Speech data acquisition: the underestimated challenge

Oliver Niebuhr, Alexis Michaud

Kieler Arbeiten in Linguistik und Phonetik (KALIPHO), Inst. f. Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS), Christian-Albrechts-Universität zu Kiel

2 Speech Data Acquisition: The Underestimated Challenge

Oliver Niebuhr
Analyse gesprochener Sprache
Allgemeine Sprachwissenschaft
Christian-Albrechts-Universität zu Kiel
niebuhr@isfas.uni-kiel.de

Alexis Michaud
International Research Institute MICA
Hanoi University of Science and Technology, CNRS, Grenoble INP, Vietnam
Langues et Civilisations à Tradition Orale, CNRS/Sorbonne Nouvelle, France
michaud.cnrs@gmail.com

The second half of the 20th century was the dawn of information technology; and we now live in the digital age. Experimental studies of prosody develop at a fast pace, in the context of an "explosion of evidence" (Janet Pierrehumbert, Speech Prosody 2010, Chicago). The ease with which anyone can now do recordings should not veil the complexity of the data collection process, however. This article aims at sensitizing students and scientists from the various fields of speech and language research to the fact that speech-data acquisition is an underestimated challenge. Eliciting data that reflect the communicative processes at play in language requires special precautions in devising experimental procedures and a fundamental understanding of both ends the elicitation process, speaker and recording facilities. The article compiles basic information on each of these requirements and recapitulates some pieces of practical advice, drawing many examples from prosody studies, a field where the thoughtful conception of experimental protocols is especially crucial

1. Introduction: Speech Data Acquisition as an Underestimated Challenge

The second half of the 20th century was the dawn of information technology; and we now live in the digital age. This results in an *"explosion of evidence"* (Janet Pierrehumbert, Speech Prosody 2010, Chicago), offering tremendous chances for the

analysis of spoken language. Phoneticians, linguists, speech therapists, speech technology specialists, anthropologists, and other researchers routinely record speech data the world over. There remains no technological obstacle to collecting speech data on all languages and dialects, and to sharing these data over the Internet. The ease and the speed with which recordings can now be conducted and shared should not veil the complexity of the data collection process, however.

Phonetics "calls on the methods of physiology, for speech is the product of mechanisms which are basically there to ensure survival of the human being; on the methods of physics, since the means by which speech is transmitted is acoustic in nature; on methods of psychology, as the acoustic speech-stream is received and processed by the auditory and neural systems; and on methods of linguistics, because the vocal message is made up of signs which belong to the codes of language" (Marchal 2009:ix). In addition to developing at least basic skills in physiology, physics, linguistics, and psychology, each of which has complexities of its own, people conducting phonetic research are expected to have a good understanding of statistical data treatment, combined with a command of one or more specific exploratory techniques, such as endoscopy, ultrasonography, palatography, aerodynamic measurements, motion tracking, electromagnetic articulography, or electroencephalography (for a description of the many components of a multisensor platform see Vaissière et al. 2010). As a result, it tends to be difficult to maintain a link between the phonetic sciences and fields of the humanities that are highly relevant for phonetic studies, and in particular for the study of prosody. Phoneticians' training does not necessarily include disciplines that would develop their awareness of the complexity and versatility of language, such as translation studies, languages, literature and stylistics, historical phonology, and sociolinguistics/ethnolinguistics. Moreover, the increasing use of digital and instrumental techniques in phonetic research is, taken by itself, a welcome development. But more and more phoneticians neglect explicit and intensive ear training, forgetting that an attentive, trained ear is the key to observations and hypotheses and hence the prerequisite for any analysis by digital and instrumental techniques. For example, we do not think that successful research on prosody can be done without the ability to produce and identify the prosodic patterns that one would like to analyse. As Barbosa (2012:33) puts it: *"The observation of a prosodic fact is never naïve, because formal instruction is necessary to see and to select what is relevant"*.

In summary, advances in phonetic technologies impose many challenges on modern phoneticians, and they can tend to replace rather than complement traditional skills. This has a direct bearing on data collection procedures. To a philologist studying written documents, it is clear that every detail potentially affects interpretation and analysis (The complexities of Greek and Latin texts are perfect examples; see, e.g., Probert 2009; Burkard 2014). Carrying the same standards into the field of speech data collection, it goes without saying that every speaker is unique, that no two recording situations are fully identical, and that human subjects participating in the experiments are no "vending machines" that produce the desired speech signals by paying and pressing a button. An experience of linguistic fieldwork, or of immersion learning of a foreign language, entails similar benefits in terms of awareness of the central importance of communicative intention (see in particular Bühler 1934, passim; Culioli 1995:15; Barnlund 2008), and of the wealth of expressive possibilities and redundant

encoding strategies open to the speaker at every instant (as emphasized, e.g., by Fónagy 2001). Researchers working on language and speech are no "signal hunters", but hunt for functions and meanings1 as reflected in the speech signal, which itself is only one of the dimensions of expression, together with gestures and facial expressions. The definition of tasks, their contextualization, and the selection of speakers are at the heart of the research process.

The diversification of the phonetic sciences is likely to continue, together with technological advances; the literature within each subfield is set to become more and more extensive, making it increasingly impractical for an individual to develop all the skills that would be useful as part of a phonetician's background. This results in modular approaches, as against a holistic approach to communication. What is at stake is no less than a cumulative approach to research. The quality of data collection is inseparable from the validity and depth of research results; and data sharing is indispensible to allow the community to evaluate the research results and build on them for further studies.

Against this background, the present article is primarily intended for an audience of advanced students of phonetics. However, it is hoped that it can also serve as a source of information for phonetic experts and researchers who have a basic understanding of phonetics but work in other linguistic disciplines, including speech technology. The present article summarizes some basic facts, methods, and problems concerning the three pillars of speech data acquisition: the speaker (§2), the task (§3), and the recording (§4). Discussion on these central topics build on our own experiences in the field and in the lab. Together, the chapters aim to convey to the reader in what sense data acquisition is an underestimated challenge. Readers who are pressed for time may want to jump straight to the Summary in section 5, which provides tips and recommenddations on how to meet the demands of specific research questions and achieve results of lasting value for the scientific community.

Given its aim, our article is both more comprehensive and introductory than other methodologically oriented papers such as those by Mosel (2006), Himmelmann (2006), Ito and Speer (2006), Xu (2011), Barbosa (2012), and Sun and Fletcher (2014), which are all highly recommended as further reading. Most readers are likely to know much if not most of what will be said. Different readers obviously have different degrees of prior familiarity with experimental phonetics; apologies are offered to any reader for whom nothing here is new.

[1] The two terms 'meaning' and 'function' tend not to be clearly separated in the literature – including in the present article, in which we simply use both terms in combination. In the long run, a thorough methodological discussion should address the issue of the detailed characterization of 'meaning' and 'function'. To venture a working definition, meanings refer to concrete or abstract entities or pieces of information that exist independently of the communication process and are encoded into phonetic signs. Functions, on the other hand, are conveyed by phonetic patterns that are attached to these phonetic signs; they refer to the rules and procedures of speech communication. If meanings are the driving force of speech communication, then functions are the control force of speech communication.

