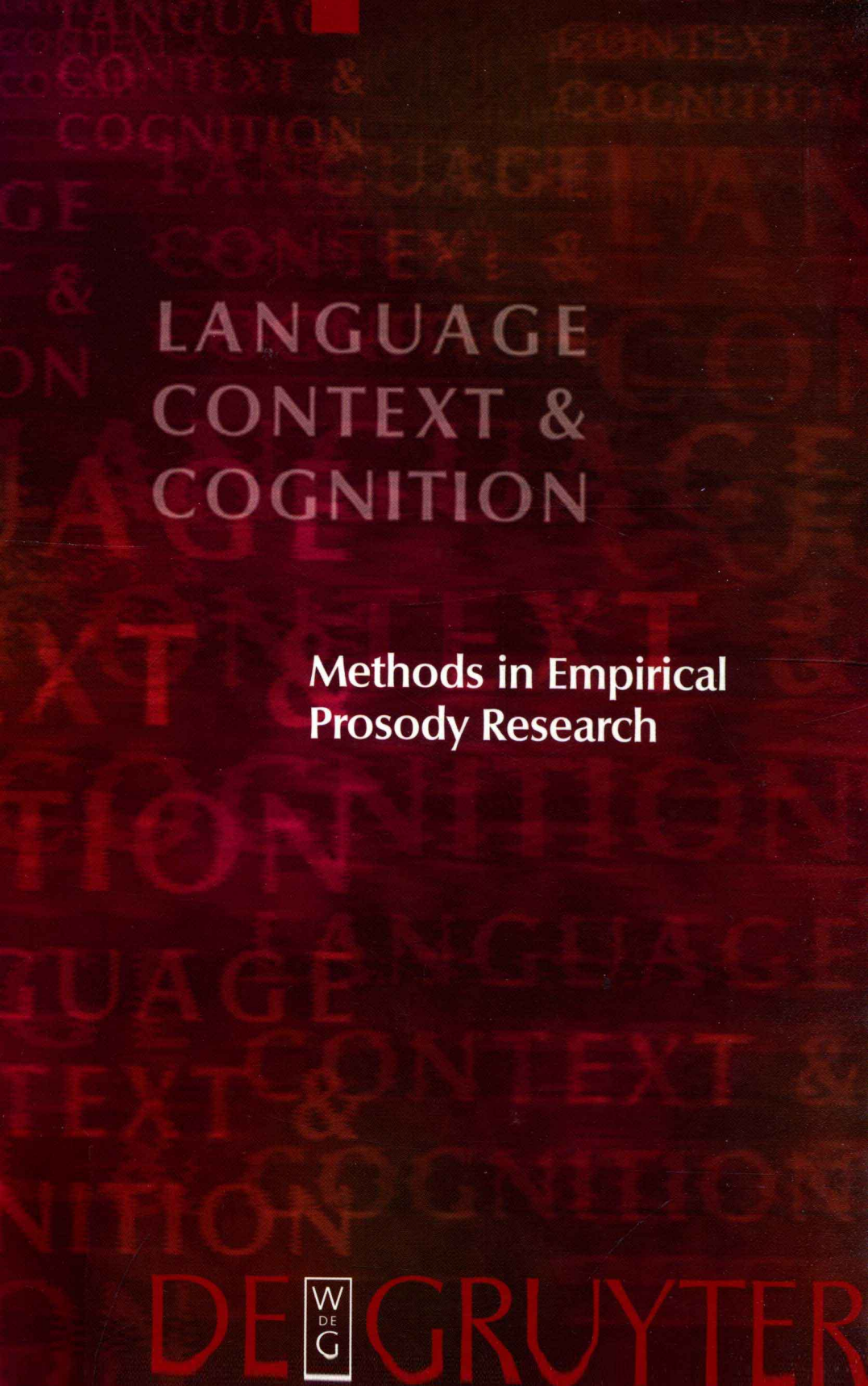
LANGUAGE
CONTEXT &
COGNITION
Methods in Empirical
Prosody Research
DE GRUYTER

