwa-lɨl mantɨl-ess-tæ.
Minsu-Subj this time a soldier-about movie-Obj make-Pst-I heard
"Minsu made a movie about a soldier this time, I heard."

A. ət'ə-n næyoŋ-i-nci a-ni?
what content-Cop-Adv P know-Int
"Do you know what is the story of that movie?"

B. kɨ kunin-e pikɨkcək sæŋæ-lɨl kɨɹi-n cakp$^h$um-iɹæ.
the soldier-Pos tragic life-Obj describe-Adj P work-I heard
"It is a work which describes the tragic life of the soldier, I heard."

5:han

B. na ipən-e uhwa-lɨl hana s'i-ess-ə.
I this time fable-Obj one write-Pst-Decl
"I wrote a fable this time."

A. ət'ə-n uhwa-inte?
what kind of fable-Int
"What kind of fable is it?"

B. han yəu-ka pyəɹak-puca-ka twe-nɨn iyaki-ya.
a fox-Subj upstart-Subj become-Adj P story-Decl
"A fox becomes an upstart in that fable."

5:kɨ

B. na ipən-e han yəu-etæhan uhwa-lɨl s'ə-ss-ə.
I this time a fox-about fable-Obj write-Pst-Decl
"I wrote a fable about a fox this time."

A. ət'ə-n yæki-nte
what kind of story-Int
"How does the story go?"

B. kɨ yəu-ka pyəɹak-puca-ka twe-nɨn iyaki-ya.
the fox-Subj upstart-Subj become-Adj P story-Decl
"The fox becomes an upstart in that fable."

6:han

B. na ipən-e kakɨk-ɨl hana mantɨl-ess-e.
I this time opera-Obj one make-Pst-Decl

"I made an opera this time."

A. ət'ə-n næyoŋ-inte?
what content-int
"What is the story of the opera like?"

B. han manyə-ka minam $c^{h}$əŋnyən-il yuhokha-nin yæki-ya.
a witch-Subj handsome young man-Obj seduce-Adj P story-Decl
"A witch seduces a handsome young man in the story."

6:ki

B. na han manyə-etæhan kakik-il mantil-əss-ə.
I a witch-about opera-Obj make-Pst-Decl
"I made an opera about a witch"

A. ət'ə-n næyoŋ-inte?
what content-int
"What is the story of the opera like?"

B. ki manya-ka minam $c^{h}$əŋnyən-il yuhokha-nin yæki-ya.
the witch-Subj handsome young man-Obj seduce-Adj P story-Decl
"The witch seduces a handsome young man in the story."

## Appendix 3

**han yes-iyaki**

acu oɹe can-e iyaki-ipnita kangwonto-e yaŋwal-iɹapulli-nin han
very long ago-AdjP story-Decl kangwonto-Loc yaŋwal- so called-AdjP a

mail-i iss-ess-ipnita. yep'in yain-i han saɹam coyoŋhi sal-ko iss-aess-ipnita.
village-Subj exist-Pst-Decl pretty woman-Subj a person calmly live-AdvP Prog-Pst-Decl

yepin yain-in əcil-ko maimsi $c^{h}$akhæ-ss-ipnita. aɹin mail saɹam-til-to
pretty woman-Top kind-Conj mind good-Pst-Decl other village person-Pl-also

caŋcikha-n maims'i-il kaci-ko iss-əss-ipnita. t'aɹasa i mail-in ki caŋcikham-iɹo
honest-AdjP mind-Obj have-AdvP Aux-Pst-Decl thus this village-Top that honesty-Instr

naɹa-an-e nəli alya-ca iss-ess-ipnita. naɹa ane ki iɹim-i hweca
country-in-Loc widely known-AdvP Aux-Pst-Decl country inside the name-Subj oftenmentioned

twe-ess-ipnita. yepin yain-in thikhi maim-i chakhæ-sə kananha-n saɹam-il
-Pst-Decl pretty woman-Top especially mind-Subj good-AdvP poor-AdjP person-Obj

po-myan kaci-n motin kess-il cu-a paɹi-kon hæ-ss-ipnita.
see-if have-AdjP all thing-Obj give-AdvP Aux-AdvP do-Pst-Decl

kiɹatan əni nal-i-ess-ipnita. $p^{h}$ok$p^{h}$uŋu-ka $c^{h}$it-ən nal-i-ess-ipnita. kamin
such one day-Subj-Pst-Decl storm-Subj fall-AdjP day-Subj-Pst-Decl black

iɹi-ka ətisanci natana-ss-ipnita. kamin iɹi-nin mu$c^{h}$ak $p^{h}$oakhæ-sa yain-kwa
wolf-Subj from nowhere appear-Pst-Decl black wolf-Top very violent-since woman-and

mail saɹam-til-il kweɹophi-ki sicak-hæ-ss-ipnita. yain-e mom-kwa maim-i
village person-Pl-Obj harass-NomP begin-do-Pst-Decl woman-Pos body-and mind-Subj

sweyakhæ ka-ss-ipnita. taɹin mail saram-til-to kweɹo-wa hæ-ss-ipnita.
get weak-Pst-Decl other village person-Pl-also show pain-Pst- Decl

i mail-in kot$^{h}$oŋ-iɹo katik c$^{h}$a iss-əss-ipnita. kam-in iɹi-nin əmc$^{h}$aŋnan
this village-Top pain-with fully filled Aux-Pst-Decl black wolf-Top tremendous

kot$^{h}$oŋ-il yəin-eke cu-ess-ipnita.
pain-Obj woman-Dat give-Pst-Decl

han tal-i cina-sə yəss-ipnita. nickail nal-i-ess-ipnita. yəp-mail
one month-Subj pass-AdvP Pst-Decl late fall day-Subj-Pst-Decl neighbor-village

c$^{h}$aŋnyan-i yoŋkamhito c$^{h}$ass-a wa-ss-ipnita. yəp-mail chaŋnyan-in yəin-e
young man-Subj bravely come-AdvP come-Pst-Decl neighbor-village young man-Top woman-Pos

c$^{h}$inku-ya-ss-ko him-i se-mya nallyaphæ-ss-ipnita. haciman iɹi-nin əni saɹam-potato
friend-Cop-Pst-and power-Subj strong-and agile-Pst-Decl but wolf-Top any person-than

him-i se-ss-ipnita. t'ohan əni nuku-potato minc$^{h}$aphæ-ss-ipnita. əni
power-Subj strong-Pst-Decl also any person-than agile-Pst-Decl any

nuku-to iɹi-lil ikil-sun əp-əss-ipnita. yəp-mail c$^{h}$aŋyan-in iɹi-wa sauki-l
person-Subj wolf-Obj defeat not able to-Decl neighbor-village young man-Top wolf-with fight

wonhæ-ss-ipnita. c$^{h}$aŋnyan-kwa iɹi-nin p$^{h}$ina-nin hyalt$^{h}$u-lil hæ-ss-ipnita. kiɹəna
want-Pst-Decl young man-Conj wolf-Top bloody-Top fight-Obj have-Pst-Decl but

c$^{h}$aŋnyan-in iɹi-e saŋtæ-nin twe-l su əp-əss-ipnita. chaŋnyan-in iɹi-e
young man-Top wolf-Pos match-Top cannot become-Pst-Decl young man-Top wolf-by

mul-lya kaii cukim-e iɹi-ke twe-əss-ipnita.
bitten-AdvP almost death-Loc reach-AdvP come to-Pst-Decl

paɹo i t'æ-y-əss-ipnita. paɹam-i mopsi pu-l t'æ-ya-ss-ipnita. nuɹ-
very this moment-Cop-Pst-Decl wind-Subj severely blow-AdjP time-Cop-Pst-Decl yellow

ən koyaŋi-ka ətisanci nat$^{h}$ana-ss-ipnita. nuɹə-n koyaŋi-nin ki c$^{h}$aŋnyan-e
AdjP cat-Subj from nowhere appear-Pst-Decl yellow-AdjP cat-Top the young man-Pos

kass-i-ess-ipnita. kam-in iɹi-nin c$^{h}$aŋnyaŋ-e mok-il mul-ko iss-əss-ipnita.
thing-Subj-Pst-Decl black wolf-Top young man-Pos neck-Obj bite-AdvP Prog-Pst-Decl

chaŋnyan-in kot$^{h}$oŋsiɹə-n pimyaŋ-il ciɹi-ko iss-əss-ipnita. kot$^{h}$oŋ-iɹo cansin-i
young man-Top painful scream-Obj make-AdvP Prog-Pst-Decl pain-due to whole body-Subj

t'ali-ko iss-ess-ipnita. nuɹə-n koyaŋi-nin iɹi-lil camsi noryo po-n hu iri-lil
tremble-AdvP Prog-Pst-Decl yellow-AdjP cat-Top wolf-Obj briefly stare-Adj P after wolf-Obj

kapcaki koŋkyakhæ-ss-ipnita. iɹi-nin mul-əss-tan c$^{h}$aŋnyan-il no-ayaman hæss-ipnita
suddenly attack-Pst-Decl wolf-Top bite-Pst-AdjP young man-Obj let go-have to-Pst-Decl
.
koyaŋi-wa c$^{h}$aŋnyan-in hamk'e iɹi-wa s'awa-ss-ipnita. kaii myəs sikan-il
cat-and young man-Top together wolf-with fight-Pst-Decl almost several hour-Obj

s'awa-ss-ipnita. hanil-e toum-i iss-ess-tanci tul-in iɹi-lil mulichi-l su ss-ess-ipnita.
fight-Pst-Decl heaven-Pos help-Sub exist-Pst-Subj two-Top wolf-Obj defeat-AdjP able to-Pst-Decl

kit'æ hanil-esə-n əksu kat-in pi-ka næɹi-ko iss-ess-ipnita.
then Heaven-Loc-Top torrential rain-Subj fall-AdvP Prog-Pst-Decl

**An Old Story**

This is a very old story. There was a village called "yeongweol" in Kangwon Province. A **pretty woman** was living there quietly. The **pretty woman** was kind and good-natured. Other villagers also had an honest mind. Therefore this village was widely known for its honesty throughout the country. People often talked about this village. The **pretty woman** was especially good-natured that she used to give everything she had when she saw a poor person.

It was one of those days. It was a day when it was storming. A **black wolf** appeared suddenly in the village. The **black wolf** was so violent that he began to harass the woman and other villagers. The body and mind of the woman became very weak. Other villagers were also in pain. This village was filled with pain. The **black wolf** gave tremendous pain to the woman.

One month passed. It was a late autumn day. A **neighbor-village-young-man** came here bravely. The **neighbor-village-young-man** was the woman's friend and powerful and agile. Yet the wolf was more powerful than any person. The wolf was also more agile than any person. Nobody was able to defeat the wolf. The **neighbor-village-young-man** wanted to fight with the wolf. The young-man and the wolf fought a bloody fight. But the young-man could not be a match to the wolf. The young-man was beaten by the wolf and in a near-death situation.

At this very moment, when it was violently windy, a **yellow cat** appeared from nowhere. The **yellow cat** belonged to the young-man. The black wolf was biting the man's neck. The young-man was screaming very painfully. The young-man's whole body was trembling from pain. The **yellow cat** stared at the wolf for a while and suddenly attacked the wolf. The wolf had to release the young-man it had been biting. The cat and the young-man fought together against the wolf. They fought for hours. Thanks to the Heaven's help, they were able to defeat the wolf. Rain was falling so torrentially from the sky at that time.

*OSU Working Papers in Linguistics* 45 (131-159)

# Longitudinal Study of the Acquisition of Taiwanese Initial Stops*

**Ho-Hsien Pan**
pan@ling.ohio-state.edu

**Abstract**: This study focuses on the Taiwanese stop system and is intended to provide information regarding the three-way voicing contrast of Taiwanese stops, as they are acquired by children. There are three voiceless unaspirated stops in Taiwanese, /p, t, k/, three voiceless aspirated stops /$p^h$, $t^h$, $k^h$/ and two voiced stops /b, g/. A longitudinal study of two Taiwanese-acquiring girls between the ages of 28 and 29 months to 33 and 40 months respectively, indicated that Taiwanese voiceless unaspirated stops were acquired first, voiceless aspirated stops secondly, and finally voiced stops. Even though the two girls acquired the three-way voicing contrast by the age of three years, their productions were not adult-like, especially for the voiced stops. There were many voiced stops produced with short lag VOTs as was found in the voiceless unaspirated stops.

## Introduction

The acquisition of stop consonant systems has been studied in various languages. It has been determined that the voiceless unaspirated stops are the first to be acquired by children in English (Macken & Barton, 1979), Spanish (Macken & Barton, 1980), Thai (Gandour et al., 1986), and Hindi ( Davis, in press; Srivastava, 1974). In general, these studies also found that if the language contrasts more than one other stop type with [b̥], [d̥], [g̊] children begin to acquire voiceless aspirated after acquiring the voiceless unaspirated stops(Gandour et al, 1986; Davis in press) and only later acquire voiced or breathy-voiced stops. The only exception is the study done by Srivastava with one Hindi-learning child, who acquired voiced stops before voiceless aspirated stops. In the study done by Gandour et al. (1986) on Thai, voiced stops were the last type to be acquired. Even after children developed the skill to produce voiced stops with lead VOT (voice onset time), their productions at the age of five were not as precise and well controlled as the adults'.