2. The speaker

2.1 Physiological, social, and cognitive factors

Individual voices differ from one another. Physiological differences are part of what Laver (1994, 27–28) refers to as the "organic level"; they are extralinguistic, but are nevertheless of great importance to analyzing and interpreting speech data. Age and body size are perfect examples for this (cf. Schötz 2006), affecting, among others, F0, speaking rate (or duration) and spectral characteristics such as formant frequencies. Physiological variables are intertwined with social variables. For instance, there are physiological and anatomical differences between the male and female speech production apparatus, which lends female speakers a higher and breathier voice as well as higher formant values and basically allows them to conduct more distinct articulatory movements than their male counterparts within the same time window (Sundberg 1979; Titze 1989; Simpson 2009, 2012). So, *"if we randomly pick out a group of male and female speakers of a language, we can expect to find several differences in their speech"* (Simpson 2009:637).

However, Simpson (2009) also stresses in his summarizing paper that gender differences in speech do not merely have a biophysical origin. Some differences are also due to learned, i.e. socially evoked behaviour, and the dividing line between these two sources of gender-related variation cannot always be easily determined. The social phenomenon of "doing gender" is well documented; it is an object of attention on the part of speakers themselves, and 'metalinguistic' awareness of gender differences in speech is widespread, particularly with respect to grammar and lexicon (cf. Anderwald 2014). Gender-related phonetic differences are less well documented. The frequent cross-linguistic finding that women speak slower and more clearly than men is probably at least to some degree attributable do "doing gender" (cf. Simpson 2009). Further, more well-defined differences between the speech of men and women are documented by Haas (1944) for Koasati, a Native American language. Sometimes women have exclusive mastery of certain speaking styles: mastering whispered speech, including the realization of tonal contrasts without voicing, used to be part of Thai women's traditional education (Abramson 1972). In languages where the differences are less codified, they are nonetheless present: Ambrazaitis (2005) found gender differences in the realization of terminal F0 falls at the ends of utterances in German and – more recently – also in English and Swedish (see also Peters 1999:63). Compared with male speakers, female speakers prefer pseudo-terminal falls that end in a deceleration and a slight, short rise at a relatively low intensity level (Ambrazaitis 2005). This pseudo terminal fall reduces the assertiveness/finality of the statement, as compared with a terminal fall. In extreme cases, this pattern might be mistaken for an actual falling-rising utterance-final intonation patterns, which has a different communicative function. Phonetically, the difference is not considerable: a rise on the order of 2 to 4 semitones for the pseudoterminal fall, of 6 semitones for a falling-rising utterance-final pattern.

Another socially-related phenomenon is the so-called 'phonetic entrainment' or 'phonetic accommodation'. That is, when two speakers are engaged in a dialogue, they become phonetically more similar to each other, particularly when the interaction is cooperative and/or when the two dialogue partners are congenial with each other (cf.

Lee et al. 2010). Phonetic entrainment can include levels and ranges of intensity and F0, voice quality (e.g., shimmer), and speaking rate (cf. Pardo 2006; Levitan and Hirschberg 2011; Heldner et al. 2010; Hirschberg 2011; Manson et al. 2013), as well as VOT patterns, vowel qualities, and speech reduction (Giles and Coupland 1991; for a summary: Kim 2012:14-29). Delvaux and Soquet (2007) provide evidence that a speaker tends to approximate the phonetic patterns of another speaker even when the latter is not present as a dialogue partner but just heard indirectly from a distance. The affected phonetic parameters are language-specific and differ, for example, between languages with and without lexical tone (Xia et al. 2014). Moroever, entrainment is not restricted to the phonetic domain. It can equally affect syntax and wording of utterances as well as body and face gestures (Nenkova et al. 2008; Reitter and Moore 2007; Ward and Litman 2007). Entrainment emerges quickly at the beginning of a dialogue, but can also increase further during a dialogue, which is why it is often conceptualized as a combination of lower-level cognitive and higher-level social skills (cf. Pickering and Garrod 2004).

A similar combination of cognitive and social factors probably accounts for the effects of musical training on linguistic habits. It is well documented that musical training affects the way the brain works, and hence constitutes an important source of cross speaker variation. Musically trained subjects outperform untrained subjects in experiments on the perception of prosody and intonation (cf. Schön et al. 2004; Pitt 1994) and speech comprehension in noise (Parbery-Clark et al. 2009), cf. Federman (2011) for an overview. Compared with the comprehensive research on perception, relatively little is known about effects of musical training on speech production. However, there is sufficient evidence to assume that musical training does affect speech production. For example, Stegemöller et al. (2008) found global spectral differences between speakers with and without musical training, and Graupe (2014) found her musically trained subjects to have a more distinct pronunciation, including larger F0 and intensity ranges and a lower speaking rate.

There are many more sources of cross-speaker variation that we cannot list here in full detail, including the individual adaptation of speakers to adverse speaking (i.e. Lombard) conditions (cf. Mixdorff et al. 2007).

2.2 Linguistic experience and linguistic skills

An individual's experience of different languages, dialects and sociolects also exerts a deep influence on the way s/he speaks. Among other spectacular experimental findings, it has been shown that one minute of exposure is enough for the ear to attune to a foreign accent (Clarke and Garrett 2004). Dialogue partners accommodate to one another in conversation; depending on their strategy to bring out or tone down social distance, dialogue partners will tend towards either convergence or divergence. This phonetic-accommodation effect described earlier in 2.1 can have farreaching consequences in the long run. It is reflected in amplified and entrenched forms in the speech of bilingual or multilingual speakers. A review entitled "The leakiness of bilinguals' sound systems" concludes that, *"although functional separation of sound systems may be both the aim and (actually quite frequent) achievement of bilinguals they are unable*

to avoid long term interference" (Watson 2002:245). Prosody is especially susceptible to the effects of language contact (Hualde 2003).

Moreover, some persons are more susceptible to influences of language contact than others. Subjects who have a good ear and a love of languages are potentially brilliant collaborators in speech data collection, able to understand tasks quickly and to apply instructions in a sensible and sensitive way. But their interest in language can adversely affect their performance when living in a language environment other than their mother tongue: speakers with high hearing sensitivity, good short-term and working memory, high attentional abilities and extensive vocabulary knowledge attune fastest to different accents (Janse and Adank 2012).

Finding out about language consultants' language abilities requires paying attention to the cultural issue of their perception of the languages they speak. Diglossia is an extremely common situation worldwide, but bilingualism is often non-egalitarian, with a prestigious national standard on the one hand and a local variety debased as 'dialect' or 'patois' on the other (on the distinction between egalitarian and nonegalitarian bilingualism: Haudricourt 1961; François 2012). In countries that enforce the adoption of a national standard, varieties other than the norm are deprecated, and speakers may consider it inappropriate to mention their mother tongue – considered as coarse, ridiculous or useless – in a curriculum vitae or a questionnaire. In China, students' résumés typically indicate Chinese as mother tongue, plus a degree of proficiency in English graded along the TOEFL scale, and sometimes other foreign languages. A native speaker of Wu or Min Chinese – branches of Sinitic that are not intelligible to speakers of Mandarin Chinese – may avoid mention of this competence in a language variety that is referred to in China as a 'dialect' and has no status as a language of culture and education. For research purposes, it may obviously be misleading to pool results from speakers whose native language has six to nine tones (see, e.g., Shen 2013) with those of native speakers of Mandarin, which has four. This is an extreme example, but issues of dialectal differences and dialect contact are well worth scrutinizing even when studying the national language of a country that has less language diversity, for instance in Europe and North America.