Methods in Empirical Prosody Research

Edited by
Stefan Sudhoff, Denisa Lenertová, Roland Meyer,
Sandra Pappert, Petra Augurzky, Ina Mleinek,
Nicole Richter, Johannes Schließer

Walter de Gruyter · Berlin · New York

∞ Printed on acid-free paper which falls within the guidelines of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Methods in empirical prosody research / edited by Stefan Sudhoff ... [et al.].
p. cm. – (Language, context, and cognition ; v. 3)
Includes index.
ISBN-13: 978-3-11-018856-1 (alk. paper)
ISBN-10: 3-11-018856-2 (alk. paper)
1. Prosodic analysis (Linguistics) – Research – *Methodology.* I. Sudhoff, Stefan, 1977– II. Series.
P224.M48 2006
414'.6–dc22

2006015632

ISBN-13: 978-3-11-018856-1
ISBN-10: 3-11-018856-2

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Printed in Germany
Cover design: Christopher Schneider, Berlin
Printing and binding: Hubert & Co., Göttingen

Preface

Research in prosody has a relatively rich empirical tradition compared with other linguistic disciplines. Located at the intersection of theoretical linguistics, psycholinguistics, and phonetics, it can draw from a great variety of methods, ranging from the systematic observation of naturally occurring data to controlled laboratory experiments. While this allows the researcher to use multiple paradigms in testing the reliability of data patterns, the criteria for choosing the adequate method(s) to investigate a given research question are often unclear. Furthermore, the gathering, treatment, and analysis of prosodic data presupposes the thorough consideration of relevant methodological aspects. With these issues in mind, members of the PhD program "Universality and Diversity: Linguistic Structures and Processes" and of the DFG research group "Linguistic Foundations of Cognitive Science: Linguistic and Conceptual Knowledge" (FOR 349) organized the workshop "Experimental Prosody Research", which was held at the University of Leipzig in October 2004. This workshop was conceived as a series of tutorials to be presented by experts in the field.

The present volume can be seen as a follow-up publication of the workshop, covering a broadened array of topics. Researchers with different backgrounds were asked to contribute state-of-the-art papers on the choice and measurement of prosodic parameters, the establishment of prosodic categories, annotation structures for spoken-language data, and experimental methods for production and perception studies (including the construction of materials, modes of presentation, online vs. offline tasks, judgement scales, data processing, and statistical evaluation). The goal of the volume is to enable researchers in linguistics and related fields to make more informed decisions concerning their empirical prosody work.

The contributions can roughly be divided into three thematic sections: The first group of papers deal with acoustic parameters and their phonological correlates. They are followed by two articles on annotation structures and their application. The third section contains contributions on selected aspects of experimental design.

The initial contribution of the first thematic section deals with strategies of speech segmentation relevant for durational analyses. In their paper "Acoustic

Segment Durations in Prosodic Research: A Practical Guide", **Alice Turk, Satsuki Nakai and Mariko Sugahara** introduce a method based on identifying clearly recognizable acoustic landmarks, so-called oral consonantal constriction events, and use numerous examples from English, Dutch, Norwegian, Finnish, and Japanese to illustrate how this primarily spectrogram-based technique can be applied in practice. The authors present a range of segment types with constrictions accurately measurable in VCV and cluster contexts, and discuss how to handle problematic cases, such as the identification of laterals. They propose guidelines for carrying out tightly controlled prosodic experiments while minimizing the influences of confounding variables such as inconsistencies in speech rate. The proposed segmentation method is argued to be valuable for optimizing the automatic segmentation tools used in the investigation of durational parameters in speech corpora.