By comparing Taiwanese data with data of other languages, this study intends to determine what acquisition patterns Taiwanese shares with other languages. For example, by comparing the acquisition of Taiwanese voiceless stops and voiced stops, we will find out whether Taiwanese voiced stops take as long to be acquired, as Thai and Hindi voiced stops do; and also whether Taiwanese voiceless unaspirated stops are indeed first to be acquired, as are the voiceless unaspirated stops in Thai, Hindi, English, Spanish, and Cantonese.

*The longitudinal study was made possible by the support of Eden Saw, Tien Hsien Chang, and Tom Chang. Their patience and passion for preserving Taiwanese are deeply appreciated. I am also grateful to the two intelligent girls, EC and SC, who made the study a pleasant experience.

Macken (1980) proposed that in order to claim that some similar acquisition patterns are observed across languages, at least two things need to be coordinated: (1) the appropriate unit of instrumental analysis and (2) a language-independent means of comparing the acquisition process across languages. Follo2wing Macken's concerns, the phonetic unit of analysis used here is VOT. The language-independent criteria for claiming a child's production as adult-like are therefore (1) a consistent relationship between the different types of stops produced by the child, (e.g., the VOT of voiceless aspirated stops must always be longer than those of voiceless unaspirated stops); (2) the mean VOTs of the child's production must approximate the mean VOT values of the adults and (3) there must be no overlap between the VOT ranges of different stop types produced by the children.

A pilot longitudinal study was carried out to trace the acquisition processes of two children over an extended period of time. Unlike pooled data with a large number of children being recorded only once, by following the same children for a long time, a complete individual acquisitional process is documented.

## Method

### Subjects

Two girls, growing up in a Taiwanese speaking environment in the U.S.A. participated in the study. EC was 28 months old when the recording started, and SC was 29 months old. All of the parents come from the central and southern part of Taiwan. The production of EC's mother ES and SC's father TC were used as a control in the adult acoustic study.

### Data Collection

Each child was recorded until were able to consistently produce the three-way voicing distinction between homorganic stops and produced VOT values comparable to those of adult speakers. In the beginning of the study when the development was rapid, both children were recorded once every two weeks. Once the child began to make a statistically significant distinction between the three different types of stops (voiceless unaspirated, voiceless aspirated, and prenasalized voiced stops), the data was collected once every month.
EC was recorded with a SONY WM-D6C walkman, and a SONY ECM-144 microphone. Recording of EC was usually done by the experimenter, and occasionally by her mother ES at her home in Columbus, Ohio. Starting at the fifth month, EC was sometimes recorded using an AIWA cassette deck in a sound treated booth in the Linguistics Laboratory at the Ohio State University. SC was recorded at home by her father TC using an AIWA walkman, with a SONY ECM-144 microphone. The recording of SC was done by her father TC at their home in Los Angeles about every 10 days for a one-year period.

In the longitudinal study the tokens were elicited by having children imitate adult production. In the beginning of the study children learned several lexical items that were not originally in their vocabulary. Later on, whenever possible, spontaneous utterances were elicited by showing objects, or pictures.

## Corpus

Table 1. corpus for EC and SC. For EC the item /be˥˧/ 'horse' chosen to replace the original item /be˧/ 'gruel' from the 31 months, because EC did not produce the target stop at all.

| | | | | | |
|---|---|---|---|---|---|
| labial | | | | labial nasal | |
| /pe˥˧/ | 'crawl' | /pŋ˧/ | 'rice' | /mẽ˧/ | 'younger sister' |
| /pʰe˧/ | 'futon' | /pʰɔŋ˥˧/ | 'bath' | | |
| /be˥˧/ | 'horse' | /bŋ˧/ | 'door' | | |
| dental | | | | dental nasal | |
| /te˧/ | 'tea' | /tŋ˥˧/ | 'long' | /nẽ˥˧/ | 'milk' |
| /tʰe˧/ | 'take' | /tʰŋ˥˧/ | 'soup' | | |
| velar | | | | velar nasal | |
| /ke˥˧/ | 'chicken' | /kŋ˥˧/ | 'light' | /ŋ˥˧/ | 'yellow' |
| /kʰe˧˩/ | 'guest' | /kʰŋ˧˩/ | 'put' | | |
| /ge˧/ | 'moon' | /gɔŋ˥˧/ | 'dizzy' | | |

## Data analysis

The child's productions were categorized according to the adult target words. Even if the child produced a [p] when imitating the adult target word with initial /pʰ/, the production was categorized as an instance of /pʰ/. A token was discarded when the child made a mistake in place of articulation. For example, tokens of /tʰe˧/ ' take' that were produced as /kʰe˧˩/ 'guest' were excluded from further analysis. The recordings were analyzed acoustically to obtain spectral and temporal measures, including VOT. The VOT was measured by using a KAY 5500 DSP sonograph, and the measurements were statistically analyzed.

### Adult Norm VOT Values

According to the results of a study on Taiwanese adult VOT production, adult mean VOT values range from 0 to 30 ms for voiceless unaspirated stops, from 40 to 100 ms for voiceless aspirated stops, and from -100 to -30 ms for voiced stops.

### Subject EC

## Statistical Results

The target item /bŋ˧/ 'door' was eliminated from the statistical analysis because it was produced as /mŋ˧/ ([məŋ˧]), with an initial nasal rather than an initial voiced stop by the adults in EC's environment. Data produced in the same month were pooled together to form a data set. There were all together, six month's worth of data recorded, representing six separate data sets. ANOVAs were done to determine the effect of voicing category, place of articulation, and following segments on each month's data. A three-way ANOVA with factors *voicing category, place of articulation* and *following segments* was not possible because of the lack of any /b/ initial tokens from 28 to 30 months. Therefore, six two-way ANOVAs with the factors *voicing category* and *following segments*, and 15 one-way ANOVAs with the factor *place of articulation* were used to analyze the VOTs of stops produced at three different places of articulation for each month.

There was a limited amount of data obtained from the voiced stop /b/ - due to the fact that EC's mother produced the word 'gruel' as [muə˥] instead of [be˥]. EC refused to produce any tokens of /be˥/ 'gruel', when asked to repeat the word. The item 'gruel' was changed to /be˧˥/ 'horse' in the 31st month. Since EC did not produce any voiced labial stops during the 28, 29, and 30 months, only three one-way ANOVAs with the factor *place of articulation* were used to analyze the VOT values of voiced stops at 31, 32, and 33 months. Since 24 ANOVAs were used to analyze EC's data, to avoid getting significant results by chances, the $\alpha$ was adjusted to 0.001 level.

Let us first consider the results of the two-way ANOVAs for each month. The results of these six ANOVAs indicated a significant main effect of voicing 2category for all six months: the 28th month ($F(2, 45) = 9.69$, $p<0.001$), the 29th month ($F(2, 83) = 14.86$, $p<0.001$), the 30th month ($F(2, 114) = 37.52$, $p<0.001$), the 31st month ($F(2, 45) = 9.69$, $p= 0.003$), the 32nd month ($F(2, 44) = 79.35$, $p<0.001$), and the 33rd month ($F(2, 101) = 112.48$, $p<0.001$).

Six post-hoc Tukey tests were done on the data of each month to check the source of the significant main effect of voicing category. The Tukey test shows that at 28 months, there is already a statistically significant distinction between the mean VOTs of aspirated stops and unaspirated stops. However, there is no distinction between the mean VOTs of the voiced stops and the voiceless unaspirated stops, or between voiced and voiceless aspirated stops. From 29 to 31 months, the unaspirated stops are significantly different from aspirated stops, and the voiced stops are also different from aspirated stops. However, there is still no distinction between stops with voicing lead and stops with short lag VOT. Starting at 32 months, the difference between voiced stops and voiceless unaspirated stops is significant. EC continued to distinguish between the three types of voicing manner at 33 months.

Results of the six ANOVAs indicated no significant main effect of following segments for any of the six data sets: the 28th month ($F(1, 45) = 0.74$, $p = 0.394$), the 29th month ($F(1, 45) = 1.12$, $p = 0.348$), the 30th month ($F(1, 45) = 0.16$, $p = 0.688$), the 31st month ($F(1, 45) = 0.74$, $p = 0.394$), the 32nd month ($F(1, 45) = 0.89$, $p = 0.352$), and the 33rd month ($F(1, 45) = 2.73$, $p = 0.102$).

Six one-way ANOVAs with the factor place of articulation were also used to analyze the VOT values of voiceless unaspirated stops at each month. Result of the six ANOVAs indicated no significant effects of place of articulation for the VOT values of voiceless unaspirated stops for the six months: 28 months ($F(2, 25) = 1.51$, $p = 0.240$), 29 months ($F(2, 32) = 4.85$, $p = 0.014$), 30 months ($F(2, 48) = 1.1$, $p = 0.343$), 31 months ($F(2, 18) = 1.68$, $p = 0.215$), 32 months, and at 33 months ($F(2, 12) = 0.61$, $p = 0.561$). Table 2 shows the mean VOTs of voiceless unaspirated labial, dental, and velar stops for each of the six months.

Table 2 EC's monthly mean VOTs broken down by voicing category and place of articulation

**28 months**

| PLACE | Unaspirated | Aspirated | Voiced |
|---|---|---|---|
| LABIAL | | | |
| VOT | 20.315 | 17.19 | -- |
| S.D. | (2.21) | (--) | |
| N | 2 | 1 | |
| DENTAL | | | |
| VOT | 15.107 | 66.01 | -- |
| S.D. | (5.92) | (33.55) | |
| N | 3 | 8 | |
| VELAR | | | |
| VOT | 22.146 | 50.39 | 28.12 |
| S.D. | (6.84) | (24.06) | (--) |
| N | 23 | 12 | 1 |

**29 months**

| PLACE | Unaspirated | Aspirated | Voiced |
|---|---|---|---|
| LABIAL | | | |
| VOT | -15.82 | 51.56 | -- |
| S.D. | (55.03) | (46.61) | |
| N | 8 | 23 | |
| DENTAL | | | |
| VOT | 21.35 | 87.19 | -- |
| S.D. | (8.08) | (35.42) | |
| N | 9 | 10 | |
| VELAR | | | |
| VOT | 34.95 | 79.58 | 24.33 |
| S.D. | (38.67) | (35.83) | (7.32) |
| N | 18 | 14 | 7 |

Table 2 (continued)

**30 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 11.989 | 83.81 | -- |
| S.D. | (25.30) | (37.56) | |
| N | 9 | 25 | |
| DENTAL | | | |
| VOT | 26.563 | 79.68 | -- |
| S.D. | (14.82) | (35.21) | |
| N | 22 | 16 | |
| VELAR | | | |
| VOT | 19.218 | 68.47 | 18.75 |
| S.D. | (34.50) | (50.90) | (30.48) |
| N | 20 | 17 | 11 |

**31 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 15.10 | 17.19 | -14.06 |
| S.D. | (16.08) | (--) | (64.08) |
| N | 6 | 1 | 5 |
| DENTAL | | | |
| VOT | 25.69 | 66.01 | -- |
| S.D. | (23.98) | (33.55) | |
| N | 9 | 8 | |
| VELAR | | | |
| VOT | 23.44 | 50.39 | 24.66 |
| S.D. | (33.08) | (24.06) | (22.14) |
| N | 9 | 12 | 10 |

Table 2 (continued)

**32 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 27.502 | 68.43 | -49.595 |
| S.D. | (22.55) | (34.03) | (37.27) |
| N | 5 | 5 | 5 |
| D2ENTAL | | | |
| VOT | 20.938 | 102.22 | -- |
| S.D. | (10.69) | (38.37) | -- |
| N | 5 | 7 | |
| VELAR | | | |
| VOT | 32.188 | 68.55 | -39.843 |
| S.D. | (12.91) | (34.14) | (19.34) |
| N | 5 | 8 | 10 |

**33 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 19.443 | 81.25 | -25.347 |
| S.D. | (9.85) | (59.33) | (28.65) |
| N | 9 | 16 | 9 |
| DENTAL | | | |
| VOT | 27.213 | 64.95 | -- |
| S.D. | (11.87) | (22.49) | -- |
| N | 12 | 14 | |
| VELAR | | | |
| VOT | 30.468 | 100.61 | -43.748 |
| S.D. | (18.65) | (36.65) | (40.53) |
| N | 14 | 15 | 18 |

Another set of six one-way ANOVAs with the factor place of articulation were used to analyze the VOT values of voiceless aspirated stops at each of the six months. Result of the six ANOVA also indicated no significant effect of place of articulation for the VOTs of voiceless aspirated stops for all six months: 28 months ($F(2, 18) = 1.68$, $p = 0.215$), 29 months ($F(2, 44) = 3.41$, $p = 0.042$), 30 months ($F(2, 55) = 0.71$, $p = 0.495$), 31 months ($F(2, 18) = 1.68$, $p = 0.215$), 32 months, and 33 months ($F(2, 17) = 2.03$, $p = 0.162$).