The second author of this paper can report first-hand on a case of code-switching in the course of data elicitation. He participated in a study of nasalization involving fiberscopic imaging of the velopharyngeal port. He unwittingly switched to Vietnamese mode when reading the logatoms /ap/, /at/ and /ak/. That day, he was accompanied by a Vietnamese speaker who was going to record a list of words designed for the study of syllables with final consonants, as an extension of studies based on electroglottographic data (Michaud 2004a) and on airflow measurements (Michaud et al. 2006). This example illustrates the fact that even in seemingly simple tasks, which in principle do not involve a high cognitive load or create particular fatigue, there can be interference between one's native language and other speech varieties, even those in which the speaker is not bilingual. It appears highly advisable to record native speakers in their home country, but even then, a sensitive inquiry into the speakers' language experience is highly advisable.

2.3 Individual strategies and preferences

To a present-day audience, it may not be necessary to emphasize the diversity of individual strategies in speech production, which is now increasingly recognized and documented. Divergences extend on a continuum from transient through idiosyncratic to cross-dialectal. They can yield decisive insights into language structures and their evolution: for instance, analyses of speaker-specific strategies shed light on how different articulatory configurations are able to generate similar acoustic outputs for vowels (Johnson et al. 1993; Hoole 1999) and consonants (Guenther et al. 1999). Investigations into connected speech processes show that some speakers do produce complete assimilations or elisions in a given context while others do not (Nolan 1992; Ellis and Hardcastle 2002; Kühnert and Hoole 2004; Niebuhr et al. 2011b). Such findings on diversity depending on speakers and styles make a major contribution to shifting connected speech processes from a cognitive level of categorical feature-based operations to a level of basically gradual articulatory interactions. In turn, this offered a fresh perspective on such fundamental issues as the role of the phoneme in speech production and perception (Gow 2003; Niebuhr and Meunier 2011; Kohler and Niebuhr 2011). Speaker-specific differences sometimes stand in a close relationship of correspondence with trading relations as brought out by perception experiments. The perception of a well-established intonational contrast in German involves an interplay of peak alignment and peak shape (Niebuhr 2007a); production data confirm that peak alignment and peak shape are both used by speakers to signal the two contrasting intonation patterns (Niebuhr et al. 2011a). However, the 34 analyzed speakers differ in the extent to which they make use of the alignment and shape cues. While the majority of speakers use both F0-peak parameters to different degrees, a small group of speakers (15%) – the "shapers" – prefer to signal the two contrastive intonation patterns by means of peak-shape differences alone. Pure "aligners" are about twice as numerous.

Similarly, making syllables perceptually salient and signalling the differences between broad, narrow, and contrastive focus accents both involve a number of articulatory and phonatory cues; and besides the fact that these cues are used in language-specific ways (cf. Andreeva and Barry 2012), there are also differences between speakers of the same language. For example, some speakers make more extensive use of changes in F0 peak range and timing, whereas others prefer using local or global changes in the duration structure or variation in articulatory dynamics and precision (cf. Hermes et al. 2008; Cangemi 2013). Although some of these differences may actually be an artefact of the elicitation task, reflecting the speaker-specific degree to which the signalling of contrastive focus is coloured by emphatic accentuation (cf. Görs and Niebuhr 2012), there is no doubt that prominence, rhythm, and focus all involve individual differences. Recent analyses of Northern Frisian prosody showed that some speakers vary the perceptual prominence of syllables by lifting the F0 maximum of the associated pitch-accent peak, whereas other flatten and hence extend the F0-peak maximum. So, in addition to distinguishing "aligners" and "shapers", it may also be necessary to search and separate "lifters" and "flatteners" (cf. Niebuhr and Hoekstra 2014). Both lifting and flattening F0 peak maxima are suitable means to make high pitch stand out in the listeners' ears, and high pitch is well known to be an attention-attracting signal. Thus, this prominencerelated example shows that we generally need a

better understanding of how acoustic parameters merge into the decisive perceptual parameters in order to explain and anticipate individual strategies and preferences.

A textbook example of a discovery based on the observation of cross-speaker differences is the study of Khmer by Eugénie Henderson. She worked mainly with one speaker, also checking the results with a second speaker; both were students at the School of Oriental and African Studies in London. Differences between the pronunciations of these two subjects helped her identify a major process of the historical phonology of Khmer: registrogenesis – the transphonologization of laryngeal oppositions on onsets.

> "The differences in usage lay chiefly in (1) the realization of the registers, and (2) the use in rapid speech of alternative forms such as those described on p. 172. Mr. Keng, as a philosophy student with literary and dramatic leanings, was aware of and interested in language from both the philosophic and aesthetic standpoints. His style of utterance was in general more deliberate and controlled than that of Mr. Mongkry, who as a student of economics was less concerned with language for its own sake. The two styles complemented each other well. Mr. Keng's style was helpful in that the different voice quality and manner of utterance of the two registers were clearly, sometimes startlingly, recognizable, even in fairly rapid speech, whereas Mr. Mongkry appeared often to make no distinction other than that of vowel quality. On the other hand, Mr. Mongkry's style of utterance was valuable for the ease and naturalness with which the alternative pronunciations proper to rapid speech were forthcoming" (Henderson 1952:149).

E. Henderson can be said to have been extremely fortunate to have come across two speakers who exemplified widely different pronunciations, one of whom was a strongly conservative speaker. However, in Pasteur's often-cited phrase (1915 [1939:131]), *"chance favours only the prepared mind"*: it is much to Henderson's credit that, rather than choosing one of the two speakers as her reference – which would have saved trouble –, she noted the differences and was able to identify the direction of change. Her findings set the key for a series of studies of register which proved fundamental to an understanding of the historical phonology of a great number of languages in East and Southeast Asia (Huffman 1976; Ferlus 1979; Edmondson and Gregerson 1993; Brunelle 2012; for a review see Michaud 2012).

2.4 The relationship between the consultants and the researcher

For field workers, the paramount importance of the relationship established with language consultants is well-recognized. *"In order to avoid disappointment and frustration, some time needs to be allocated for identifying [the consultants'] strengths and weaknesses, and most important, they themselves need some time to overcome shyness and insecurity and discover their own talents and interests"* (Mosel 2006:72; see also Mithun 2001). The example of Henderson's study of Khmer illustrates the fact that this process of mutual understanding does not necessarily take a very long time: rather, it is an issue of the investigator's attention to human subjects' personality, and to the stylistic preferences that contribute to shaping their pronunciation. Some phonetici-

ans choose to have the least possible contact with the experimental subjects taking part in their experiments; to us, this appears as a misguided interpretation of the notion of scientific objectivity. Scientific objectivity by no means requires the investigator to overlook such important parameters.

Good communication between the investigator and the participants in an experiment is essential to assessing to what extent differences found across the different participants' data sets reflect ingrained speaker-specific strategies, and to what extent they reflect different understandings of the tasks to be performed. The default hypothesis is that the experimental condition is the same for all speakers, and that differences in the recorded data therefore reflect cross-speaker differences. But different subjects may have interpreted the instructions differently, so that the differences in the data reflect in part the stylistic choices that they adopted: an experiment providing more precise guidance, such as a more explicit contextualization of the communicative setting that the experiment aims to simulate, may bring out greater closeness between speakers. In order to spot and interpret speaker-specific strategies, and to adjust the data collection procedure accordingly, the experimenter requires a trained ear, as well as a good command of the investigated language.

