sponse times, CLARIFYUND the longest, and CLARIFYPERC between these. The differences are statistically significant (one-way within-subjects ANOVA; F=7.558; dF=2; p<0.05).

4 Conclusions and discussion

In the present study, we have shown that users of spoken dialog systems not only perceive the differences in prosody of synthesized fragmentary grounding utterances, and their associated pragmatic meaning, but that they also change their behavior accordingly in a human-computer dialog setting. The results show that two annotators were able to categorize the subjects' responses based on pragmatic meaning. Moreover, the subjects' response times differed significantly, depending on the prosodic features of the grounding fragment spoken by the system.

The response time differences found in the data are consistent with a cognitive load perspective that could be applied to the fragment meanings ACCEPT, CLARIFYPERC and CLARIFYUND. To simply acknowledge an acceptance should be the easiest, and it should be nearly as easy, but not quite, for users to confirm what they have actually said. It should take more time to reevaluate a decision and insist on the truth value of the utterance after CLARIFIYUND. This relationship is nicely reflected in the data.

Although we have not quantified other prosodic differences in the users' responses, the annotators felt that there were subtle differences in e.g. pitch range and intensity which may function as signals of certainty following CLARIFYPERC and signals of insistence or uncertainty following CLARIFYUND. More neutral, unmarked prosody seemed to follow ACCEPT. When listening to the resulting dialogs as a whole, the impression is that of a natural dialog flow with appropriate timing of responses, feedback and turntaking. To be able to create spoken dialog systems capable of this kind of dialog flow, we must be able to both produce and recognize fragmentary grounding utterances and their responses. Further work using more complex fragments and more work on analyzing the prosody of user responses is needed.

References

Clark, H.H., 1996. *Using language.* Cambridge: Cambridge University Press.
Edlund, J. & M. Heldner, 2005. Exploring Prosody in Interaction Control. *Phonetica 62(2-4),* 215-226.
Edlund, J., D. House & G. Skantze, 2005. The effects of prosodic features on the interpretation of clarification ellipses. *Proceedings of Interspeech 2005,* Lisbon, 2389-2392.
Filipsson, M. & G. Bruce, 1997. LUKAS – a preliminary report on a new Swedish speech synthesis. *Working Papers 46,* Department of Linguistics and Phonetics, Lund University.
Wallers, Å., J. Edlund & G. Skantze, 2006. The effect of prosodic features on the interpretation of synthesised backchannels. *Proceedings of Perception and Interactive Technologies,* Kloster Irsee, Germany.

The Prosody of Public Speech – A Description of a Project

Eva Strangert[1] and Thierry Deschamps[2]

[1]Department of Comparative Literature and Scandinavian Languages, Umeå University
eva.strangert@nord.umu.se
[2]Department of Philosophy and Linguistics, Umeå University
thierry.deschamps@ling.umu.se

Abstract
The project concerns prosodic aspects of public speech. A specific goal is to characterize skilled speakers. To that end, acoustic analyses will be combined with subjective ratings of speaker characteristics. The project has a bearing on how speech, and prosody in particular, can be adjusted to the communicative situation, especially by speakers in possession of a rich expressive repertoire.

1 Introduction

This paper presents a new project, the purpose of which is to identify prosodic features which characterize public speech, both read and spontaneous. The purpose is moreover to reveal how skilled public speakers use prosody to catch and keep the attention of their listeners, whether it be to inform or argue with them. Combined with acoustic analyses of prosody, subjective ratings of speakers will contribute to our knowledge of what characterizes a "good" or "skilled" speaker. Thus, the project, though basically in the area of phonetics, has an interdisciplinary character as it also addresses rhetoric issues.

The idea of approaching public speech has grown out of previous work in the field of prosody including the recently completed project "Boundaries and groupings – the structuring of speech in different communicative situations", see Carlson et al. (2002) as well as studies dealing specifically with the prosody of public speech, see below. Additional motivation for the new project is the growing interest today in public speech, and rhetoric in particular.