Result of the three ANOVAs indicated no significant effect of place of articulation for the VOTs of voiceless aspirated stops for 31 to 33 months: 31 months ($F(1, 19) = 1.68$, $p = 0.215$), 32 months ($F(1, 13) = 0.46$, $p = 0.509$), 33 months ($F(1, 25) = 1.47$, $p = 0.236$ ).

Three regressions were performed to determine if there was a significant relationship between *age* and *VOT development* of voiceless unaspirated stops, between *age* and *VOT development* of voiceless aspirated stops, and between *age* and *VOT development* of voiced stops. The regressions show significant linear relationships between age and VOT values of unaspirated stops, and aspirated stops (Table 3).

Table 3 EC summary table of ANCOVA analysis examining effect of age on VOT

| | EFFECT OF AGE ON VOT | |
|---|---|---|
| | Linear Trend | Linear & Quadratic Trend |
| UNASPIRATED | | |
| R-Square | 0.438 | |
| Sig. level | P<0.0001 | |
| ASPIRATED | | |
| R-Square | 0.771 | |
| Sig. level | P<0.0001 | |
| VOICED | | |
| R-Square | 0.088 | 0.396 |
| Sig. level | P=0.081 | P<0.0001 |

As the age increases, the VOTs of /p, t, k, $p^h$/ increase, the VOTs of /$t^h$/ decrease, the VOTs of /$k^h$/ stay relatively the same (Figure 1). For voiced stops, there is no significant linear relationship between age and VOT values. By multiplying the age to the second order, a significant quadratic relationship was found. The curve fitted through the mean VOTs of /g/ in Figure 1 shows the relationships between age and VOT developments. At the 31 and 32 months, the /b/ tokens were sampled with mean VOT around -30 ms. At the 33 months the mean VOT value of /b/ increased toward -15 ms as shown in Figure 1. There was a quadratic trend type of VOT development through the three months. As for /g/ the mean VOTs stayed relatively the same around 30 ms from 28 to 30 months. At 31 months, the mean VOT value suddenly decreased to around 0 ms, and the mean VOT values continued to decrease during 32 months.

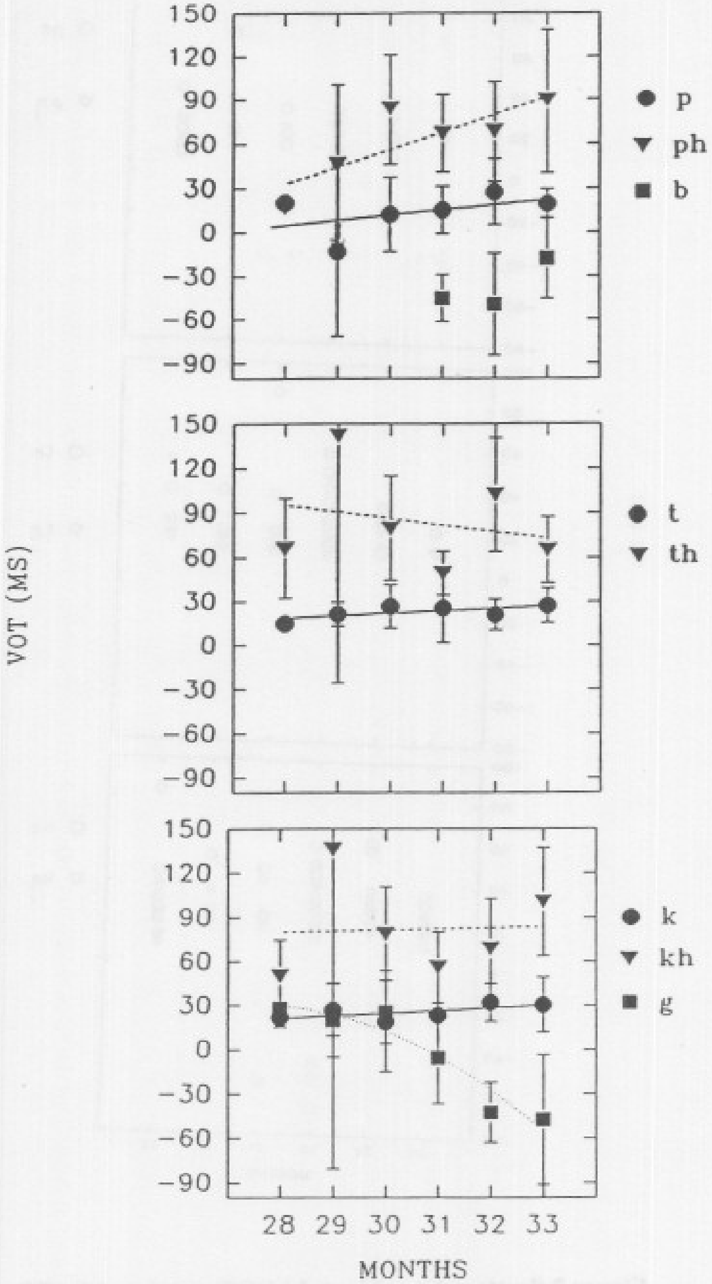


**Figure 1 VOT development for EC, means and standard deviations**

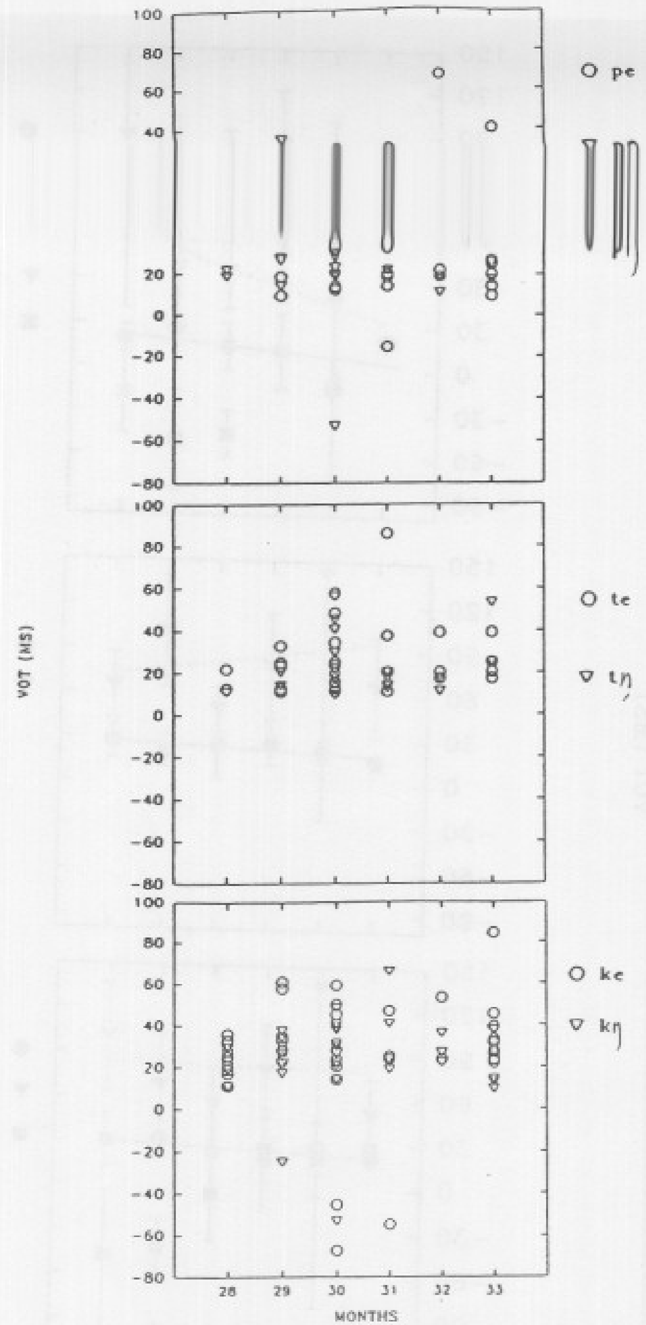



Figure 2 Scatter plots of /p, t, k/ VOTs produced by EC

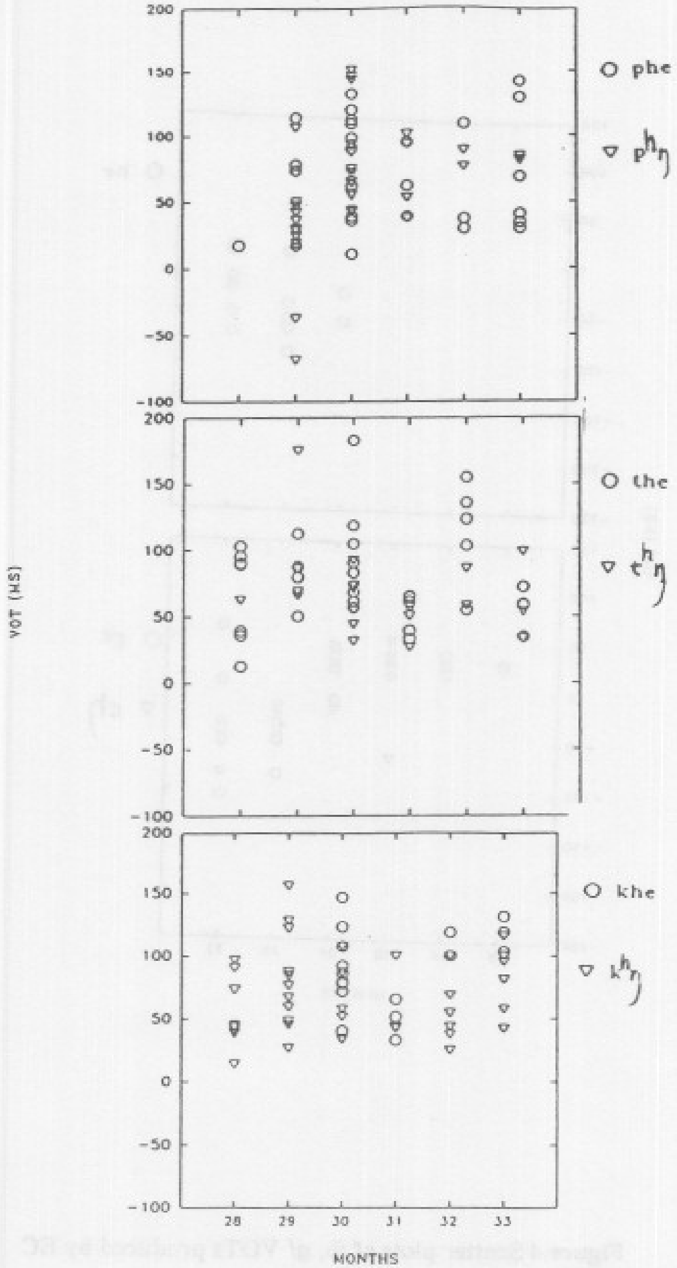



**Figure 3 Scatter plots of /$p^h$, $t^h$, $k^h$/ VOTs produced by EC**

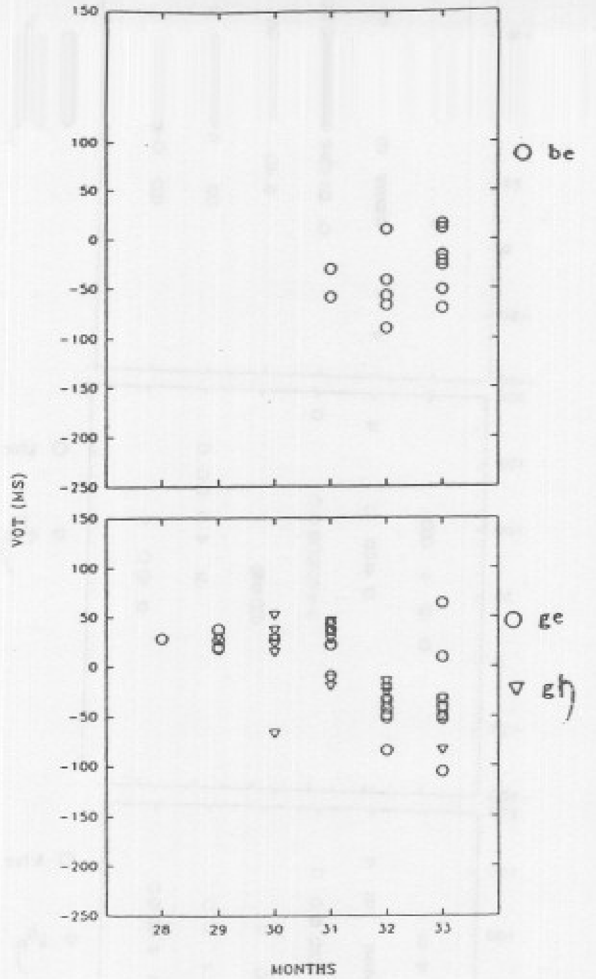



**Figure 4 Scatter plots of /b, g/ VOTs produced by EC**

**General Tendencies of EC's Data**

At the beginning of this study, EC's VOTs for unaspirated stops ranged between 0 to 40 ms. These were lengthened during the 30th and 31st month, when EC produced aspirated stops with VOTs exceeding the adult upper VOT limit for aspirated stops, which is 100 ms. The production of extra long aspiration will be referred to as hyperaspiration hereon. In the 32nd month, the VOTs were shortened. Some tokens of unaspirated stops were produced as nasals during 29 to 31 months.