A sensitive definition of tasks also benefits greatly from exchanges with the subjects, especially when adapting a setup originally devised for another language. For instance, a study of Vietnamese prosody (Dung et al. 1998) initially calqued studies of Germanic or Romance languages, and attempted to contrast segmentally identical declarative and interrogative sentence pairs such as "Bao đi Việt Nam" ('Bao goes to Vietnam') and "Bao đi Việt Nam?" ('Is Bao going to Vietnam?'). This setup neglected the central role played by particles in conveying sentence mode in Vietnamese. Recordings were carried out in France. Bilingual speakers who had been living in France for many years had no difficulty in producing and differentiating the sentence pairs, as in French, whereas speakers who had just arrived and had little command of French practically refused to read such interrogative sentences. *"Either they spontaneously added a final particle, à, or they pronounced them in a very emphatic, exclamatory way, or on the contrary like the declarative counterparts"* (Dung et al. 1998:401). In less extreme cases, speakers may simply comply with the instructions, silencing any misgivings they may have – unless the investigator takes care to discuss the experimental setup with them.

Applying an experimental setting with new speakers, and on a new language, requires thorough re-examination of the method (a highly recommended reading on this topic is Vaissière 2004). An extreme case in point is that of an experiment on tone identification, performed under fieldwork conditions: the investigators calqued a procedure that had been used for a national language, playing a signal from which segmental information had been masked and requiring listeners to select one of several real words (presented in written form) constituting minimal sets or quasiminimal sets. Speakers of the language under investigation had some command of the national language in which the words were presented to them, but the task of recognizing the written words, translating them mentally, and matching the tone pattern of the heard stimulus with their tonal representation of the minimal sets in their native language proved highly challenging, so that the participants' performance was poor; data for some of them had to be discarded altogether.

A side advantage of experimental setups based on a sensitive cultural and socio-linguistic contextualization is that they stand up to the highest ethical standards, and answer the concerns embodied in the guidelines for human subject research of the investigators' home institutions and funding agencies. Distress caused by culturally/socially inappropriate recording tasks could be considered as a form of abuse of human subjects. Beyond formal guidelines, which can hardly anticipate the range of actual situations, the responsibility for relating with language consultants in the best possible way is ultimately the researcher's own; and ethical concerns coincide with scientific concerns. *"Ethical issues are embedded in a host of other '- ical' issues, such as methodological and technological ones"* (Grinevald 2006:347): adopting a socially/culturally appropriate behaviour, valuing the knowledge that the consultants share with the investigator, giving a fair compensation for their time and effort, explaining the process of data collection and research, and preserving the collected data, are both ethical imperatives and important aspects of successful data collection.

3. The task

3.1 Recording settings and the issue of "laboratory speech"

The out-of-the-way setting of a recording booth can be conducive to out-of-the-way linguistic behaviour, in cases where the speaker lacks a real addressee or a real communicative task to perform. We were keenly reminded of this when participating as subjects in an experiment on foreign-accented English: the task consisted in telling the story of "Little Red Riding Hood" under two conditions, once with a child of age 10 present in the booth to serve as audience, and once alone in the recording booth, without an audience. Being familiar with recording studios, we did not expect to be deeply influenced by these different settings, but the difference proved considerable. It was excruciatingly difficult to flesh out a narrative without an audience; this was reflected in numerous disfluencies. Phonetic studies confirm that speakers behave differently when reading isolated sentences or monologues than when reading turns of dialogue together with another speaker. A comparison of read monologues and dialogues by Niebuhr et al. (2010) shows that, even though the style of the script was the same in both cases – an informal style, intended as a close approximation of conversational speech –, the read dialogues were prosodically closer to spontaneous dialogues (as recorded and analyzed by Mixdorff and Pfitzinger 2005) in terms of F0 level, declination and variability, speaking rate and phonation mode.

Researchers in phonetics are often aware of the potential distance between linguistic behaviour in the lab and outside, witness this reflection found in the introduction to "Intonation systems: A survey of twenty languages":

> "The majority of the work reported in this volume is based on the analysis of a form of speech which has come to be known, sometimes rather disparagingly, as "laboratory speech", consisting of isolated sentences pronounced out of context, usually read rather than produced spontaneously (…). An obvious question which needs to be answered is how far does variability in the situations in which speech is produced influence the results obtained

> under these conditions? To what degree do generalisations obtained from isolated sentences apply to more spontaneous situations of communication? (…) There is obviously still a great deal of work to be done in this area before we can even begin to answer these questions." (Hirst and Di Cristo 1998:43)

While it is often difficult to assess in detail the influence exerted by laboratory conditions, it is clear that the settings of a recording exert a decisive influence on a speaker's performance. The subjects taking part in an experiment have some expectations about the experiment, and some representations about what a phonetics laboratory may look like. They may feel called upon to adopt a specific style, in unpredictable ways. When confronting a microphone, some speakers may adopt a more formal, deliberate style of speech than the investigator aims to capture; visual details such as the distance to the microphone, and the presence of a pop shield hiding the microphone from view, all contribute to shaping the subject'sexperience, relating to highly personal factors such as their fondness or dislike for public addresses, and their degree of self-confidence in oral expression.

People maintaining online databases and language archives often find it difficult to elicit reasonably detailed metadata from the researchers: information about the speakers, the recording tasks, the time of recording… Researchers have a lot on their plate, and the task of documenting their data sets may appear to them as a distraction from research – no matter how interested they are in these data, whose importance to research they acknowledge in principle. For instance, an archivist asking researchers to indicate the time of day when each recording was made is likely to be considered too fussy. Yet there is evidence that this parameter exerts an influence on speech. Görs (2011) created a corpus of more than 30 German speakers, who read texts (i) early in the morning; (ii) at noon; and (iii) late in the evening. She found systematic prosodic differences as a function of the time of day. In the morning, speakers show a slower speaking rate and a lower average F0, as well as stronger glottalization at prosodic boundaries. Speaking rate and average F0 increase at noon; the same applies to the level of speech reduction. In the evening, average F0 is lower again; the speaking rate remains high, but with fewer speech reductions; and voice quality is overall breathier, among other differences.

The personal experiences and findings summarized in this chapter boil down to the self-evidence that communication is context-sensitive. 'Spontaneous speech' is not a homogeneous category; and 'naturalness' is not a straightforward criterion to assess speech data, since speech can be said to constitute a natural response to a particular setting, even in the case of "unnatural" settings. Ultimately, this means that speech recordings from one setting cannot be more natural than speech recordings from another setting. Speech data of any kind can be described as a natural response to the settings under which they were recorded; in this sense, 'naturalness' is not a relevant criterion to evaluate speech recordings (Wagener 1986). What matters is the investigator's in-depth understanding of the communication setting. Speech data recorded at the linguist's initiative under highly controlled laboratory conditions can offer an appropriate basis for research, provided the researcher ensures that the communication setting is well-defined, and is clarified to the speakers' satisfaction.

3.2 The range of recording tasks and cross-task differences

Salient differences are observed across different types of elicited speech material. The investigator should be aware of the implications of the choice of materials. For classifying and comparing basic types of speech material, it appears convenient to start from a six-way typology. It relies on the fact that recordings can be made with and without a dialogue partner and on a read or spontaneous (i.e. unscripted) basis[2]. In addition to these two binary parameters, isolated words/logatoms – typically monosyllabic nonce words like [bab], [pap] and [pip] – and isolated sentences should be regarded as two separate subtypes of read monologues. In this light, one may distinguish six (4+2) types of speech materials: isolated logatoms or words; isolated sentences; read monologues; read dialogues; unscripted monologues; and unscripted dialogues.