Dik J. Hermes' article "Stylization of Pitch Contours" discusses paradigms for the reduction of pitch contours to their perceptually essential properties by eliminating irrelevant details such as microprosodic fluctuations and interruptions due to unvoiced speech segments. While containing as little information as possible, the resulting representation must correspond to the mental image of a perceived pitch contour, i.e., a resynthesis based on the stylized contour should be perceptually equivalent to a resynthesis on the basis of the original contour. Hermes describes two different stylization methods. In the first, a pitch contour is reduced to as small a number of continuous straight lines (or curves) as possible without changing the perceived intonation, resulting in a "close-copy" stylization. Resting on the assumption that the perception of pitch contours depends, above all, on the pitch of the syllabic nuclei, the second method displays pitch contours as sequences of tones aligned with the syllable structure. The two approaches to the stylization of pitch contours are discussed with respect to their theoretical foundations, practical characteristics, possibilities and limitations.

Christophe d'Alessandro's chapter "Voice Source Parameters and Prosodic Analysis" focuses on the various sound parameters reflected in voice quality and discusses the difficulties related to their acoustic analysis. The author approaches voice source phenomena from the perspective of their linguistic (phonological), expressive, and phonostylistic functions. While voice source parameters are assumed to be of secondary importance compared to f0 in read speech and other laboratory speech conditions, they play a significant role in expressive communication situations, according to d'Alessandro. The range of parameters discussed in the paper include the degree of periodicity (e.g. in unvoiced, whispered or voiced speech), the voice open quotient (e.g. for strangled tones or lax voice), and the voice spectral tilt. The author points out that their investigation requires specific, and sometimes intricate signal analysis techniques. After presenting a summary of phonation types and voice quality dimensions in addition to a detailed review of voice source models, the

author describes techniques for the analysis of aperiodicities, voice pressure, vocal effort, and voice registers.

In his article "Prosody beyond Fundamental Frequency", **Greg Kochanski** argues that a description of the prosodic ability of the human voice must include a number of acoustic properties other than f0 and duration, such as loudness, slope of the speech spectrum, timbre, and degree of voicing. His point is based on information-theoretical estimates for the amount of information that must be carried by the prosodic properties of the sound and for the capacity of the individual acoustic channels. Kochanski discusses implications of this account both for the linguistic theory of prosody and experimental techniques, and algorithms for the relevant acoustic measurements. The experimental methods evaluated by the author relate to the questions as to (i) how much information can be transmitted (channel capacity), (ii) how a linguistic feature is encoded, and (iii) how much and which information is actually transmitted.

Klaus J. Kohler's paper "Paradigms in Experimental Prosodic Analysis: From Measurement to Function" evaluates present-day experimental prosody research, contrasting the prevalent Autosegmental-Metrical approach with the framework of Function-Oriented Experimental Phonetics. In Kohler's opinion, the former has neglected crucial ingredients of prosodic phonology such as time, communicative function, and the listener. Several perception experiments on the timing of peaks and valleys are discussed, the results of which are interpreted as evidence for a contour rather than a tone sequence model of intonation. As for the second factor, Kohler argues against the subordination of function to form and the restriction to linguistic (as opposed to paralinguistic) function. All prosodic categories must moreover be perceptually confirmed. The author introduces the Kiel Intonation Model (KIM), which rests on these assumptions, as an alternative model of prosodic phonology.

The next chapter, entitled "Information Structure and Prosody: Linguistic Categories for Spoken Language Annotation", shifts the focus to annotation structures and their application. **Stefan Baumann** deals with the annotation of prosody and information structure in West Germanic languages. While a spoken-language annotation system for prosody has previously been established (GToBI), a system for the annotation of information structure is still a desideratum. After a discussion of information structural dimensions such as theme/rheme, given/accessible/new, and focus/background, the paper introduces a multi-layer annotation system developed in the MULI (Multilingual Information Structure) project. The MULI system combines an annotation according to the information structural dimensions with an annotation of prosody, as amply exemplified by data from German.