As for the aspirated stops, in the 28th month, except for /pʰ/, the VOT values were within the adult norm. But in the 29th month, the individual VOT values exceeded the adult norm, which is 100 ms. EC kept on increasing the VOTs of aspirated stops from 28 through out 30 months. Depending on her familiarity with a word, the VOT peak was reached either in the 29th or the 30th month. In words that she was familiar with, e.g. /kʰɲ˧/ 'put' in Figure 3, the VOT peak was reached earlier than in unfamiliar words, like /pʰe/ 'futon'. By 31 months EC shortened the hyperaspiration for the aspirated stops.

EC started to produce /b/ with lead VOTs in the 31st month. At 33 months the VOT values of /b/ increased again. From 28 to 30 months most of the /g/ tokens were produced with short lag VOTs. During the 31 and 32 months, the VOTs of /g/ suddenly dropped. By the 33rd month, there was a clear distinction among the three voicing contrast of stops.

Macken (1980) proposed three categories for VOT development. During category I, there is no distinction between the aspirated and unaspirated stops. In category II, there is a distinction, but the VOT for both the aspirated and unaspirated stops fall within the adult norm for unaspirated stops (1;7 to 1;9 Tom; 1;5 to 1;9, Tessa; 1;7 to 1;10, Jane). In category IIIA, the children hyperaspirated the stops with long lag VOT (1;9, Tom, 1;9 to 2;0 Tessa, 1;10 to 2;0, Jane), They began to shorten the VOT back to the adult norm (category IIIB) (1;11 to 2;1, Tom). Generally speaking the data for EC agree with the category III pattern defined by Macken.

**Subject SC**

**Statistical Results**

Data produced in the same month were pooled. Two-way ANOVAs with the factors voicing category and following segments, and one-way ANOVAs with the factor place of articulation were used to analyze the VOTs of each of the 12 months data. Results of twelve two-way ANOVAs (voicing category * following segments) indicated significant main effects of voicing category for the data of the 12 months, as shown in Table 4.

Table 4 Significant main effects of voicing category for each months' data

| Months | Effects of voicing category |
|---|---|
| 29 months | (F (2 , 58) = 4.32, P = 0.018) |
| 30 months | (F (2, 183) = 75.47, P< 0.001) |
| 31 months | (F (2, 105) = 62.30, P< 0.001) |
| 32 months | (F (2, 107) = 108.91, P< 0.001) |
| 33 months | (F (2, 119) = 50.73, P< 0.001) |
| 34 months | (F (2, 95) = 48.27, P< 0.001) |
| 35 months | (F (2, 94) = 55.85, P< 0.001) |
| 36 months | (F (2, 42) = 20.63, P< 0.001) |
| 37 months | (F (2, 175) = 59.64, P< 0.001) |
| 38 months | (F (2, 178) = 90.64, P< 0.001) |
| 39 months | (F (2, 173) = 36.85, P< 0.001) |
| 40 months | (F (2, 64) = 57.47, P< 0.001) |

For data of every month, post-hoc tests were used to check if there are significant differences between VOTs of stops with different voicing categories. The post-hoc Tukey tests showed a significant difference between voiceless aspirated and unaspirated stops from the 29th to the 40th month. There were significant differences between the VOTs of voiceless aspirated stops and voiced stops. A significant difference between mean VOTs of voiceless unaspirated stops and voiced stops did not show up until the 34th month. However, at 39 and 40 months there were again no longer any significant differences between the mean VOTs of voiced stops and voiceless unaspirated stops.

Results of the 12 two-way ANOVAs (voicing categories * following segment) indicated no significant main effects of the following segment for VOTs for any months except for the 33rd month. There was a significant effect of following segments on VOTs at 33 months ( $F(1, 119) = 11.07$, $p = 0.0012$ ).

Results of 12 one-way ANOVAs indicated that there were no significant main effects of place of articulation for the VOTs of voiceless unaspirated stops for any months except for the 32nd and 38th month. There were significant main effects of place of articulation for the VOTs of voiceless unaspirated stops for the 32nd month ($F (2, 52) = 7.00$, $p = 0.002$). and for the 38th month ($F (2, 77) = 10.09$, $p < 0.001$). Table 5 shows the monthly mean VOTs of voiceless unaspirated labial stops, dental stops, and velar stops.

A post-hoc test showed that at 32 and 38 months, the VOTs of voiceless unaspirated velar stops, were significantly different from those of voiceless unaspirated alveolar and labial stops.

Results of the 12 one-way ANOVA indicated that there were no significant main effects of place of articulation for the VOTs of voiceless aspirated stops for any months except for the 31st and the 39th month. There were significant main effects of place of articulation for the VOTs of voiceless aspirated stops produced at the 31st month ($F (2, 32) = 12.29$, $p< 0.001$), and the 39th month ($F (2, 64) = 12.8$, $p< 0.001$).

Table 5 SC's monthly mean VOTs broken down by voicing category and place of articulation effect of places of articulation

**29 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 16.620 | 60.594 | -- |
| S.D. | (5.24) | (40.98) | |
| N | 11 | 9 | |
| DENTAL | | | |
| VOT | 23.213 | 47.851 | -- |
| S.D. | (11.25) | (22.18) | |
| N | 7 | 8 | |
| VELAR | | | |
| VOT | 25.293 | 58.854 | -- |
| S.D. | (11.61) | (20.85) | |
| N | 16 | 9 | |

**30 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 24.672 | 71.420 | 13.907 |
| S.D. | (11.99) | (34.31) | (5.77) |
| N | 19 | 28 | 10 |
| DENTAL | | | |
| VOT | 22.823 | 65.381 | -- |
| S.D. | (12.82) | (31.30) | |
| N | 29 | 26 | |
| VELAR | | | |
| VOT | 30.834 | 67.463 | 33.306 |
| S.D. | (13.50) | (26.03) | (23.21) |
| N | 30 | 23 | 19 |

Table 5 (continued)

**31 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 25.915 | 111.36 | 21.615 |
| S.D. | (25.64) | (28.87) | (13.97) |
| N | 12 | 11 | 12 |
| DENTAL | | | |
| VOT | 15.075 | 56.76 | -- |
| S.D. | (6.20) | (21.73) | |
| N | 17 | 13 | |
| VELAR | | | |
| VOT | 28.312 | 73.865 | 25.954 |
| S.D. | (22.10) | (31.21) | (9.88) |
| N | 17 | 11 | 18 |

**32 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 17.450 | 89.316 | 13.869 |
| S.D. | (6.53) | (25.63) | (3.58) |
| N | 12 | 12 | 8 |
| DENTAL | | | |
| VOT | 17.118 | 63.635 | -- |
| S.D. | (6.58) | (18.02) | |
| N | 22 | 11 | |
| VELAR | | | |
| VOT | 26.935 | 69.923 | 26.249 |
| S.D. | (12.83) | (30.80) | (17.43) |
| N | 21 | 12 | 15 |

Table 5 (continued)

**33 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 18.751 | 66.586 | 7.618 |
| S.D. | (15.19) | (36.08) | (16.50) |
| N | 17 | 13 | 16 |
| DENTAL | | | |
| VOT | 24.541 | 66.184 | -- |
| S.D. | (22.13) | (36.24) | |
| N | 17 | 14 | |
| VELAR | | | |
| VOT | 21.604 | 78.280 | 11.167 |
| S.D. | (6.17) | (38.12) | (33.01) |
| N | 23 | 10 | 15 |

**34 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 24.689 | 67.883 | -7.924 |
| S.D. | (17.46) | (27.60) | (30.40) |
| N | 10 | 9 | 14 |
| DENTAL | | | |
| VOT | 14.510 | 48.924 | -- |
| S.D. | (4.65) | (26.73) | |
| N | 14 | 16 | |
| VELAR | | | |
| VOT | 20.899 | 66.875 | -4.296 |
| S.D. | (10.35) | (28.89) | (22.46) |
| N | 16 | 10 | 12 |

Table 5 (continued)

**35 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 11.831 | 61.062 | 0.976 |
| S.D. | (4.85) | (28.62) | (17.08) |
| N | 7 | 17 | 8 |
| DENTAL | | | |
| VOT | 19.401 | 52.675 | -- |
| S.D. | (8.77) | (29.93) | |
| N | 19 | 14 | |
| VELAR | | | |
| VOT | 21.87 | 61.408 | -3.282 |
| S.D. | (7.73) | (16.47) | (30.43) |
| N | 21 | 15 | 10 |

**36 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 15.623 | 62.480 | -14.304 |
| S.D. | (4.13) | (68.48) | (22.72) |
| N | 3 | 2 | 13 |
| DENTAL | | | |
| VOT | 13.802 | 51.560 | -- |
| S.D. | (9.24) | (35.36) | |
| N | 6 | 2 | |
| VELAR | | | |
| VOT | 15.108 | 62.498 | -3.023 |
| S.D. | (7.05) | (38.93) | (25.07) |
| N | 3 | 4 | 15 |

Table 5 (continued)

**37 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 9.233 | 60.750 | -3.571 |
| S.D. | (17.78) | (44.01) | (27.67) |
| N | 22 | 19 | 35 |
| DENTAL | | | |
| VOT | 17.313 | 51.485 | -- |
| S.D. | (9.06) | (31.91) | |
| N | 25 | 20 | |
| VELAR | | | |
| VOT | 21.237 | 67.018 | 3.173 |
| S.D. | (20.24) | (29.72) | (27.26) |
| N | 22 | 19 | 32 |

**38 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 12.579 | 64.835 | 0.500 |
| S.D. | (6.073) | (27.20) | (15.07) |
| N | 20 | 15 | 25 |
| DENTAL | | | |
| VOT | 16.965 | 74.217 | -- |
| S.D. | (16.56) | (33.32) | |
| N | 28 | 24 | |
| VELAR | | | |
| VOT | 27.392 | 91.407 | 10.623 |
| S.D. | (11.00) | (42.21) | (46.08) |
| N | 32 | 20 | 20 |

Table 5 (continued)

**39 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 13.750 | 78.219 | 5.981 |
| S.D. | (11.61) | (43.00) | (22.39) |
| N | 15 | 17 | 29 |
| DENTAL | | | |
| VOT | 10.119 | 31.581 | -- |
| S.D. | (18.75) | (25.38) | |
| N | 21 | 27 | |
| VELAR | | | |
| VOT | 22.498 | 67.121 | 17.559 |
| S.D. | (18.61) | (31.59) | (31.85) |
| N | 26 | 23 | 21 |

**40 months**

| PLACE | Voicing Categories | | |
|---|---|---|---|
| | Unaspirated | Aspirated | Voiced |
| LABIAL | | | |
| VOT | 15.626 | 97.034 | 9.374 |
| S.D. | (11.40) | (36.62) | (2.61) |
| N | 6 | 10 | 6 |
| DENTAL | | | |
| VOT | 17.709 | 71.309 | -- |
| S.D. | (7.93) | (42.55) | |
| N | 9 | 11 | |
| VELAR | | | |
| VOT | 18.404 | 86.875 | -4.786 |
| S.D. | (12.30) | (35.97) | (19.70) |
| N | 9 | 10 | 9 |

Post-hoc tests indicated that the VOTs of voiceless unaspirated labial, dental, and velar stops are different from each other at the 31st month. The VOTs of voiceless unaspirated dental stops are significantly different from those of voiceless unaspirated labial and velar stops at the 39 months.

SC did not produce any voiced stops at 29 months. Eleven one-way ANOVAs were used to check the main effect of place of articulation for the VOTs of voiced stops produced from 30 to 40 months. Results of these 11 one-way ANOVAs indicated that there were no significant main effects of place of articulation for the VOTs of voiced unaspirated stops from the 31st month on. There was a significant effect of place of articulation on voiced stops produced at the 30th month ($F(1, 27) = 6.66$, $p = 0.016$). Table 5 shows the monthly mean VOTs of voiced labial and velar stops.

For each of the three data sets a regression was used to check if there is a significant relationship between age and the VOT developments. The regressions found significant linear relationships between age and the VOTs of voiceless unaspirated, voiceless aspirated, and voiced stops. There were gradual decreases of VOTs for /p, b, t, k, g/. There were gradual increase of VOTs for /$p^h$, $t^h$, $k^h$/. Table 6 shows the result of regression analyses.