The methods behind these six types can be rated along various dimensions, among which we will discuss five: (i) degree of control over experimental variables (i.e. dependent and independent variables) as well as other variables (control variables); (ii) event density: the number of analyzable tokens per time unit; (iii) expressiveness; (iv) communicative intention: the speaker's concern to actually convey a message; and (v) homogeneity of behaviour: the probability that the elicitation condition is defined in such a way that it leads speakers to behave in a comparable way. The diagram in Figure 1 is an attempt to represent, how the six types of methods perform in terms of these five dimensions. The performance is given as a simple relative ranking from 1 (worst) to 6 (best), based on notes and findings in the literature and on our own experience.

First of all, Figure 1 brings out the evident fact that there is no ideal method/material that performs best on all dimensions. Methods based on read speech allow for a high degree of control and yield a relatively high event density. However, they tend to dampen speaker involvement and hence do not perform well in terms of expressivity and communicative intention. This is particularly true for read monologues – such as newspaper texts (Amdal and Svendsen 2006) or prose (Zellers and Post 2012) – as well as for readings of logatoms and isolated sentences. Isolated logatoms allow for an even greater control in terms of prosody than isolated sentences (cf. Cooke and Scharenborg 2008); and as they are shorter, the event density is also higher. Dialogues increase expressiveness, informality and communicative intention (see, e.g., the evidence from Dutch presented by Ernestus 2000), and the presence of a dialogue partner stabilizes the speech behaviour of the recorded subject (Fitzpatrick et al. 2011). A richer semantic-pragmatic context has the same stabilizing effect (this is referred to as the "richness principle" in Xu 2012). This is another reason why one speaker's homogeneity of

[2] 'Unscripted speech' is in our opinion a more precise term than 'spontaneous speech', because all it means is that speakers do not produce predetermined utterances. The attribute 'spontaneous' can easily be misinterpreted as 'impulsive', 'instinctive', or 'automatic' and hence associated with a particularly emotional or agitated way of speaking, which can, but need not be applicable. However, 'spontaneous speech' is the more established term, which is why we use the two terms interchangeably, focussing on 'unscripted' speech in the context of the present section (3.2) because of its focus on the nature of the tasks entrusted to the speakers.

behaviour paradoxically increases from isolated logatoms and sentences through read monologues and dialogues to unscripted dialogues.

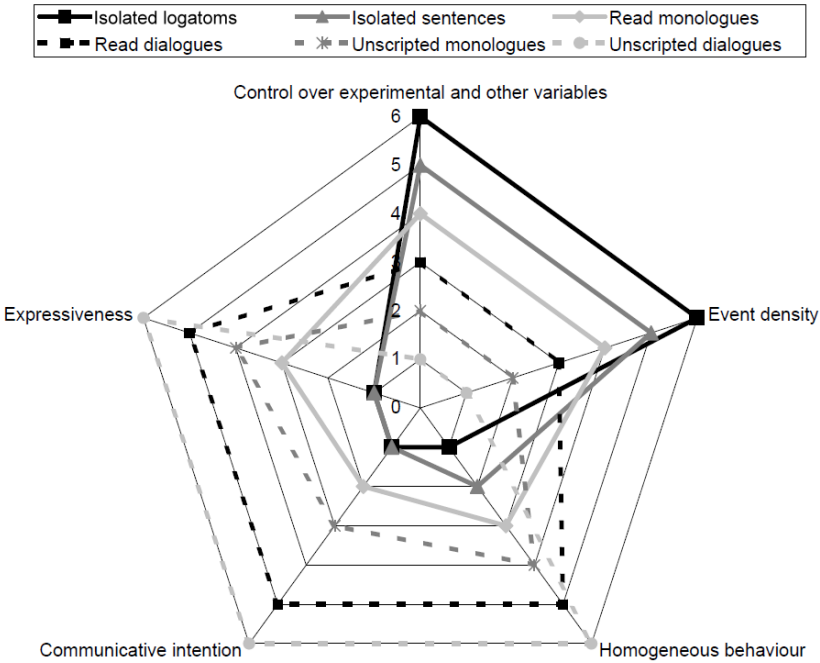

Figure 1. Ranking of six basic types of recording tasks on five dimensions that represent characteristics of experimental designs and speech communication.

A homogeneous speech behaviour is not only important for cross-speaker comparisons, but also with regard to replicability. Some research questions require the same speech material to be recorded twice with different recording devices. For instance, to study nasality, fiberscopy and airflow provide complementary perspectives, but recording both at the same time is impractical. To overcome this difficulty, it is possible to record nasal airflow and images of the velopharyngeal port separately, in two sessions, and to time-align the two data sets on the basis of landmarks on the acoustic signals recorded under both setups. For these post-aligned data to yield valuable results, the subjects' performance under both setups must be as close as possible to complete identity.

The disadvantages of dialogues are their lower event density (i.e. conducting a study becomes much more time-consuming) and the lack of control over experimental and other variables, particularly in the case of unscripted dialogues. Prototypical examples of unscripted dialogues are free conversations without any guidelines or topic specifications (cf. CID, Bertrand et al. 2008) or recordings of TV or radio broadcasts (cf. RUNDKAST, Amdal et al. 2008). Broadcast speech often constitutes the backbone of the large databases used in speech processing; in-depth statistical treatment of these databases sheds new light on sound systems (see, e.g., Gendrot and Adda-Decker

2007), but a limitation for prosody research is that broadcast speech is tilted towards a relatively narrow range of styles.

On the whole, read dialogues (like the KIESEL corpus described in Niebuhr 2010) seem to strike a reasonable compromise between the conflicting demands of symmetry, on the one hand, and ecological validity, on the other. Read dialogues combine an informal, expressive speaking style – which can be enhanced by using a corresponding orthography and font type – with relatively high degrees of communicative intention, homogeneous behaviour, event density, and a relatively high degree of control over experimental and other variables. The control of the experimenter in read dialogues extends to the semantic-pragmatic context, which is why read dialogues are ranked higher than unscripted monologues in terms of the homogeneous behaviour of speakers. Moreover, the control over the semanticpragmatic context can also be exploited to elicit specific melodic signs on key words. Finally, by selecting and combining appropriate dialogue partners, experimenters can make use of phonetic entrainment, in order to direct the speech behaviour of the two speakers into a certain direction. For example, if a privy dialogue partner is instructed to speak in a highly-reduced or very expressive fashion, then this speech behaviour is likely to rub off to some degree on the naïve dialogue partner. This entrainment strategy is of course also applicable to unscripted dialogues. Even if there is no privy dialogue partner, it is possible for the investigator to elicit different speech behaviour from his/her subject simply by matching the same speaker with different dialogue partners.

Figure 1 only constitutes a convenient means to represent several parameters, in order to compare widely different types of methods and materials. In detail, there is of course much more to say about each of these methods, and about strategies to improve each type of elicited speech material along several dimensions.

For instance, in unscripted dialogues, the event density can be increased, and controlling elements added: this is the famous case of ‘Map tasks’ (Anderson et al. 1991), which make it possible to introduce key words. A privy dialogue partner in a Map task recording can furthermore trigger certain communicative actions of the speaker. For example, in the study of Görs and Niebuhr (2012), a privy dialogue partner repeatedly pretended misundertandings, allowing for the elicitation of key words with narrow-focus intonation patterns. A privy dialogue partner can also help a speaker overcome the intimidating effects of recording-booth settings, before or during the actual recordings (see, e.g., Torreira et al. 2010). Most map task data have been elicited in Indo-European languages like American and Australian English, German, and Italian. However, given its success, its transfer to a wide range of languages appears feasible and promising.