The second contribution of this thematic section, "Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data" by **Dafydd Gibbon,** focuses on a methodology for deriving hierarchical models

of timing from large amounts of annotated speech. The author argues for a data-driven approach based on high-quality corpora of naturally occurring phonetic material, demonstrating how a representation of timing at the prosodic level can be induced automatically. A tree-building algorithm is used to scan the corpus for local differences in syllable duration and to construct an elaborate Time Tree. The latter is compared to a syntactic bracketing of the same material, resulting in a numeric Tree Similarity Index. The close match between iambic patterns in the Time Tree and syntactic constituents points to the distinction between closed class and open class items. Gibbon provides an in-depth discussion of the theoretical underpinnings of Time Tree induction, reaching from time in phonetics to local and global linear models of rhythm.

The third thematic section, related to aspects of experimental design, is introduced by **Fred Cummins'** article "Probing the Dynamics of Speech Production". The chapter concentrates on two novel experimental approaches to the study of macroscopic timing phenomena in speech, Speech Cycling and Synchronous Speech. They are developed to unveil internal characteristics of the speech production system by intervening in the speaking process, more specifically external forcing and mutual entrainment. In the first method, speech production is influenced by means of periodic auditory signals cueing phrase onsets or stress positions. The second technique involves two speakers reading a prepared text in synchrony with one another, which leads to a considerable reduction of interspeaker variability by suppressing unpredictable idiosyncratic and expressive features. Cummins argues that these methods are especially valuable for the examination of dynamic properties of speech, such as rhythmic organization, phrasing, and pausing.

In the next paper, "Using Interactive Tasks to Elicit Natural Dialogue", **Kiwako Ito and Shari R. Speer** focus on the use of scripted vs. spontaneous speech in prosodic experiments. Three studies examining PP attachment ambiguities are reviewed with respect to the varying degree of naturalness of the dialogue, and the extent to which disambiguating prosodic cues are produced. The authors show that experiments in which the speakers plan their utterances themselves yield the best results for both factors. They demonstrate how problems that arise with highly natural dialogue as well as with highly scripted speech can be avoided by using interactive speech tasks. Experiments on the interaction between the information status and tonal behavior within adjective-noun sequences serve to illustrate the tree decoration task. This real-world object manipulation task developed by the authors makes it possible to constrain the set of target phrases and to guide the speakers to produce certain contrasts. Due to the naturalness of the elicited data, this experimental design does not only provide reliable prosodic cues, but can also evoke prosodic patterns which have not been observed in reading experiments.

Concentrating on the distinction between presentational (H*) and contrastive (L+H*) pitch accents in English, **Duane G. Watson, Christine A. Gunlogson and Michael K. Tanenhaus** demonstrate that the eye-tracking method can be adopted as a useful paradigm in prosody research. In their chapter, entitled "Online Methods for the Investigation of Prosody", they illustrate how empirical research can shed light on theoretically grounded controversies concerning the categorical distinction between distinct accent types and their discourse function. By analyzing eye fixations in a restricted visual context, the authors show that the accent types have different distributions: While the presentational accent is compatible with a broad range of interpretations, the use of contrastive accents is much more restricted. According to the authors, eye-tracking experiments indicate that prosodic information can be used as a predictor for subsequent structures in real-time language comprehension. Possible extensions of this technique for use in related issues, such as prosodic boundary placement, are discussed in the paper.

Toni Rietveld and Aoju Chen's contribution, "How to Obtain and Process Perceptual Judgements of Intonational Meaning", focuses on perceptual studies of intonational meaning in Germanic languages. Intonation is used to convey linguistic and paralinguistic meaning, which is often assumed to correspond to the phonological representation of the pitch contour (expressing discrete form-function relations) and its phonetic implementation (signaling gradient meaning differences), respectively. The authors address the question of how to obtain and process perceptual judgements of contrasts of the latter kind. In the first part of the paper, four unidimensional scaling methods – Equal Appearing Interval Scale (EAI), Paired Comparisons (PC), Direct Magnitude Estimation (DME), and Visual Analogue Scale (VAS) – as well as a multidimensional scaling procedure are discussed. In the second part, the suitability of the EAI, DME, and VAS are evaluated on the basis of two experimental studies of the perception of friendliness and pitch register in Dutch. The results suggest that the VAS is most sensitive among the scales taken into consideration in capturing the perceived paralinguistic meaning differences, but that the EAI scale might be more suitable for dealing with the perception of properties that are more directly related to pitch, such as register.