Table 6 SC summary table of ANCOVA analysis examining effect of age on VOT

| EFFECT OF AGE ON VOT | |
|---|---|
| | Linear Trend |
| UNASPIRATED | |
| R-Square | 0.648 |
| Sig. level | P<0.001 |
| ASPIRATED | |
| R-Square | 0.781 |
| Sig. level | P<0.001 |
| VOI | |
| R-Square | 0.060 |
| Sig. level | P<0.001 |

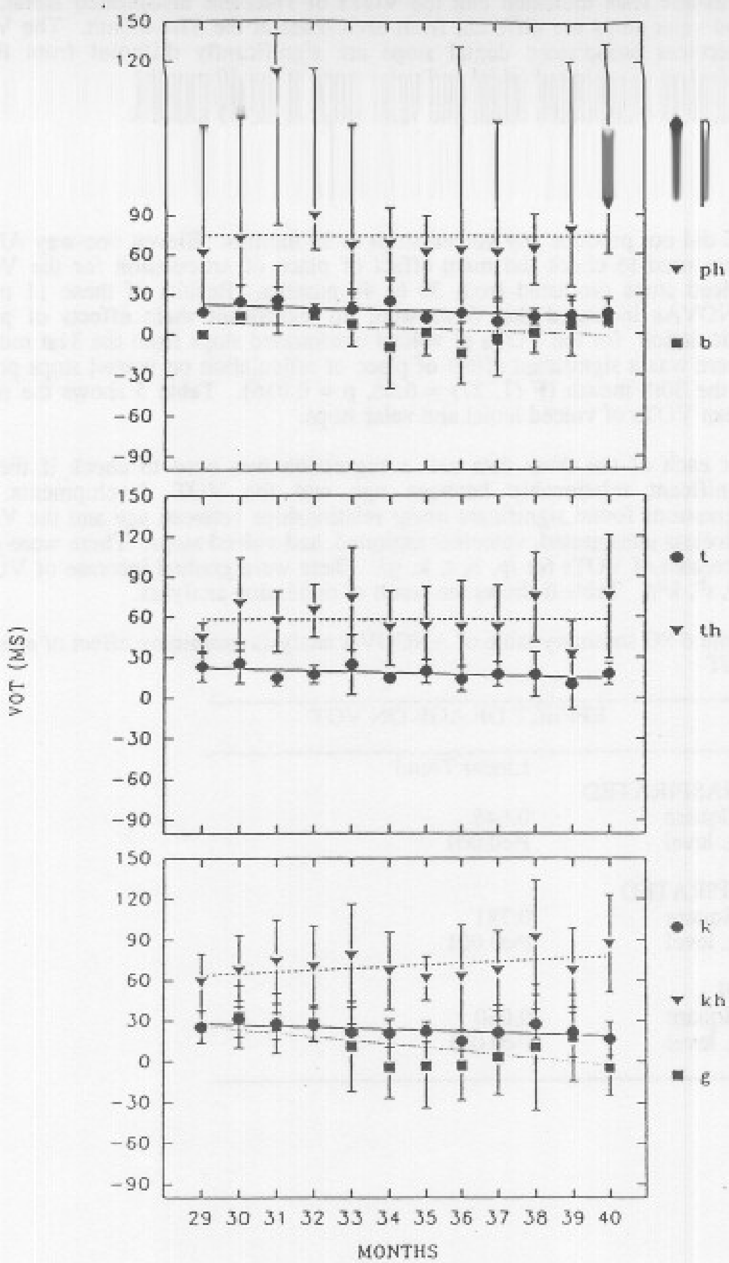



**Figure 5 VOT development for SC, means and standard deviations**

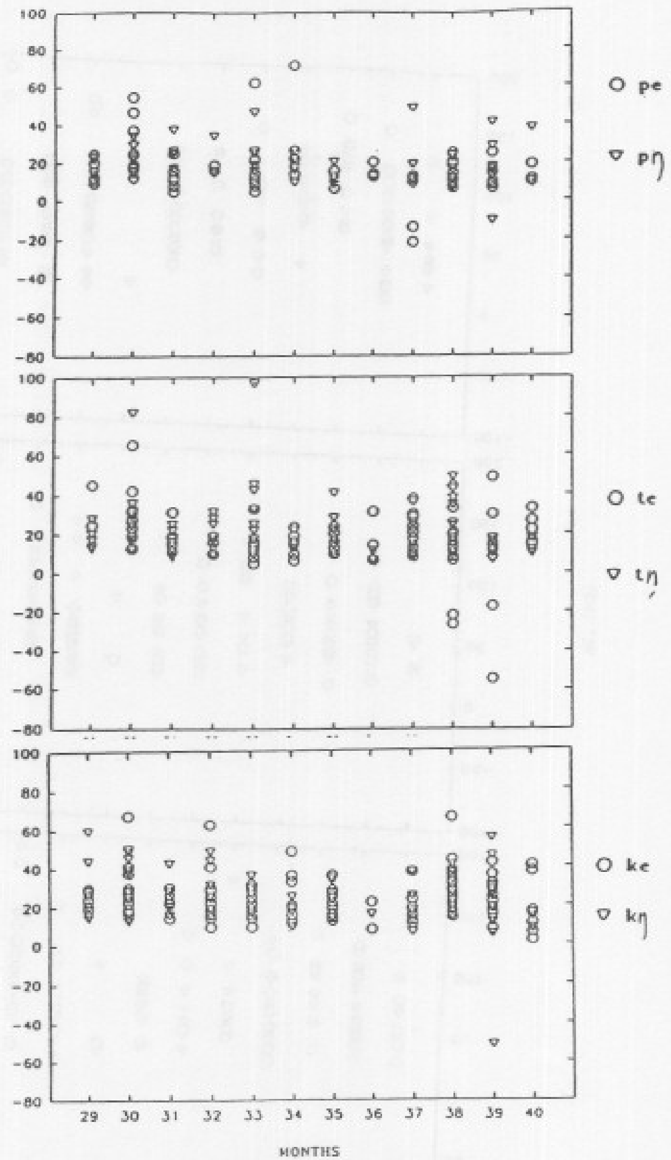



**Figure 6 Scatter plots of /p, t, k/ VOTs produced by SC**

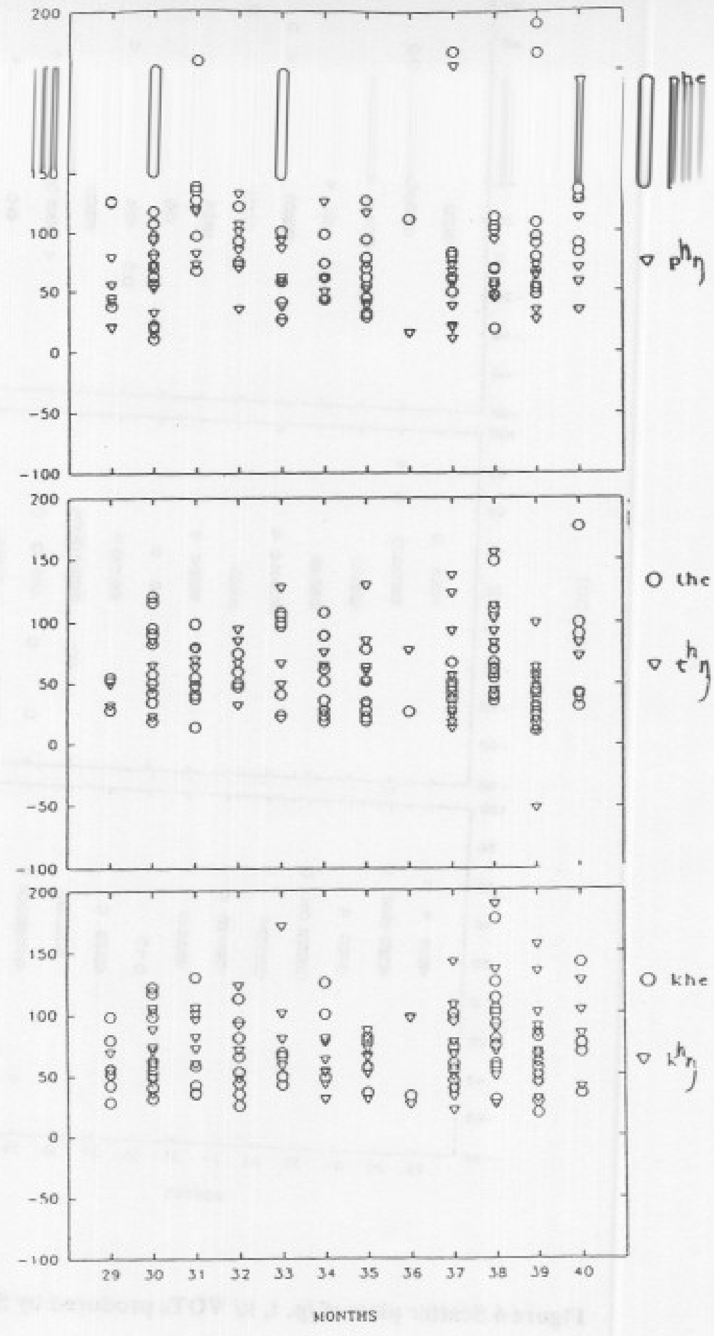



Figure 7 Scatter plots of /$p^h$, $t^h$, $k^h$/ VOTs produced by SC

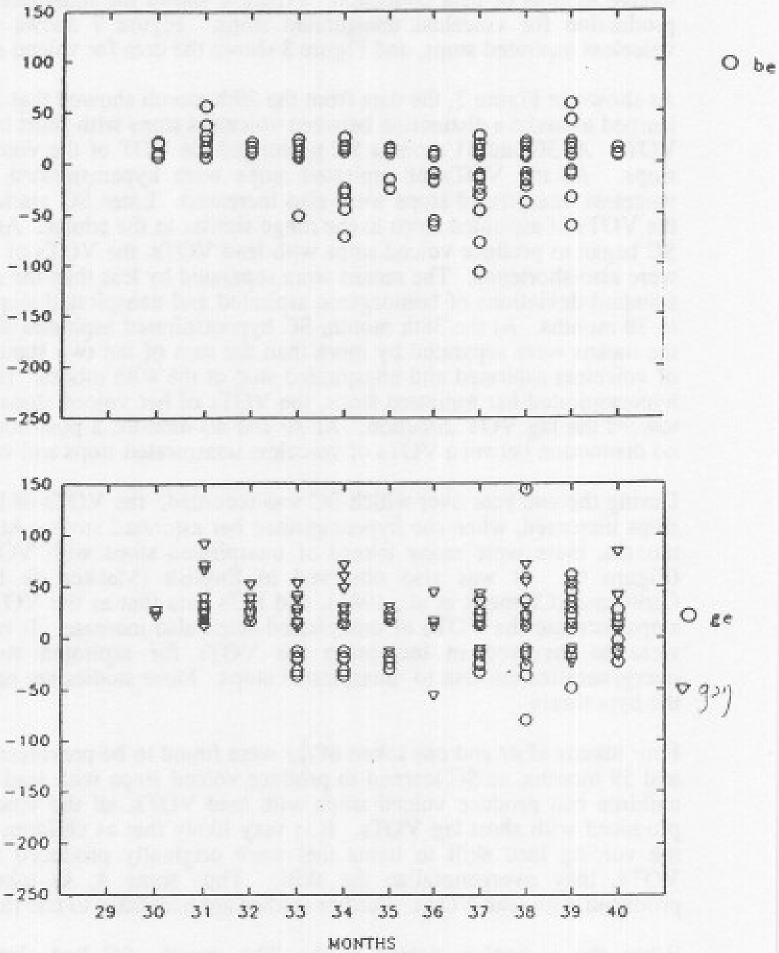



**Figure 8 Scatter plots of /b, g/ VOTs produced by SC**

## General Tendencies of SC's Data

Figure 5 shows the means and standard deviations of SC's VOTs during the twelve months of data collection. Figure 6 shows the individual tokens of SC production for voiceless unaspirated stops. Figure 7 shows the tokens for voiceless aspirated stops, and Figure 8 shows the data for voiced stops.

As shown in Figure 5, the data from the 29th month showed that SC had already learned to make a distinction between voiceless stops with short lag and long lag VOT. At 30 and 31 months SC prolonged the VOT of the voiceless aspirated stops. As the VOTs of aspirated stops were hyperaspirated, the VOTs of voiceless unaspirated stops were also increased. Later SC gradually shortened the VOTs of aspirated stops to the range similar to the adults'. At 34 months, as SC began to produce voiced stops with lead VOTs, the VOTs of aspirated stops were also shortened. The means were separated by less than the sum of the two standard deviations of homorganic aspirated and unaspirated stops, from the 35 to 38 months. At the 38th month, SC hyperaspirated aspirated stops again, and the means were separated by more than the sum of the two standard deviations of voiceless aspirated and unaspirated stop at the 40th month. However, as SC hyperaspirated her aspirated stops, the VOTs of her voiced stops also increased toward the lag VOT direction. At 39 and 40 months, a post-hoc test indicated no distinction between VOTs of voiceless unaspirated stops and voiced stops.