Tasks similar to the Map task are the ‘Shape-Display task’ (cf. Fon 2006), or the ‘Appointment-Making task’ and the ‘Videotask’, which were used in the collection of German data by Simpson et al. (1997), Peters (2005), and Landgraf (2014). While the appointment-making scenario generates a high number of day, time, and place expressions, the Videotask scenario, in which the speakers first see two slightly divergent video clips of their favourite TV series and then confer to find the differences, additionally exploits an emotionally charged common ground between the speakers in order to enhance their expressiveness. The Videotask idea can be implemented with very different types of broadcasts, including cartoons, and it has been

shown for the latter type of broadcasts that the Videotask is basically able to trigger phonetic entrainment, just as real everyday dialogues (cf. Mixdorff et al. 2014).

The appointment-making scenario and the Videotask scenario are representatives of two different strategies that have been used to elicit unscripted dialogues: role-play tasks and quiz tasks. The appointment-making scenario is a typical role-play task. Other role-play tasks are a 'Sales Conversation' (cf. Ernestus 2000) and a 'Design-Team Project Meeting' (cf. http://corpus.amiproject.org/). Further typical quiz tasks are, for example, the 'Picture-Difference Task' (cf. Turco et al. 2011) and the 'Joint Crossword Puzzle Solving' used by Crawford et al. (1994). The Map task belongs to yet another type of task that may be called 'Instruction-Giving task'. Other examples are the 'Tree-Decoration task' (cf. Ito and Speer 2006), the 'Picture-Drawing task' of Spilková et al. (2010), the 'Card Task' (Maffia et al. 2014), and the 'Toy Game'. Unlike the Map task, the Toy Game has already been 16 successfully applied for eliciting natural conversation and prosody in the field. It is a simple, portable set up developed in conjunction with the 'Dene Speech Atlas' (http://ling.rochester.edu/people/mcdonough/dnld/JMcDonough/dene-speechatlas.html)

In the Toy Game, two players sit on opposite sides of a table with an occlusion between them. On a table in front of each player is a sheet of paper and some small objects (toy animals, cups, fences, etc.). The sheets of paper have three shapes drawn on it: circle, square and triangle. Both sheets are the same. The goal of the game is for both players to have the same arrangements on their sheets. How players proceed can be given some leeway, but in general players accomplish the task by taking turns asking questions, starting with one player, then the second player gets a turn. Recording begins well before the game begins, because the interaction that takes place in agreeing on names for objects is extremely useful for later analysis. The Toy Game is typically played three times with increasing complexity. The first game is a short warm up game. In the second game, the two players still have the same small number of items but in different arrangements. In the third game, there is an increase in both the number of types of toys and the number of tokens of each toy.

It may be assumed that role-play tasks perform better in terms of event density and experimental control (richness of semantic-pragmatic context). But they are outperformed by quiz and instruction-giving tasks with respect to expressiveness and communicative intention. In instruction-giving tasks, it seems also easier to foist the speaker on a privy dialogue partner, who then takes the role of the instruction receiver.[3]

[3] It need not be forgotten in this context that the experimenter's creativity to control the behavioural responses of subjects needs to be channelled by ethical considerations. These are formalized as ethical guidelines at some research institutions, but in practice the bulk of the real responsibility rests with the investigator. Foisting a privy dialogue partner on speakers is a type of deception; instructions that lack crucial information or deliberately provide misinformation in order to distract the subjects from the actual aim of the experiment are also problematic. Such strategies are common practice in many fields of research, most prominently in psychology, and tend to be considered as ethically acceptable, so long as the behavioural responses are interpreted in view of the distractor strategy. Other elicitation scenarios can create serious conflicts by implicitly

Quiz tasks are basically applicable for speakers of very different cultures in the field and in the lab. Moreover, when the different stimuli are shown (simultaneously or subsequently) to the same speaker, quiz tasks can also be used to elicit monologues, which is not possible with the role-play or instruction-giving tasks.

In a similar way as for unscripted dialogues, unscripted monologues can be based on retelling picture stories (cf. Iwashita et al. 2001; Mosel 2006, 2011) in order to include key words and/or a semantic-pragmatic context frame that ensures homogeneity of the speech behaviour. Alternatively, speakers can be asked to recite lyrics, poems or traditional texts that they know; this can make them feel more comfortable compared with previously unknown picture stories, but recitation constitutes a highly specific activity, often associated with specific styles. These tasks and similar other tasks as well as suitable elicitation material like "The pear story" or "Frog, where are you?" are explained in more detail in the book "Questionnaire on Information Structure (QUIS): Reference Manual" by Skopeteas et al. (2006). When eliciting monologues, it is useful for the speaker to have an addressee. Even if s/he does not say anything, subjects feel more comfortable and produce speech in a different way when the act of speaking is a social activity (cf. 3.1 and Fitzpatrick et al. 2011).

Another way to elicit expressive unscripted monologues is to record speakers during or after computer games (cf. Mixdorff 2004). This creates a fairly specific semantic-pragmatic context frame and allows for the elicitation of key words. Johnstone et al. (2005) even controlled the outcome of the games (win or failure) to stimulate positive and negative emotions. Similar manipulations of the environmental conditions were used by Maffia et al. (2014) for eliciting expressiveness and emotions during Card-Task dialogues.

Finally, in accordance with the conclusion in 2.1, Figure 1 suggests that investigating a research question on speech production should involve several recordings tasks, starting from isolated logatoms or sentences, through read monologues or dialogues to unscripted dialogues. The 'ShATR Corpus' (Crawford et al. 1994) and the 'Nijmegen Corpus of Casual French' (Torreira et al. 2010) are good representatives of such a multiple-recordings strategy. Pilar Prieto's "Grup d'Estudias de Prosodia" (GrEP) has developed various innovative methods for the contextual elicitation of prosodic and gestural patterns. Some of these methods are summarized in Prieto (2012) and are worth considering for those who are interested in studying the many interrelations between prosody and gestures, since most traditional tasks are not suitable for this purpose.

3.3 Within-task differences

In addition to cross-task differences, there exist multifarious within-task differences. The present survey focuses on artefactual within-task differences. The examples below show the usefulness of developing an awareness of these pitfalls, in order to devise strategies to overcome them.

---

forcing speakers to choose between violating linguistic or cultural norms and questioning the authority of the experimenter. This point is further detailed in 3.1.

A typical example of an intentional within-task difference is when speakers are asked to perform the same reading task with the same material at different speaking rates. Typically, speakers are asked to read a set of isolated sentences at their normal rate and then additionally very slowly and/or as fast as they can. The sentences are either directly repeated at different rates, or the rate differences are produced blockwise. Such a within-task difference is used among others as a means to get a dynamic view of the realization and timing of intonation patterns, of speech rhythm, and of connected-speech processes such as assimilation. A frequent (and not always explicitly noted) by-product is that the F0 level is raised when speaker produce the presented sentences as fast as they can (cf. Kohler 1983; Steppling and Montgomery 2002; Schwab 2011). Interestingly, a raised F0 level is also typical of speech produced under high cognitive load or (physical or mental) stress (cf. Scherer et al. 2002; Johannes et al. 2007; Godin and Hansen 2008). It is reasonable to assume that the instruction to read and produce sentences at the fastest possible rate requires a higher cognitive load and puts speakers under stress. This requires great caution when examining data collected through deliberate speaking-rate variation: part of the phonetic differences between the speaking-rate categories set up by the experimenter may reflect differences in cognitive load and stress level. This illustrates the introductory statement that speakers are no "vending machines": the implications of instructions for within-task differences must be carefully considered.