The project should also be seen in the perspective of the significance given to the areas of speaking style variation and expressive speech during the last decades. This research is theoretically important, as it increases our knowledge of how human speech can be optimally adjusted to the specific situation, and it contributes to learning about the limits of human communicative capacity. Public speech offers a possibility to study speech that can be seen as extreme in this respect. In politics and elsewhere when burning issues are at stake and where often seriously committed individuals are involved, a rich expressive repertoire is made use of. In this domain, prosody has a major role.

2 Background

Common to textbooks in rhetoric is their focus on those aspects which do not concern the manner of speaking, although it is included in the concept of "rhetoric". The emphasis is rather on argumentation and planning of the speech act, the rhetoric process, as well as the linguistic form; correctness, refinement, and clarity are demanded. The descriptions of how to speak are considerably less detailed and very often even vague. The recommendations of

today are mostly similar to those given two thousand years ago; the voice of a skilled speaker should be "smooth", "flexible", "firm", "soft", "clear" and "clean" (Johannesson, 1990/1998, citing Quintilianus' (ca AD 35-96) "Institutes of Oratory").

As far as phonetically based investigations are concerned, Touati (1991) analyzed tonal and temporal characteristics in the speech of French politicians. The analyses were undertaken with a background in earlier studies of political rhetoric and, in addition, other types of speech in public media in Sweden, see Bruce & Touati (1992). Other studies of public speech based on Swedish include Strangert (1991; 1993) both dealing with professional news reading. A study by Horne et al. (1995) concerned pausing and final lengthening in broadcasts on stockmarket reports on Swedish Radio.

Analyses of interview speech made within the "Boundaries and groupings" project also have relevance here. The purpose in this case was not to study public speech per se. However, the results, in particular as concerns fluency and pausing (see e.g. Heldner & Megyesi, 2003; Strangert 2004; Strangert & Carlson, 2006), may be assumed to reflect the fact that the speech was produced by a very experienced speaker. A recent study with focus on "the skilled professional speaker" (Strangert 2005) approaches problems sketched for the current project.

Braga & Marques (2004) focused on how prosodic features contribute to the listeners' attention and interpretation of the message in political debate. The conception of a speaker as "convincing", "powerful" and "dedicated", is assumed to be reflected in (combinations of) prosodic features, or "maximes". The study builds on the idea put forward by Gussenhoven (2002) and developed further by Hirschberg (2002) of universal codes for how prosodic information is produced by the speaker and perceived by the listener. Wichmann (2002) and Mozziconacci (2002) belong to those dealing with the relations between prosody (f0 features in particular) and what can be described as "affective functions"; a comprehensive survey of expressive speech research can be found in Mozziconacci (2002).

Wichmann (2002) makes a distinction between "ways of saying" (properties or states relating to the speaker) and "ways of behaving" (the speaker's attitude to the listener). "Ways of saying" includes first, how the speaker uses prosody in itself – stress and emphasis, tonal features, speech rate, pausing etc. – and second, the emotional coloring of speech (e.g. "happy", "sad", "angry") as well as states such as "excited", "powerful" etc. Examples of "ways of behaving" are attitudes such as "arrogant" and "pleading". In addition, the speaker may use other argumentative and rhetorical means. All these functions of prosody make it a complex, nuanced and powerful communicative tool.

To study the affective functions of prosody, auditive analyses must be combined with acoustic measurements (see e.g. Mozziconacci, 2002). Also, listeners' impressions have to be categorized appropriately. A standard procedure is to have listeners judge samples of speech. However, human speech very often conveys several states, attitudes and emotions at the same time and this without doubt is true for the often quite elaborated speech produced in the public domain. This complexity is examined in a study by Liscombe et al. (2003) through the use of multiple and continuous scales for rating emotions. In their study, the subjective ratings are also combined with acoustic analyses of prosodic features.

3 Work in progress

As a first step, we made a survey asking 22 students of logopedics at Umeå University what kind of qualities they looked upon as important for a person regarded as a "good speaker". The students wrote down as many characteristics (in Swedish) as they could, guided only by the definition of a good speaker as "A person who easily attracts listeners' attention through her/his way of speaking."