A closely related issue is the topic of **Carlos Gussenhoven**'s paper, "Experimental Approaches to Establishing Discreteness of Intonational Contrasts". The author discusses methods that allow dissociating discrete from gradual differences between intonational pitch contours. The former are attributed to paralinguistically meaningful variation between different pronunciations of the same phonological contour, and the latter to distinct phonological categories. A number of experimental techniques are reviewed. The author suggests that it is better for participants to judge the appropriateness of a given contour as an imitation of another contour (passable-imitation task) than to assign the contour to abstract categories (categorical perception task). The discussion of the

consequences of alternative paradigms (the semantic difference task, the semantic scaling task, and the imitation task) leads to the conclusion that tasks should more explicitly engage listeners' intuitions about phonological identity in order to minimize the interference of purely phonetic factors in responses to phonologically equivalent contours.

In the final contribution, "Phonetic Grounding of Prosodic Categories", **Katrin Schneider, Britta Lintfert, Grzegorz Dogil, and Bernd Möbius** argue that prosodic categories emerge from probability distributions, which correspond to regions in the parametric phonetic space. According to the authors, their emergence results from the application of a speaker-internal analysis-by-synthesis process. Two methodological points relevant to this theoretical reasoning are discussed: the testing of the categorical status of prosodic events, and the mapping of prosodic categories onto the continuous acoustic parameters of the speech signal. With respect to the first issue, the authors demonstrate the design, stimulus generation, and evaluation of two experimental paradigms – categorical perception and the perceptual magnet effect – using the perception of boundary tones in German as an example. A computational model for the emergence of prosodic categories is introduced in connection with the second issue. This model is based on the probability distributions of phonetic parameters, the relevance of which is illustrated by a statistical frequency evaluation of a speech corpus.

As mentioned earlier, the idea for the present compendium grew out of the Leipzig workshop on "Empirical Prosody Research". We would like to thank the numerous participants of the workshop for their fruitful discussion and, above all, the presenters for their contributions and tutorials: Stefan Baumann, Aoju Chen, Carlos Gussenhoven, D. Robert Ladd, and Bert Remijsen. The workshop would not have been possible without generous funding by the DFG research group "Linguistic Foundations of Cognitive Science: Linguistic and Conceptual Knowledge" (FOR 349), and the DFG PhD program "Universality and Diversity: Linguistic Structures and Processes", both at the University of Leipzig. Their support is gratefully acknowledged. We would also like to thank Anita Steube, head of the research group and editor of "Language, Context, and Cognition", for her support and for the inclusion of this volume into the series. Our appreciation also goes out to Sebastian Hellmann for his technical assistance, and to David Dichelle for his proofreading work.

Leipzig, April 2006

The editors

Contents

Alice Turk, Satsuki Nakai (Edinburgh) &
Mariko Sugahara (Kyoto)*

Acoustic Segment Durations in Prosodic Research: A Practical Guide

1 Introduction

Carefully designed durational experiments are promising tools for testing and formulating theories of prosodic structure, its relationship with grammar, and its phonetic implementation. If properly designed, they allow for tight control of prosodic variables of interest, and can yield reliable durational measurements. Results from these tightly controlled experiments can then be used to form hypotheses about the way segment durations vary in more natural speech situations, or can be used to test hypotheses based on observations of natural speech corpora.