During the one year over which SC was recorded, the VOTs of her unaspirated stops increased, when she hyperaspirated her aspirated stops. At 30, 38 and 39 months, there were many tokens of unaspirated stops with VOT over 40 ms (Figure 6). It was also observed in English (Macken & Barton, 1979), Cantonese (Clumeck et al., 1981), and EC's data that as the VOTs of aspirated stops increase the VOTs of unaspirated stops also increase. It is as if children were so involved in increasing the VOTs for aspirated stops, that they overgeneralize the task to unaspirated stops. More studies are necessary to test the hypothesis.

Four tokens of /t/ and one token of /k/ were found to be prevoiced during the 38 and 39 months, as SC learned to produce voiced stops with lead VOT. Before children can produce voiced stops with lead VOTs, all the voiced stops were produced with short lag VOTs. It is very likely that as children learn to apply the voicing lead skill to items that were originally produced with short lag VOTs, they overgeneralize the skill. Thus some /t, k/ tokens were also produced with lead VOTs. Further studies are necessary to test this hypothesis.

When the recording started at the 29th month, SC had already begun to distinguish between voiceless aspirated and unaspirated stops. She continued to increase the VOTs of aspirated stops until the 31st month the VOTs are shortened back. During the 35th month to 37th month the lower limit of the VOTs for $/k^h/$ and $/p^h/$ began to decrease. There was more overlap between voiceless aspirated and unaspirated stops. In the 38th month SC hyperaspirated the aspirated stops again to distinguish between voiceless aspirated and unaspirated stops.

SC did not produce voiced stops with lead VOTs until the 33rd month. Before then all the voiced stops had been produced with short lag VOTs. Even though there was a statistically significant three way distinction during the 34th to 38th month, SC started again to produce most voiced stops with lag VOTs at the 39th and 40th month, and so that the difference between VOTs of voiced stops and voiceless unaspirated stops was no longer significant. The difference is probably lost because of SC's exposure to English. During 39 and 40 months, SC began to speak English, and occasionally addressed in English, too.

## Conclusion

The developmental patterns of voicing distinction in stop consonants were very different in SC and EC. When the recording started neither EC nor SC produced stops with lead VOTs. It took EC 6 months to acquire the three-way voicing contrast. Even at 31 months, when EC had learned to produce voiced stops with lead VOTs, the mean VOTs for voiced stops continued to decrease. There was less and less overlap between the standard deviations of homorganic stops. She seemed to go through a new "stage" every month. By the 33rd month, a clear three way voicing distinction of stops was achieved. As for SC, even after she learned to produce voiced stops with lead VOTs, the mean VOTs of voiced stops remained relatively the same and did not decrease very much. There was always overlap between the standard deviation bars for the homorganic stops of different types. Even though post-hoc Tukey comparison showed significant differences between the mean VOTs of voiceless unaspirated stops and voiced stops during the 34th to 38th month, the two stops were indistinguishable again in the 39th and 40th month. This loss of a three-way voicing distinction is very likely attributed to the exposure of English. Even though for EC a clear three-way voicing distinction was maintained at 33 months. It is very likely that the distinction could have been lost had the study continued for a longer period, because like SC, EC also began to learn English at 37 months.

The data of both SC and EC are consistent with the description of category III proposed by Macken & Barton (1980). In category III the VOTs of voiceless aspirated and unaspirated stops produced by children are similar to adult norms. In category IIIA, the VOT mean of aspirated stops are considerably longer. In category IIIB the means of aspirated stops are shorten back towards adult means. The same phenomenon was also observed here. Both SC and EC hyperaspirated their aspirated stops before the VOT was shortened again. It is hypothesized that hyperaspiration might be the general strategy that children use when they try to develop a distinction between voiceless aspirated and unaspirated stops.

It was noted that EC and SC produced some unaspirated stops with lead VOTs. This was also observed in English data. It was found that some English initial /b, d, g/ [b̥, d̥, g̊] tokens were produced with lead VOTs (Macken & Barton, 1979). In English, however, this is compatible with the adult target; although we do not understand the pattern of variability completely, it is clear that both voiced [b, d, g] and voiceless unaspirated [p, t, k] (or /b, d, g/) are possible variants of the English stop type (Lisker & Abramson, 1964). In Taiwanese, by contrast [b, d, g] are not acceptable productions of the voiceless unaspirated stops [p, t, k], since these two series contrast with each other.

There was another phenomenon that was noted not only in SC and EC's data, but also in acquisitional studies of English (Macken & Barton, 1979) and Cantonese data (Clumeck et al., 1981). Whenever aspirated stops were hyperaspirated, the VOTs of unaspirated stops also increased.

One thing noted from EC's data, and not reported in other languages was that word familiarity may influence a child's production of that word. Before using the word /be˥˧/ 'horse', which is a word that EC was familiar with, no tokens of /b/ were elicited. As soon as the item was used, many tokens of /b/ were elicited. Since she did not refuse to say the word any more. Also when

hyperaspiration started, familiar words were first hyperaspirated before unfamiliar words. However, Macken (1980) did suggest that children's acquisition of voicing contrast appears first in the place of articulation that was used most in the children's lexicon. In other words, if children are more familiar with words starting with /b/, they learn the voicing contrast first for the bilabial sounds. In the current study, the more familiar place of articulation is confounded with the more familiar words. In any case the two studies agree that familiarity could influence children's production and acquisition of a sound. It was proposed in Gerken's (1994) speech production model that there is a limited pool of resources that the speaker can assign at each level of representation, such as the syntax, morphology, phonology, and phonetic levels. An utterance that requires more resources at one level will have less resources remaining at other levels. Therefore when asked to imitate after an unfamiliar word, there will be more resources assigned in order to perceive the word correctly, and to sort out the necessary motor controls to produce the word. Thus, less resources remain for producing the unfamiliar word.

Generally speaking, as proposed by Macken (1980) children learned to produce voiceless aspirated stops by hyperaspirating the aspirated stops. After the distinction between voiceless aspirated and unaspirated stops is established, the VOT of aspirated stops is shortened to be within the adult norm. Then children learn to produce voiced stops with lead VOT. Even though a child has acquired the skill to prevoice stops, it takes a long time for the production to become completely adult like. Gandour et al. (1986) found that even five-year-olds do not produce voiced stops in an adult like manner.

In sum, the data presented here show patterns that are similar to acquisition patterns observed in other languages. As in other languages with voiceless unaspirated stops, Taiwanese voiceless unaspirated stops are the first stops to be acquired by children. Just as English-learning and Cantonese-learning children do, when Taiwanese children first try to distinguish voiceless aspirated stops from unaspirated stops, they hyperaspirate voiceless aspirated stops. As in the Clumeck et al.'s (1981) Cantonese and Macken & Barton's (1979) English data, when the VOTs of these Taiwanese aspirated stops are lengthened, the VOTs of unaspirated stops also increase. As in Gandour et al's (1986) Thai data, Taiwanese voiced stops are the last to be acquired by children.

According to Gandour et al. (1986), Thai-learning five-year-olds still do not produce voiced stops in a completely adult-like manner. In order to trace the process of how children's production of voiced stops gradually became adult-like, a cross-sectional study sampling from a large number of children from various ages is necessary.

**References**

Allen, G. D. (1985). "How the young French child avoids the pre-voicing Problem for word-initial voiced stops," *Journal of Child Language*. **12**, 37--46.

Clumeck, H., Barton C., Macken M., & Huntington D. (1981) "The aspiration contrast in Cantonese word-initial stops: Data from children and adults," *Journal of Chinese Linguistics* 9, 210-224.

Davis, K. (in press). "Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English," *Journal of child Language.*

Gandour, L., Petty, S. H., Dardarananda, R., Dechongkit, L., & Mukngoen S., (1986) "The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants," *Journal of child Language*, **13**, 561-572.

Gerken (1994) "Sentential processes in early child language: Evidence from the perception and production of function mophemes," in J.C. Goodman & H.C. Nussbaum (eds.), *The development of speech perception,* Cambridge, MA: The MIT Press, 271-338.

Lisker, L., & Abramson A. S. (1964) "A cross-language study of voicing in initial stops: acoustical measurements," *Word,* 20, 384-422.

Macken, M. A., & Barton D. (1979) "The acquisition of the voicing contrast in English: A study of voiced onset time in word-initial stop consonants," *Journal of Child Language,* 7, 41-74.

Macken, M. A., & Barton D. (1980) "The acquisition of the voicing contrast in Spanish: Phonetic an phonological study of word initial stop consonants," *Journal of Child Language,* **7**, 433-458.

Macken. M. A. (1980). "Aspects of the acquisition of stop systems: A cross-linguistic perspective," in G.H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (eds), *Child phonology,* **Vol.1**, 143-167.

Srivastava, G. P. (1974) "A child's acquisition of Hindi consonants," *India Linguistics*, **35,** 112-118.

*OSU Working Papers in Linguistics* 45 (160-165)

# Perceptual Evidence of Tonal Coarticulation

**Shu-hui Peng**
peng@ling.ohio-state.edu

**Abstract:** The current study examines the extent to which tonal coarticulation is perceptible. Ten tokens each of two Taiwanese syllables (high tone and mid-falling tone) were placed in three contexts in which the following tone was high, mid, or mid-falling. The task of the listeners was to identify the whole phrase in a three-way forced choice after listening to the first syllable. Acoustic analysis showed that the high tone was contextually more stable than the mid-falling tone. Anticipatory assimilation of F0 was found between the mid-falling tone and its following tone. The results of the identification test indicated that contextual variability was perceptually detectable by listeners and contributed to the recognition of the following tone. Consistent with the patterns of assimilation, the following tone was more predictable when the phrase started with a mid-falling tone than when it started with a high tone. As is the case with segmental coarticulation, tonal coarticulation changes phonetic features of neighboring tones which contribute to the recognition of the tone. However, the variability of the tonal features are constrained by the phonological system of the language to maintain phonological contrast.

## Introduction

By representing assimilation as feature-spreading, non-linear phonologies recognize that these processes are rooted in patterns of segmental coarticulation attested in every language that has been studied instrumentally (e.g. Boyce, 1990). Production studies of languages such as Vietnamese (Han & Kim, 1974), Mandarin (Shih, 1988; Shen, 1990), Yoruba (Laniran, 1992), and Taiwanese (Lin, 1988; Peng, 1994) suggest that tonal coarticulation, like segmental coarticulation, is universal. Previous studies showed that tonal assimilation can be anticipatory or perseveratory in terms of direction of feature spreading. Anticipatory coarticulation refers to the phenomenon that tonal features of one tone assimilate to features of the following tone. Perseveratory assimilation is the carry-over effect of one tone on its following tone. However, previous studies also indicated that the effects of tonal coarticulation are sometimes dissimilatory. Tonal features of tones are changed to be less similar to each other in tone sequences.

The present study extends a production study of tonal coarticulation in Taiwanese (Peng, 1994). Taiwanese has seven distinctive tones: high, mid, high-falling, mid-falling, mid-rising, and two entering tones. All of the Taiwanese tones, except the entering tones, were examined in different tonal contexts in the previous production study. The productions of native Taiwanese speakers showed that anticipatory tonal assimilation occurred between contour tones (high-falling tone, mid-falling tone and mid-rising tone) and the following tone. For example, the F0 offset of the high-falling tone was higher in the context of high tone than in the context of mid-falling tone. Furthermore, in some cases, the pitch range of the target tone shifted according to the pitch height of the following tone. The coarticulation patterns between level tones (high tone and mid tone) and the following tone were dissimilatory in the production of some subjects. The pitch contour of the high tone and the mid tone ended at a higher F0 value when followed by a low-onset tone, e.g. mid-falling tone, than when followed by a high-onset tone, e.g. high tone.

The current study examines the extent to which tonal coarticulation is perceptible. That is, can a tone be recognized from its influence on the F0 of a preceding syllable excised from context in the way that a following /i/ vs. /u/ can be recognized in the spectrum of an /s/ or /ʃ/ excised from a CV syllable (Yeni-Komshian & Soli, 1981)? Taiwanese underlying high-falling and mid tones in the context of high, mid and mid-falling tones were investigated in production and perception. Because of tone sandhi, these two target syllables in context are actually high tone and mid-falling tone. Acoustic analysis was done to show pitch contours of the target syllables. The target syllables were excised from the phrases to be used as stimuli in the perception test. The task of the listeners in the perception test was to identify the whole phrase in a three-way forced choice after listening to the target syllable.

## Method

### *Materials*

The six two-syllable phrases used in the experiment are shown in the Appendix. The first syllable was the target syllable which was either the high-falling (51) or mid (33) tone underlyingly, changed to the high tone (55) and mid-falling tone (21) respectively after tone sandhi. The target syllable was followed by high tone, mid tone or mid-falling tone. All the phrases were read ten times. There were sixty tokens in total. The first syllables were excised from the phrases to be used as stimuli in the identification test.

### *Subjects*

All phrases were read by one female native Taiwanese speaker. Nineteen native Taiwanese speakers participated in the perception test (11 female, 8 male). None of them reported having speech or hearing problems. All the subjects also speak Mandarin and English.