Furthermore, fatigue and boredom can soon seep in when going through an experimental task. This can lead to unintentional within-task differences. Repetition detracts greatly from illocutionary force. This point is brought out by the thesis of Kohtz (2012). Her aim was to investigate, if and how subtypes of the common sentence-list elicitation task affect the production of nuclear accent patterns. To this end, nine speakers read two individually randomized lists of 200 sentences presented separately. The sentences in both lists ended in disyllabic sonorous target nouns produced with rising nuclear accents on the initial (lexically stressed) syllable. The target nouns in list A were embedded in the classic carrier sentence "The next word is ___". The sentences of list B had a lexically more variable NP-VP-PP structure (such as "The cat sleeps on the sofa"), in which the target nouns occurred at the end of the PP. Kohtz found consistently higher F0 variability and intensity level for list B than for list A. The speaking rate on the other hand was higher and consistently increased for the list-A than for the list-B sentences. However, the most crucial within-task difference applied to both lists and is displayed in Figure 2. The more sentences the speakers produced, the earlier and more stably aligned were the rising nuclear accents in the sentence-final target words. Only 50 sentences were already sufficient to halve the standard deviations for the alignment of rise onset and peak maximum relative to their respective segmental landmarks, i.e. the beginning of the accented syllable or its vowel. The standard deviations of the final 20 sentences were up to 85% smaller than those of the initial 20 sentences.

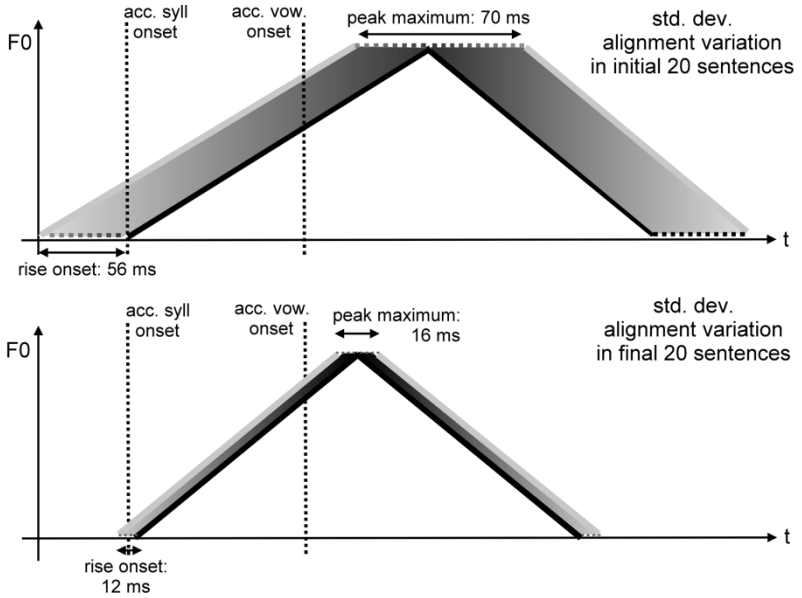

Figure 2. Schematic illustration of the fading alignment variation of F0 rise onset and peak maximum from the initial 20 to the final 20 sentences in the study of Kohtz (2012) on German intonation patterns.

Speech production is essentially a muscular task; and as for every muscular task, repetitive training in a controlled, undisturbed environment reduces variability and increases precision, efficiency, and speed. The successive changes in the nuclearaccent patterns of Kohtz must be seen in this light. Lists of 200 isolated similar sentences are an ideal training ground. It gradually reduces speech production to a mere muscular exercise and allows speakers to (unconsciously) train production and timing of their rising nuclear accents, undisturbed by syntactic and prosodic variation, communicative purposes, and interference of voiceless segments so that they can achieve an extraordinarily high level of precision and constancy.

So far, studies like that of Kohtz (2012), which critically evaluate and compare elicitation methods, are still rare. However, such studies are highly relevant to assess research results, and even to look back on the development of strands of research within phonetics/phonology. Among other implications, Kohtz's findings raise the thought-provoking issue of the extent to which the great amount of attention attracted by 'segmental anchoring' (cf. Ladd 2003) is due to the nature of the data sets under examination: segmental anchoring is primarily (and even somewhat exclusively) investigated on sets of read sentences. It is possible that the essence of segmental anchoring, i.e. an extraordinarily high level of precision and constancy in accent-contour alignment, does not show up for other kinds of elicitation tasks or for shorter lists of sentences. Initial evidence in favour of this possibility comes from the study of Welby and Loevenbruck (2006) on the segmental anchoring of rising accent contours in

French. Welby and Loevenbruck elicited a small corpus of less than 50 isolated sentences, as well as a paragraph corpus with a similarly small number of target sentences that were framed by syntactically and phonologically diverse context sentences. Although the target sentences in this procedure are equally carefully controlled as in all other studies on segmental anchoring, it is obvious that the procedure does not allow speakers to intensively train the production of their accent productions. Accordingly, the alignment patterns found by Welby and Loevenbruck showed "*a fair amount of variability […] within and across speakers […], in contrast to the very stable 'segmental anchors' found for other languages*". Moreover, "*comparisons between the two corpora also reveal intra-speaker variability […]. There were almost no significant results for a given speaker that held across the two corpora"* (Welby and Loevenbruck 2006:110). This led to the assumption that tonal alignment in French is guided by wider anchorage areas in the segmental string. However, the actual reason why Welby and Loevenbruck found anchorage areas rather than specific segmental anchor points may be a methodological one.

In conclusion, the advantages of sentence list elicitations are undeniable. They yield a high number of relevant tokens from any number of speakers in a short amount of time and with a high degree of segmental and prosodic control (cf. Figure 1). Yet, it seems fairly evident that the "instrument of spoken language" can become blunt if your list contains too many sentences. Sentences that share a similar morphosyntactic structure may further accelerate the erosive process.

4. The recording

4.1 The necessity to use professional recording equipment

On a technical note, one must emphasize the necessity of using professional recording equipment and of making and sharing sustainable recordings, as opposed to the widespread practice of recording "disposable data". It is useful to think in terms of future uses of the data, beyond one's immediate research purposes. Sampling at 16,000 Hz may seem fine for drawing spectrograms, since a display from 0 to 5,000 or 8,000 Hz is sufficient for the study of vowels, and for spectrogram reading. But if at some later date the original team of researchers (or other colleagues to whom they kindly communicate their data) wish to look at fine phonetic details, for instance allophonic/intonational variation in the realization of fricatives such as [s] and [ʃ], the sampling rate will prove too low: when separating [s] and [ʃ] on the basis of center ofgravity measurements, it was found that the best results obtained when frequencies up to 15,000 Hz were included (Niebuhr et al. 2011b). Acoustic data may also be used at some point to conduct perception tests, and a sampling rate of 44,100 Hz is designed to capture all the frequencies that the human ear can perceive. Adopting such a rate (or a higher one) therefore seems advisable, especially since digital storage of files in this format is now technically easy.