In this paper, we discuss methodological issues relating to such studies. In the first part of the paper, we outline principles of reliable and accurate acoustic speech segmentation that allow us to make inferences about the durations of consonantal constrictions and surrounding, mostly vocalic, intervals. In doing so, we discuss the relative segmentability of a range of segment types, in the hope that this will help researchers to design materials with the maximum likelihood of accurate segmentation. In the second part of the paper, we discuss additional methodological issues relating to the design of durational experiments. These include ways of designing materials to control for sources of known durational variability, and methods for eliciting prosodic contrasts.

We thank Matthew Aylett, Simon King, Peter Ladefoged, Jim Scobbie, Laurence White, Ivan Yuen, and especially Stefanie Shattuck-Hufnagel and Jim Sawusch for discussion of ideas presented here, and Bert Remijsen for detailed comments on a pre-final version of this chapter. We are also grateful to two anonymous reviewers for their useful comments, to Sari Kunnari for help in collecting the Finnish data, and to Kari Suomi and Richard Ogden for helpful information regarding Finnish phonology and phonetics. This work was supported by Leverhulme, and British Academy grants to the first and second authors, and an AHRC grant to the first and third authors.

2 Principles of acoustic speech segmentation

Segmenting the speech signal into phone-sized units is somewhat of an artificial task, since the gestures used to produce successive speech sounds overlap to a great degree, as illustrated in Browman and Goldstein (1990) and elsewhere. For example, the closing gesture tongue movement for /g/ in the phrase *Say guide walls* (Figure 1) begins before the end of the preceding vowel /e/, as evidenced by the rising F2 formant transition for this vowel.[1] This situation of articulatory overlap makes it difficult to determine the point in the acoustic signal where the vowel ends and the consonant begins. Nevertheless, there are often salient acoustic landmarks that correspond straightforwardly to recognisable articulatory events (Stevens, 2002). In particular, although we know that movement towards consonantal constriction begins earlier, abrupt spectral changes coincide with the onsets and releases of oral consonantal constrictions for the production of stops, fricatives, and affricates, as illustrated in Figure 1 (/s, g, d, z/) and Figure 2 (/s, p/).

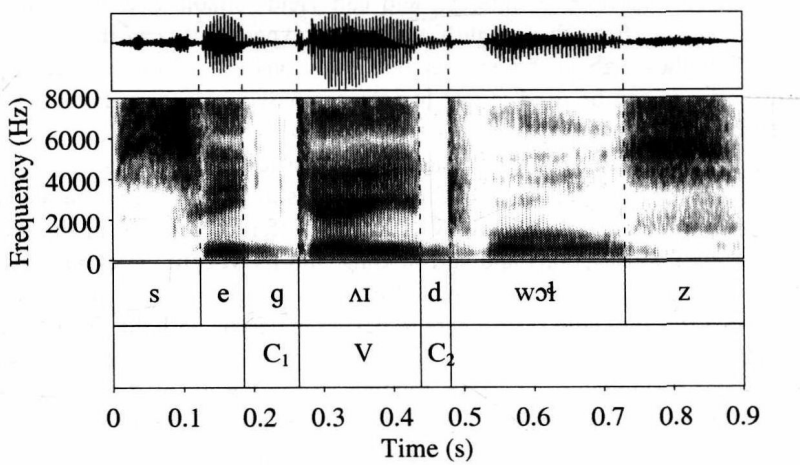

Figure 1: *Say guide walls,* spoken by a female Scottish English speaker

[1] /e/ is monophthongal in Scottish English.

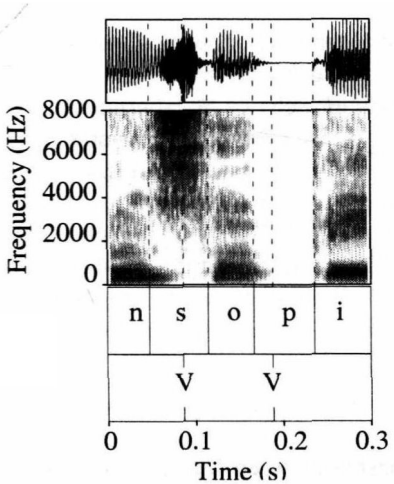

Figure 2: A fragment (underlined) from *MINUSTA "san" sopii kohtaan tuhat-kaksisataa* 'I THINK "san" fits [#] 1200', spoken by a female Northern Finnish speaker. *San* is a nonsense word. *V* in the second label tier indicates the offset of voicing for [s], and [p].