### *Recording & Acoustic Measurements*

Tokens were recorded using a TEAC V-427C stereo cassette deck in a sound-proof booth at the Ohio State University Linguistics Laboratory. The subject was asked to keep a steady speech rate and equal loudness in reading the phrases.

Duration and fundamental frequencies of the target syllables were measured. Duration was measured from wide-band spectrograms on a Kay DSP Sona-Graph 5500. The duration of each target syllable was measured from the release of the initial stop consonant to the offset of the diphthong. The fundamental frequency of each phrase was analyzed with the *Waves* ™ signal editor (Entropic Research Laboratory, INC., 1993). The F0 of each target syllable was measured at three different points along the F0 contour: onset, midpoint and offset.

*Procedure*

The identification test consisted of two sessions: high tone and mid-falling tone. Before each session, subjects were asked to read the three two-syllable phrases starting with the target syllable. There was a practice session before each test session. The practice session contained nine trials selected from all following tone conditions including high tone, mid tone and mid-falling tone. All the first syllables were presented to subjects four times in random order over headphones. There were 120 trials in each test session. Subjects heard the target tone at the beginning of each trial and could listen to it as many times as necessary by clicking on a button shown on the computer screen in front of them. There were also three other buttons labeled with the three phrases, Chinese characters. The task of the subject was to identify the phrase in a three-way forced choice after listening to the first syllable.

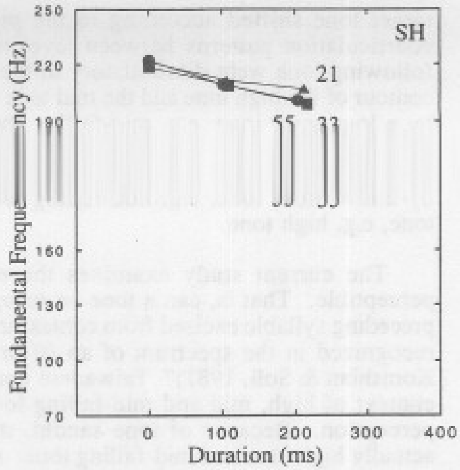



**Fig. 1.** Mean durations and F0s of the high target tone followed by different tones.

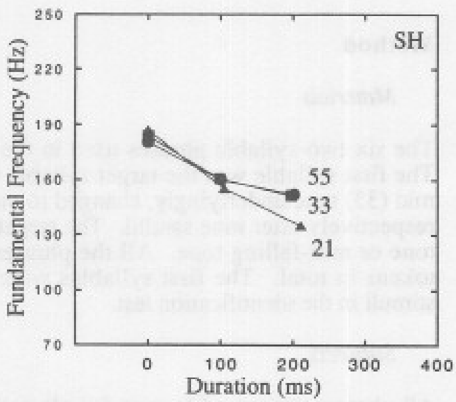



**Fig. 2.** Mean durations and F0s of the mid-falling tone followed by different tones.

## Results

*Production data*

Acoustic analysis showed similar patterns of assimilation to those found in previous studies (Lin, 1988; Peng, 1994). Pitch heights of tones were affected by tonal environments. Measurements of duration and F0 indicated that the high tone was contextually more steady than the mid-falling tone in the productions of this speaker. Figure 1 shows mean durations and F0 values of the high tone in different tonal environments. Three different F0 measures are shown: onset, mid-point and offset of syllable. The effects of the following tone on the duration and F0 of the high tone were not significant. Although the F0 contour of the high tone ends at a slightly higher F0 value in the environment of mid-falling tone than in the other two tonal environments, the difference was not statistically significant.

Mean duration and F0 values of the mid-falling tone are shown in Figure 2. The effect of the following tone on duration of the mid-falling tone was marginally significant, [$F(2, 27) = 4.929$, $p < 0.05$]. The duration of the mid-falling tone was slightly longer when followed by a mid-falling tone than when followed by a mid tone. Tonal context had a significant effect on the F0 of the mid-falling tone at the offset of the F0 contour, [$F(2, 27) = 22.444$, $p < 0.001$]. The F0 value of the mid-falling tone was lower when followed by the mid-falling tone than when followed by the other two tones.

*Identification data*

The results of identification tests for the high tone and the mid-falling tone are shown in Figures 3 and 4. I values derived from the lower bound of the Green's area measure (Green, 1964) represent the measurement of signal detectability. The sensitivity of subjects to the signals they heard ranged from zero to one. A value of one refers to perfect sensitivity or performance. For example, if subjects identify a phrase with complete accuracy and without any false alarm response, then the I value for the phrase is one.

Figure 3 shows the I values with standard error for the high tone in different environments of the following tone. The effect of response type was significant, [$F(2, 54) = 22.801$, $p < 0.001$]. Tukey post-hoc tests showed that the I value for the high tone in the context of the mid tone was higher than in the context of the high tone which was in turn higher than in the context of the mid-falling tone. However, t-tests indicated that the performance of subjects was better than chance ($I = 0.33$) only in identifying the phrase with high tone followed by mid tone, [$F(1, 18) = 2.397$, $p < 0.05$]. Figure 4 shows the I values of the mid-falling tone in different tonal environments. The effect of response type was significant, [$F(2, 54) = 8.736$, $p < 0.01$]. The performance of subjects was better in the environment of the mid-falling tone than in the other two environments. t-tests indicated that performance for all response types was better than chance, [high tone $F(1, 18) = 5.175$, $p < 0.001$; mid tone $F(1, 18) = 3.669$, $p < 0.01$; mid-falling tone $F(1,18) = 2.528$, $p < 0.05$].

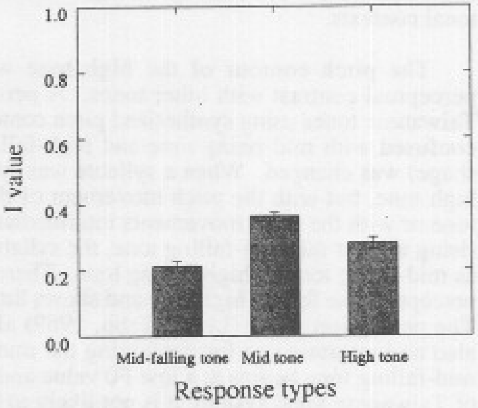



**Fig. 3.** Mean I values of different response types for the high tone.

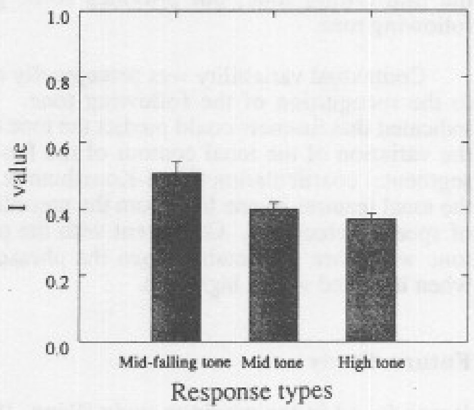



**Fig. 4.** Mean I values of different response types for the mid-falling tone.

## Discussion

Just as segmental coarticulation does, tonal coarticulation changes phonetic features of neighboring tones which contribute to the recognition of the tone. However, the variability of the tonal features are constrained by the phonological system of the language, to maintain phonological contrast. The mid-falling tone showed more acoustic variation according to tonal context than the high tone. Duration and F0 values of the mid-falling tone changed according to the tonal quality of the following syllable. The mid-falling tone was longer in duration and lower in F0 at the offset of the F0 contour when preceding mid-falling tone than when preceding the high tone and the mid tone. The effect was stronger on F0 than on duration. Anticipatory assimilation of F0 was found between the mid-falling tone and its following tone. The F0 height of the mid-falling tone at the offset of the F0 contour assimilated to the onset F0 height of its following tone. The high tone, on the contrary, did not show much variation in both duration and F0 values in different tonal contexts.

The pitch contour of the high tone was contextually stable to maintain perceptual contrast with other tones. A perception study (Lin & Repp, 1989) of Taiwanese tones using synthesized pitch contours showed that high tone was easily confused with mid-rising tone and high-falling tone if its pitch movement (tone shape) was changed. When a syllable was synthesized with the pitch height of the high tone, but with the pitch movement of the mid-rising tone or the high-falling tone or with the pitch movements intermediate between the high tone and the mid-rising tone or the high-falling tone, the syllable tended to be perceived by listeners as mid-rising tone or high-falling tone. Therefore, pitch movement is an important perceptual cue for the high tone and shows little variation in different tonal contexts. The perception study (Lin & Repp, 1989) also showed that pitch movement was also an important cue for perceiving the mid-falling tone. However, because the mid-falling tone begins at a low F0 value and ends at the bottom of the pitch range of Taiwanese tonal system, it is not likely to be perceived as other tones unless the tone shape is drastically changed to be flat as the mid tone or rising as the mid-rising tone, or the F0 onset is raised as high as the high tone or high-falling tone. The change of F0 offset due to tonal assimilation does not affect the tonal quality of the mid-falling tone, but provides some perceptual cues for recognizing the following tone.

Contextual variability was perceptually detectable by listeners and contributed to the recognition of the following tone. The results of the identification test indicated that listeners could predict the tone of the second syllable in a phrase from the variation of the tonal contour of the first syllable. As found in perception of segmental coarticulation (Yeni-Komshian & Soli, 1981), listeners could anticipate the tonal features of one tone from the preceding tone possibly to increase efficiency of speech perception. Consistent with the patterns of assimilation, the following tone was more predictable when the phrase started with a mid-falling tone than when it started with a high tone.

## Future Study

It was found in the previous study (Peng, 1994) that tonal dissimilation occurred between the high tone or the mid tone and the following tone. One may wonder whether F0 cues derived from tonal dissimilation will contribute to the perception of the high tone or the mid tone as found in tonal assimilation. A perception study using synthesized pitch contours will indicate the extent to which tonal dissimilation is perceptually detectable. The synthesized pitch contour of the target tone will have a lower F0 offset when followed by a high-onset tone than when followed by a low-onset tone. This perception study will show whether listeners can predict one tone from the F0 cues in the preceding tone generated by tonal dissimilation.

## References


Boyce, S. (1990) Coarticulatory organization for lip rounding in Turkish and English. *Journal of Acoustical Society of America*, 88 (6), 2584-2595.

Entropic Research Laboratory, INC. (1993) *Waves+ 5.0*. AT&T Bell Laboratories. Washington, D. C..

Green, D. M. (1964) General prediction relating yes-no and forced choice results. *Journal of Acoustical Society of America*, 36, 1042.

Han, M. & Kim, K.-O. (1974) Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, 2, 223-232.

Laniran, Y. (1992) *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yoruba*. PhD Dissertation, Cornell University.

Lin, H.-B. (1988) *Contextual Stability of Taiwanese Tones*. PhD Dissertation, University of Connecticut.

Lin, H.-B. & Repp, B. (1989) Cues to the perception of Taiwanese tones. *Language and Speech*, 32(1), 25-44.

Peng, S.-H. (1994) Effects of Prosodic Position and Tonal Context on Taiwanese Tones. *OSU Working Papers in Linguistics*, 44, 166-190.

Shen, X. (1990) Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18, 281-295.

Shih, C. (1988) Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.

Yeni-Komshian, G. & Soli, S. (1981) Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of Acoustical Society of America*, 70 (4), 966-975.


## Appendix

/kau$^{51}$ taŋ$^{55}$/ --> [kau$^{55}$ taŋ$^{55}$] nine years
/kau$^{51}$ tai$^{33}$/ --> [kau$^{55}$ tai$^{33}$] nine generations
/kau$^{51}$ tã$^{21}$/ --> [kau$^{55}$ tã$^{21}$] nine burdens

/kau$^{33}$ kaŋ$^{55}$/ --> [kau$^{21}$ kaŋ$^{55}$] time-consuming
/kau$^{33}$ kau$^{33}$/ --> [kau$^{21}$ kau$^{33}$] thick
/kau$^{33}$ kua$^{21}$/ --> [kau$^{21}$ kua$^{21}$] thick lid

*OSU Working Papers in Linguistics* 45 (166-169)

# Temporal Structure of an Estonian Lament: A Case Study*

**Jaan Ross**
jankel@pea.kki.ee
**Ilse Lehiste**
ilsele@ling.ohio-state.edu

**Abstract:** The acoustic structure of a South-Eastern Estonian lament was analyzed. The lament was recorded in 1972 as performed alternatingly by a soloist and a choir (of several female voices). In this lament, the basic structural unit consists of two four-foot verse lines of alternating trochaeic and dactylic feet (2 + 3 + 2 + 3 syllables). The first verse of the couplet has invariant verbal content throughout the whole lament. The two-line unit is first performed by the soloist and, at the end of the second line, taken over and repeated by the choir. There are 11 structural units in the whole lament. Due to an almost complete one-to-one correspondence between syllables in the text and notes in the melody, the overall number of notes in the lament is

$$n = (2 + 3 + 2 + 3) \times 2 \times 2 \times 11 = 440$$

reduced by occasional pauses where the performer has to take a breath. Durations of all notes are perceived as equal by a listener, the exception being the very end of each structural unit where the last two notes are sung at a faster speed. The possible influence of a number of factors on syllable/note durations was tested by an ANOVA. These factors included word stress, metrical accent, modus of performance (solo or ensemble), variability of verbal contents (first or second line of the structural unit), position in the verse line, and position of the structural unit in the whole lament. The results were significant for performance modus, syllable/note position, and (somewhat less) for word stress. A closer study of differences between the performances of the soloist and the choir reveals a marked ritenuto at the end of the second verse line sung by the soloist, probably aimed at signaling to the choir that it has to step in.