7 characteristics were given on average, with a range between 4 and 11. In addition to personality/emotional and attitudinal/interaction features, the labels given also reflected

opinions about speech per se (articulation, voice characteristics and prosody), cf. Wichman (2002). Thus, even if both the personality and the attitudinal features are transferred to the listeners through speech, the subjects did not refrain from having opinions about the speech itself. Table 1 shows the distribution of labels after grouping into the three categories.

With this as a background we will proceed by having subjects judge short passages of speech (spontaneous and prepared) for multiple speaker characteristics. These will include not only positively valued qualities like those listed here; also other qualities, including more negatively colored ones, need to be covered in an effort to characterize speaker behavior. We are currently in the process of developing a test environment for this experiment. In this work we lean on previous efforts (see Liscombe et al., 2003). Combined with acoustic analyses we expect the multiple ratings to give insight into how different acoustic/prosodic features contribute to the impression of skilled – and less skilled – speaking behavior.

Table 1. Characteristics of "a good speaker" grouped into three categories based on 22 subjects' written responses (see text). Labels in Swedish with English translations.

Speaker characteristics		Number of responses
Speech features		
tydlig artikulation	clear articulation	7
god röststyrka, röstläge	sufficient volume, voice level	6
icke-monoton röst	non-monotonous voice	4
variation i röststyrka, röstläge	variation of volume, voice level	3
rätt betoning, fokusering	adequate prominence and focus	2
väl avvägd pausering	well-adjusted pausing	2
bra taltempo, ej för snabbt	well-adjusted tempo, not too fast	2
varierat taltempo	varied speech tempo	1
talflyt	fluency	1
varierad prosodi, uttrycksfullhet	varied prosody, expressiveness	3
Personality features		
inlevelse, entusiasm, engagemang	involvement, enthusiasm, commitment	16
humor, lättsamhet	sense of humour	12
karisma, utstrålning	charisma, appeal	6
lugn, avslappnad stil	calm, relaxed style of speaking	5
personlighet	personality, individuality	4
positivt inställning	positive attitude	3
ödmjukhet, självinsikt	sense of humility	2
tydlighet	distinctness, authority	2
självförtroende	self-confidence	1
övertygelse	conviction	1
Interaction features		
förmåga att knyta an till lyssnarna	ability to relate to audience	8
nivåanpassning relativt lyssnarna	choosing the right communicative level	6
lyhördhet	sensitivity	3
vilja till interaktion	ability to interact with audience	3
utan överlägsenhet	respectful, non-arrogant style	2

As the project in addition aims to characterize also other aspects of public speaking, a variety of representative speech samples will be collected. In analyses of this material, fluency, pausing, prominence, emphasis and voice characteristics will be central. Among the questions we seek answers to are: What types of strategies are used for holding the floor? How does speech perceived as fluent and disfluent respectively differ acoustically? How are prominence and emphasis used in speech in media? What are the prosodic characteristics of agitation? Answers to these questions, we believe, will add to our understanding of human communicative capability and will also be useful in modeling speaking style variation. Knowledge gained within the project may further be expected to be practically applicable.