We propose that acoustic segment durations should be determined by the intervals that these oral consonantal constriction events define. Oral constriction criteria are preferable to criteria based on the onset or offset of voicing, since oral constriction criteria can be used comparably for many different classes of speech sounds, including voiced and voiceless oral stops, fricatives, affricates, and nasal stops. Although oral constriction and voicing criteria might be thought to be interchangeable in some cases, e.g. at the onsets of voiceless obstruent constrictions in vowel-voiceless obstruent sequences, e.g. as between word-medial [o] and [p] in *sopii*, voicing often persists after the onset of the phonologically voiceless constriction (e.g. [s] and [p] in Figure 2). In situations of this type, the oral constriction onset criterion is clearly preferable.

As can be clearly seen in Figure 2, oral constriction criteria can yield very different segment durations than criteria based on voicing. Similar discrepancies are observed in situations where aspirated voiceless stop offsets are measured. These potential differences also make a strong case for being explicit about segmentation criteria in reports, and above all for application of consistent segmentation criteria.

The duration of an interval between a C_1 constriction release landmark and a following C_2 constriction onset landmark in a C_1VC_2 sequence (e.g. the [ʌɪ] interval in Figure 1) is often described as the duration of a "vowel". We follow this convention here. However, this interval is not exclusively vocalic, since the so-called "vowel" duration includes formant transitions and burst noise that cue the identity of the surrounding consonants, in addition to any aspiration

from preceding voiceless aspirated stops. This point should be kept in mind when interpreting labels for such intervals.

We argue that the judicious choice of experimental materials can yield reliable, accurate durational measurements, if the materials contain alternations of salient oral consonantal constrictions and sonorant segments such as vowels. In particular, constriction onsets and releases are relatively easy to identify in: (1) stop consonants, e.g. [p, t, k, b, d, g], sibilants, e.g. [s, ʃ, z, ʒ], and affricates, e.g. [tʃ, dʒ] in VCV contexts and (2) non-homorganic clusters (clusters containing consonants of different places of articulation) differing in manner. We will discuss segmentation criteria for these sequences below.

2.1 Relative segmentability

There is a clear relationship between segmentation reliability and the strength of conclusions that can be drawn from experiments that use segmentation as part of their methodology. In order to ensure confidence in results of durational experiments, we recommend that materials be designed with the highest possible number of target segments whose durations can be reliably and accurately estimated.

In the following sections, we present detailed segmentation criteria for sequences of segments shown in Table 1. These criteria derive from the theory of the relationship between articulation and acoustics (see Stevens, 2002), and from our experience in segmenting American English, Standard Scottish English, and, to a lesser extent, Southern Standard British English, Standard Dutch, Northern Finnish and Standard Japanese. Although grounded in general acoustic theory as it relates to speech production, there may be language specific factors, such as allophonic variation, assimilation or coarticulation patterns that may make some of these specific criteria less applicable for particular languages or language varieties.

Table 1 includes 1) phones that we have found to be reliably segmentable in most contexts, 2) phones which we have found to be reliably segmentable in restricted contexts, 3) phones which we have found to be less reliably segmentable in most contexts, and 4) others which are to be avoided whenever possible. The phone classes mentioned in Table 1 conform to definitions given in Ladefoged (2001); we have defined additional terms not explicitly mentioned there. Not all phone types are included; we only discuss cases that we have had sufficient experience with to describe with confidence.

It should be noted that nasal stops are the most appropriate class of segments for experiments where both duration and F0 are of interest. Obstruents are known to raise or lower F0 in adjacent pitch periods depending on their voicing specification, and are therefore less appropriate for F0 analyses.