## Introduction

Lamenting is a way of passionate expression of grief which may accompany fundamental changes in human life, such as death or marriage. Lamenting is a pre-Christian tradition and some of its well-known examples have been recorded from Papua New Guinea and Hong Kong. Remnants of lamenting, however, are still preserved in remote Eastern European regions such as Karelia or parts of southeastern Estonia, the territory known as Setu.

* This study was made possible through a Fulbright scholarship granted to the first author in 1992.

Some ethnomusicologists consider laments as possessing special functions and performance contexts, characterizing them as not quite speech and not quite singing. At the same time, lament tradition does not seem to be very homogeneous. Care should be taken in order not to overlook differences which may occur between lament traditions even when they have been observed in regions geographically close to each other. While it is generally agreed that Karelian laments are rather distant from orally performed folksongs, even demonstrating certain signs of sacred language which may not be understood by listeners (Tolbert 1990), the Setu laments, on the contrary, are close to the song tradition and could be considered as a part of it (Salve 1993). Like Estonian folksongs, a Setu lament uses elements from the common pool of the pre-existent repertoire, and combines these with an improvisational part which links the performance to the specific occasion.

The current paper studies the temporal structure of a Setu funeral lament which was recorded in 1972 under semi-field conditions, i.e. in a natural environment upon request by the interviewer, but not at the actual funeral. As an approximation, this lament may be considered as an isochronous sequence of temporal events. A tendency toward neutralization of quantity oppositions during lamenting has been described by Ross and Lehiste (1994). For the purposes of this paper, a temporal event may be defined either as a syllable or as a note (or tone), since there is one-to-one correspondence between syllables and tones in the lament. Below we will focus on deviations from the quasi-isochronous sequence of syllables/tones which occur during the performance and will try to understand the reasons for these deviations.

## Material

The analyzed lament was performed alternately by a soloist and a chorus of a few voices, all female. The basic structural unit of the lament consists of two four-foot verse lines of alternating trochaeic and dactylic feet (2 + 3 + 2 + 3 syllables). The first verse line of the couplet has invariant verbal content throughout the whole lament (it constitutes an address formula directed at the deceased), while the second line develops the "story" to be told. The two-line unit is first performed by the soloist and, at the end of the second line, taken over and repeated by the choir. There are eleven structural units in the whole lament. The predicted overall number of syllables (or tones) in the lament is

$$n = (2 + 3 + 2 + 3) \times 2 \times 2 \times 11 = 440$$

The actual number of syllables/tones measured was equal to 410. It was somewhat less than predicted because of replacement of some syllables/tones by pauses when the performer had to take a breath, and in some cases due to technical reasons.

## Methods

The duration of syllables (tones) was measured using a Kay Elemetrics Signal Analysis Workstation (Model 5500), with a sampling frequency of 20 kHz. Segmentation of the sound signal was performed using two time-synchronized spectrographic representations (with bandwidths of 59 and 400 Hz respectively) of successive four-second portions of sound. The time resolution was estimated not to exceed 20 msec.

## Results and Discussion

The distribution of syllable/note durations in the lament is shown in Figure 1.

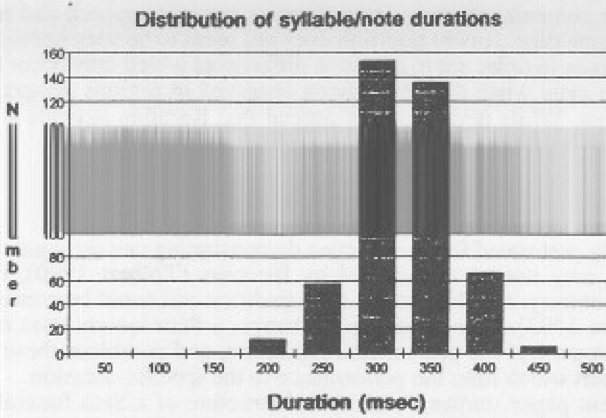



Figure 1. Distribution of syllable/note durations in the analyzed lament.

The average syllable/tone duration in this lament is about 300 msec. One-way analysis of variance was performed where the influence on syllable/tone durations of the following six variables was studied:

-position within a line (of a total of ten positions)
-position of a two-line structural unit within the lament (of a total of eleven)
-position of a line within the structural unit (first or second)
-relative metrical accent strength (strong or weak)
-performance type (solo or ensemble)
-presence or absence of word stress on the syllable.

Metrically strong positions are numbers 1, 3, 6, and 8 in a verse line, as the first syllables of trochaeic or dactylic verse feet.

Significant effects on syllable/tone durations were found by its position within the line ($F(9,400) = 2.56$, $p < .01$), by performance either solo or by choir ($F(1,408) = 12.8$, $p < .001$), and by word stress ($F(1,408) = 4.49$, $p < .05$).

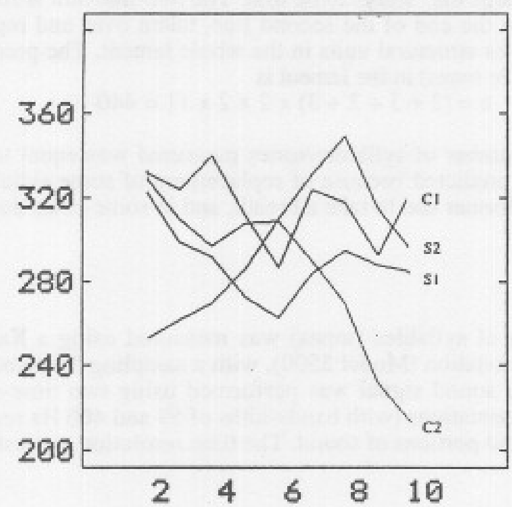



Figure 2. Tempo curves for four different verse line types. X-axis: syllable/tone position within a line; y-axis: smoothed average, in msec, of syllable/tone duration in given position.

Figure 2 presents tempo curves for four different verse line types in the lament: the line with invariant text performed solo (S1), the line with changing words performed solo (S2), the line with invariant text performed by ensemble (C1), and the line with changing words performed by ensemble (C2). Syllable/tone durations were averaged over 11 structural units (consisting of S1 + S2 + C1 + C2) as well as over two successive syllables/tones (1 + 2, 2 + 3, 3 + 4 etc.); the xaxis thus presents smoothed averages of syllable/tone durations in given positions within the line. Visual inspection shows that the tempo curve for S2 (the soloist's second line) differs from those of other lines, demonstrating monotonous increase of syllable/tone durations (reduction in tempo) until the 7th or 8th position in the line. This suggests that the most powerful reason for deviating from the isochronous syllable/tone sequence of the lament is a rather pragmatic one: to signal to the choir that it is their turn to join in, which happens before the last metric foot of the soloist's line. This appears also to be the reason for the significant effect on syllable/tone duration of its position within the line and of performance by either solo or choir.

This conclusion is strengthened by a two-way analysis of variance, which shows interaction of syllable/tone position both with solo/ensemble performance ($F(9,390) = 4.76$, $p < 6.82$, $p < .0001$), as well as interaction between the two latter variables, solo/ensemble and variant/invariant words ($F(1,406) = 5.90$, $p < .05$). Moreover, it has been noticed that lament performers tend to use all available means, including bodily movements, in order to make explicit the moment when other performers should join the performance (Sarv 1994).

As regards the role of word accent, found to be significant at the .05 level, the outcome is not as clear. All Estonian words are accented on the first syllable. The first syllable of polysyllabic words participates in the three-way quantity opposition which affects both the duration of the accented syllable and the duration of the second syllable. The durational ratios between the first and second syllable are approximately 2/3 for the short quantity, 3/2 for the long quantity, and 2/1 for the overlong quantity (Lehiste 1968). The ANOVA compared accented word-initial syllables with all other syllables in the lament. This did not take into consideration the occurrence of secondary stress on the third or fourth syllable of four- or five-syllable words, nor the fact that monosyllabic words do not participate in the quantity opposition and may or may not carry sentence-level stress. A preliminary calculation of the average duration of accented first syllables of polysyllabic words ($N = 108$) gave a value of 296 msec (s.d. 54.66), and of unstressed syllables in polysyllabic words ($N = 248$), 312 msec (s.d. 140.18). In the given lament, the ratio between the two syllables is 0.95, which is closest to the ratio found in words of Q1, the short quantity (2/3, or 0.67). Neutralization of quantity oppositions in laments is discussed extensively in Ross & Lehiste 1994. In the current case, it appears that at least partial neutralization in the direction of Q1 has taken place.

## References


Lehiste, I. (1968). *Vowel quantity in word and utterance in Estonian*. Congressus Secundus Internationalis FennoUgristarum. Helsinki: Societas Finno-Ugrica. Pp. 293-303.

Ross, J., and Lehiste, I.(1994). *Lost prosodic oppositions: A study of contrastive duration in Estonian funeral laments*. To appear in Language and Speech.

Salve, K. (1993). *Tâhelepanekuid setu surnuitkude zanridevahelistest ja geograafilistest seostest*. Akadeemia 5, 2605-2621.

Sarv, V. (1994). Personal communication.

Tolbert, E. (1990). *Women cry with words: Symbolization of affect in the Karelian lament*. Yearbook for Traditional Music, 22, 80-122.

Figure 3 presents motion curves for four different verse line types in the manner: the line with tenseline was performed solo (X1), the line with changing words, performed solo (X2), the line with tenseline was performed by ensemble (C1), and the line with changing words performed by ensemble (C2). Syllable/tone durations were measured over 13 utterances using ... of $S1 + S2 + C1 + C2$.



They present smoothed averages of syllable/tone durations in given positions within the line. Visual inspection shows that the tempo curve for X2 (the soloist's second line) differs from those of other lines, demonstrating a conspicuous increase of syllable/tone durations (reduction in tempo) until the 7th or 8th position in the line. This suggests that the most powerful reason for deviating from the isochronous syllable/tone sequence of the lament is a rather pragmatic one: to signal to the other choir singers their turn to join in, which happens before the last metric foot of the soloist's line. This appears also to be the reason for the significant effect on syllable/tone durations of its position within the line and of performance by either solo or choir.

This conclusion is strengthened by a two-way analysis of variance, which shows interaction of syllable/tone position with a fit solo/ensemble performance ($F(9,250) = 4.79$, $p < 0.02$, $p < .0001$), as well as interaction between the two latter variables, soloist/ensemble and verse/changing words ($F(1,250) = 5.90$, $p < .01$). Moreover, it has been noticed that lament performers tend to use all available means, including bodily movements, in order to make explicit the moment when other performers should join the performance (Salo 1991).

As regards the role of word accent, found to be significant at the .05 level, the outcome is just as clear. All Estonian words are accented on the first syllable. The first syllable of polysyllabic words participates in the three-way quantity opposition which affects both the duration of the accented syllable and the duration of the second syllable. The durational ratios between the first and second syllable are approximately 2/3 for the short quantity, 3/2 for the long quantity, and 2/1 for the overlong quantity (Lehiste 1960). The ANOVA compared accented word-initial syllables with all other syllables in the lament. This did not take into consideration the occurrence of secondary stress on the third or fourth syllable of four- or five-syllable words, nor the fact that monosyllabic words do not participate in the quantity opposition and they may not carry sentence-level stress. A preliminary calculation of the average duration of accented first syllables of polysyllabic words ($N = 108$) gave a value of 296 msec (s.d. 94.66), and of unaccented syllables in polysyllabic words ($N = 245$), 312 msec (s.d. 120.24). In the given laments, the ratio between the two syllables is 0.95, which is closest to the ratio found in words of Q1, the short quantity (2/3, or 0.67). Neutralization of quantity oppositions in laments is discussed extensively in Ross & Lehiste 1994. In the current case, it appears that at least partial neutralization in the direction of Q1 has taken place.

## References


Lehiste, I. (1960). Vowel quantity in word and utterance in Estonian. Congressus Secundus Internationalis Fenno-Ugristarum. Helsinki: Societas Fenno-Ugrica. Pp. 293–303.

Ross, J. and Lehiste, I. (1994). Lost prosodic oppositions: A study of contrastive duration in Estonian funeral laments. To appear in Language and Speech.

Salve, K. (1991). Eestirootsikeel and ... [illegible] ... Academica 5. Tallinn.

Sarv, V. (1993). Personal communication.

Tolbert, E. (1990). Women cry with words: Symbolization of affect in the Karelian lament. Yearbook for Traditional Music, 22, 80–105.