Acknowledgements

References
Braga, D. & M.A. Marques, 2004. The pragmatics of prosodic features in the political debate. *Proc. Speech Prosody 2004*, Nara, 321-324.
Bruce, G. & P. Touati, 1992. On the analysis of prosody in spontaneous speech with exemplification from Swedish and French. *Speech Communication 11*, 453-458.
Carlson, R., B. Granström, M. Heldner, D. House, B. Megyesi, E. Strangert & M. Swerts, 2002. Boundaries and groupings – the structuring of speech in different communicative situations: a description of the GROG project. *TMH-QPSR 44*, 65-68.
Gussenhoven, C., 2002. Intonation and interpretation: Phonetics and phonology. *Proc. Speech Prosody 2002*, Aix-en-Provence, 11-13.
Heldner, M. & B. Megyesi, 2003. Exploring the prosody-syntax interface in conversations. *Proc. ICPhS 2003*, Barcelona, 2501-2504.
Hirschberg, J., 2002. The pragmatics of intonational meaning. *Proc. Speech Prosody 2002*, Aix-en-Provence, 65-68.
Horne, M., E. Strangert & M. Heldner, 1995. Prosodic boundary strength in Swedish: final lengthening and silent interval duration. *Proc. ICPhS 1995*, Stockholm, 170-173.
Johannesson, K., 1998/1990. *Retorik eller konsten att övertyga*. Stockholm: Norstedts Förlag.
Liscombe, J., J. Venditti & J. Hirschberg, 2003. Classifying subject ratings of emotional speech using acoustic features. *Proc. Eurospeech 2003*, Geneva, 725-728.
Mozziconacci, S., 2002. Prosody and emotions. *Proc. Speech Prosody 2002*, Aix-en-Provence, 1-9.
Strangert, E., 1991. Phonetic characteristics of professional news reading. *PERILUS XII*. Institute of Linguistics, University of Stockholm, 39-42.
Strangert, E., 1993. Speaking style and pausing. *PHONUM 2*. Reports from the Department of Phonetics, University of Umeå, 121-137.
Strangert, E., 2004. Speech chunks in conversation: Syntactic and prosodic aspects. *Proc. Speech Prosody 2004*, Nara, 305-308.
Strangert, E., 2005. Prosody in public speech: analyses of a news announcement and a political interview. *Proc. Interspeech 2005*, Lisboa, 3401-3404.
Strangert, E., & R. Carlson. 2006. On modeling and synthesis of conversational speech. *Proc. Nordic Prosody IX, 2004*, Lund, 255-264.
Touati, P., 1991. Temporal profiles and tonal configurations in French political speech. *Working Papers 38*. Department of Linguistics, Lund University, 205-219.
Wichmann, A., 2002. Attitudinal intonation and the inferential process. *Proc. Speech Prosody 2002*, Aix-en-Provence, 11-22.

Effects of Age on VOT: Categorical Perception of Swedish Stops by Near-native L2 Speakers

Katrin Stölten
Centre for Research on Bilingualism, Stockholm University
Katrin.Stoelten@biling.su.se

Abstract
This study is concerned with effects of age of onset of L2 acquisition on categorical perception of the voicing contrast in Swedish word initial stops. 41 L1 Spanish early and late learners of L2 Swedish, who had carefully been screened for their 'nativelike' L2-proficiency, as well as 15 native speakers of Swedish participated in the study. Three voicing continua were created on the basis of naturally generated word pairs with /p t k b d g/ in initial position. Identification tests revealed an overall age effect on category boundary placement in the nativelike L2 speakers' perception of the three voicing continua. Only a small minority of the late L2 learners perceived the voicing contrast in a way comparable to native-speaker categorization. Findings concerning the early learners suggest that most, but far from all, early L2 speakers show nativelike behavior when their perception of the L2 is analyzed in detail.

1 Introduction
From extensive research on infant perception it has become a well-known fact that children during their first year of life tune in on the first language (L1) phonetic categories, leaving them insensitive to contrasts not existing in their native language (e.g. Werker & Tees, 1984).

In a study by Ruben (1997) it was found that children who had suffered from otitis media during their first year of life showed significantly less capacity for phonetic discrimination compared to children with normal hearing during infancy when they were tested at the age of nine years. Such findings do not only demonstrate the importance of early linguistic exposure, they have also been interpreted as an indication for the existence of a critical period for phonetic/phonological acquisition which may be over at the age of one year (Ruben, 1997).

In research of age effects on language acquisition one classical issue is concerned whether theories of a critical period can be applied to second language (L2) acquisition. The question is whether the capacity to acquire phonetic detail in L2 learning is weakened or lost due to lack of verbal input during a limited time frame for phonetic sensitivity, or whether a nativelike perception and an accent-free pronunciation is possible for any adult L2 learner.

The present study is part of an extensive project on early and late L2 learners of Swedish with Spanish as their L1. The subjects have been selected on the criterion that they are perceived by native listeners as mother-tongue speakers of Swedish in everyday oral communication. Thereafter, the candidates' nativelike L2 proficiency has been tested for various linguistic skills. The present study focuses on the analysis of the nativelike subjects' categorical perception of the voicing contrast in Swedish word initial stops.

Both Swedish and Spanish recognize a phonological distinction between voiced and voiceless stops in terms of voice onset time (VOT) but they differ as to where on the VOT continuum the stop categories separate. In contrary to languages like Swedish and English,