	Boundary between consonant and vowel in CV or VC sequences, where consonants are:	Boundary between two members of a consonant cluster, where phones in clusters are:
Reliably segmented in most contexts	Oral stops, e.g. [p, b, t, d, k, g] Sibilants, e.g. [s, ʃ, z, ʒ] Affricates, e.g. [tʃ, dʒ]	Oral stops, nasal stops, sibilants, and affricates in the following sequences, when these differ in place and manner of articulation: Sonorant consonant*-oral stop Sonorant consonant-sibilant Sonorant consonant-affricate Oral stop-sonorant consonant Sibilant-sonorant consonant Affricate-sonorant consonant Sibilant-oral stop Oral stop-sibilant Nasal stop-sibilant Sibilant-nasal stop *Sonorant consonant = approximants or nasal stops, e.g. [l, j, w, m]
Reliably segmented in some contexts	Nasal stops, e.g. [n, m] Weak voiceless fricatives, e.g. [f, θ]	
Less reliably segmented		Weak voiceless fricatives Nasal or voiceless stops in homorganic nasal-stop or stop-nasal clusters, e.g. [mp, pm]
To be avoided	Central and lateral approximants, e.g. [w,l]; [h] Weak voiced fricatives, e.g. [v, ð]	Voiceless and voiced consonants in homorganic clusters, e.g. [st], [mb] Consonants in clusters sharing manner of articulation, e.g. [pk], [bt], [mn], [sʃ] Stop-affricate clusters

Table 1: Relative segmentability of consonants in VCV and cluster contexts

2.2 Segmentation criteria

The detailed segmentation criteria that we present in the following sections are all based on the more general strategy of finding constriction onsets and re-

leases, as described above. Most of the criteria we discuss are based on spectral characteristics most easily seen in spectrograms. Waveforms can also be useful for segmentation since they show dips and rises in amplitude, which often correspond to the onsets of constrictions and their release. However, amplitude dips can sometimes be gradual on waveform displays, particularly when constrictions are voiced, and some types of frication noise can be difficult to distinguish from aspiration noise on waveform displays. For these reasons, we prefer to rely primarily on spectrograms for first-pass segmentation decisions within an accuracy of 5-10 ms, and on waveforms for more fine-grained segmentation decisions, once general boundary regions have been defined.

Note that segmentation accuracy will necessarily depend on factors other than segmentation criteria, namely 1) the sampling rate used in digitising the signal, and 2) the spectrogram analysis window size, assuming that spectrograms are used for segmentation, and 3) the degree to which each successive analysis window overlaps (frame shift). For example, a sampling rate of 16,000 Hz will yield accuracy within .0625 ms if segmenting on the waveform, but a spectrogram analysis window size of 5 ms (for 200 Hz wideband analysis) will limit the accuracy of spectrogram-based criteria to within this 5 ms window. The reduced accuracy of spectrogram-based criteria as compared to those based on the waveform supports the use of the waveform for final fine-grained segmentation once segment boundaries have already been determined within 5-10 ms.

When using visual displays for segmentation, it is easier to see gross spectral changes when these are zoomed out, or contain longer stretches of speech. We recommend more zoomed out spectrogram displays to determine general boundary regions, and more zoomed in waveform displays for determining exact boundary locations.

In the following sections, we discuss segmentation criteria in rough order of relative segmentability, as organised in Table 1.

2.2.1 Consonants in VCV contexts

Oral stops

In our experience, canonical variants of oral stops are generally easy to segment (see [g, d] in Figure 1, [p] in Figure 2). The onset of stop closures in VCV contexts are associated with 1) a decrease in overall amplitude, and 2) cessation of all but the lowest formant and harmonic energy. Although some stop closures are also accompanied by the cessation of voicing, many voiced stops and even some phonemically voiceless stops have voicing that continues through part or all of the stop closure (see [t] and the second [p] in Figure 3). In addition, for some vowel-voiceless stop sequences, voicing can stop earlier