The same "the more the merrier" principle applies to data other than audio as well. Electroglottography (Fabre 1957; Abberton and Fourcin 1984) allows for highprecision measurements of the duration of glottal cycles, and for obtaining other information about glottal behaviour, such as the glottal open quotient/closed quotient. In view of the

most common uses of the electroglottographic signal, a sampling rate of 16,000 Hz would seem to be more than enough. Rather unexpectedly, this turns out to be too low for some research purposes. Research on the derivative of the electroglottographic signal brings out the significance of peaks on this signal: a strong peak at the glottis-closure instant, marking the beginning of the closed phase, and a weaker one at opening, marking the beginning of the open phase (Henrich et al. 2004). The closing peak is referred to as DECPA, for Derivative- Electroglottographic Closure Peak Amplitude (Michaud 2004b), or as PIC, for Peak Increase in Contact (Keating et al. 2010). When measuring the amplitude of this peak, signals with a sampling frequency of 16,000 Hz do not provide highly accurate information. The peak is abrupt; at a sampling rate of 16,000 Hz, widely different values are obtained depending on the points where the samples are taken at digitization. Technically, a high-precision measurement of DECPA can be obtained with a signal sampled at 44,100 Hz, with interpolation of the signal in the area of the closure peak (following a recommendation by Wolfgang Hess, p.c. 2004).

Likewise, concerning bit-depth, 24-bit may seem way too much by present-day standards, but it can still reasonably be considered, since it gives a great margin of comfort for digital amplification of portions of the signal that have extremely low volume. One bit more improves the signal-to-noise ratio by 6 dB. So, especially for recordings outside the laboratory and/or if speakers are a little further away from the microphone, as in the case of Figure 3, 24-bit should be the rule rather than the exception.

Figure 3. Recordings of narratives elicited at Mr. Vi Khăm Mun's home in Tuong Duong, Nghe An, Vietnam.

The choice of microphone is particularly critical for your recording. Compared with normal, omnidirectional microphones, head-mounted microphones or microphones with

cardioid or shotgun characteristics are usually more expensive. However, the investment pays off, as the directionality of these microphones helps keep the two channels of dialogue partners distinct, even if the dialogue partners are not physically separated, but sit face-to-face one or two meters away from each other, as is shown in Figure 4. Moreover, irrespective of whether dialogues or monologues are recorded, directional microphones help reduce environmental noise in your recordings, especially under fieldwork conditions. Head-mounted microphones have the further advantage that the distance between speaker/mouth and the microphone remains constant. Hence the intensity level is independent of the speaker's head or body movements. Microphones should always be equipped with a pop filter and pointed to the larynx rather than the mouth of the speaker, cf. Figure 4. The orientation towards the larynx has no negative effect on the recorded speech signals, but further contributes to dampen popping sounds. It is also possible to use a windshield.

Figure 4. Recording of scripted dialogues conducted in the sound-treated room of the Faculty of Engineering, Kiel University, May 2014.

Some institutions do not have high-fidelity portable devices for field workers; others only have short supplies of them, sometimes without technical staff to manage these fragile equipments, and so they do not lend them to students. Equipment is expensive. However, whenever possible, students should consider investing in their own equipment, which they will know well, and which will be available to them at any time. The purchase of recording equipment could be considered on a par with the purchase of a personal computer: while it may appear as an unreasonable demand on a student budget (especially in countries with low living standards), having one's own equipment typically increases the quality of one's data, and of the research based on them. For fieldwork, the cost of the equipment should be weighed against the overall expenses of the

field trip(s) and of the months or years of research time spent annotating and analyzing the data. In principle, speech data have an endless life expectancy, and endless reusability. Seen in this light, the common practice of recording MP3 files – with lossy compression – from a flimsy microphone (such as the internal microphone of a laptop or telephone) hardly appears as an appropriate choice.

Monitor your recording carefully, especially if you are not yet thoroughly familiar with the equipment. Prevention is better than cure; in the case of audio recordings, there is simply no way to 'de-saturate' a clipped signal in a satisfactory way, or to remove reverberation as one would brush off a layer of dust. Sound engineers have to make choices and compromises in the complex process of tidying historic music recordings; for the acquisition of new data, you should get a good signal from the start, and limit its processing to volume amplification, without special effects.

Needless to say, these remarks about speech signals can be extended to other types of data, such as video recordings.

4.2 Selection of subjects

"It is a truism but worth repeating that different informants have different talents. Some are truly excellent at explaining semantic subtleties, while others have deep intuitions about the sound structure of their language" (Dimmendaal 2001:63). For some experimental purposes, subjects with an awareness of linguistic structures or even of linguistic theory may be appropriate; for other purposes, subjects with such an awareness are best avoided. Producing spontaneous speech in the lab and producing spontaneous-sounding read speech both require a certain extroversion, fluency, language competence, and self-confidence; speakers should be pre-selected accordingly.

When dialogues are to be elicited, a deliberate selection and pairing of speakers is also important with respect to phonetic entrainment. Additionally, if the dialogue partner is a good friend, this greatly helps creating a relaxed, informal atmosphere for the recording.

Concerning the speaker sample, while four or five speakers constitute a good begining for a reasonable sample of a well-defined social group, it should be kept in mind that they do not represent the full complexity of the language at issue (in particular, its sociolectal complexity). During analysis, compare within-subject means, and – if necessary – create sub-samples before you calculate overall means for each measurement.

Let your speakers/informants fill out questionnaires that collect as detailed information (metadata) as possible – not just the three usual suspects: age, gender, and home town. Rather, type and amount of musical experience, level of education as well as smoking habits and the like should also be asked for. You could even include the question "How do you feel today?", stating explicitly that answering this and all other questions is voluntary, and that the investigator will be responsible for ensuring that these pieces of information are not made public.

For practical reasons, phoneticians often record speakers living away from the area where the target language (or dialect) is spoken. The consequences of language contact can be reduced by selecting people who have recently arrived from their homeplace, but the investigator should remain on the watchout for effects of language contact nonetheless and be additionally very specific about linguistic life course in the question-

naire. It seems safer to select subjects who experienced the smallest possible amount of language/dialect contact. There is little hope of factoring out interferences between languages when employing bilingual or multilingual speakers in phonetic studies, since the type and extent of interference varies according to numerous parameters that include age and the place of residence (Watson 2002:243).

In order to assess the expressiveness or extroversion of your speakers, we suggest making use of existing scales and questionnaires from the field of psychology (e.g., Gross and John 1997). Concerning recording settings, provide copious detail, including pieces of information that may seem anecdotal or irrelevant, such as the time of day.

4.3 Special precautions when using written prompts

Written prompts are a major source of artefacts. To linguists working on languages without a written tradition, it is obvious *that "speaking and writing are conceptually different activities, and so is a language in its spoken and written form"* (Mosel 2006:70). Linguists working on national languages may have less awareness of this central point. A study of colloquial Khmer reports that, *"in a pilot experiment, it was determined that participants had a difficult time producing colloquial variants when presented with visual primes – imagine being presented with the written sentence <I am not going to...> but being instructed to produce 'I ain't gonna...' – so instead a system was devised where the experimenter prompted the participant orally with the Standard Khmer form, whereupon the participant would provide the colloquial variant"* (Kirby 2014). Such precautions are of the greatest importance to obtain reasonably homogeneous data, otherwise the speaker's behaviour may fluctuate in unpredictable ways.

Keeping these difficulties in view, it is possible to create dialogue texts that integrate common reduction phenomena in the orthographic representation. Let your carefully selected and paired dialogue partners practice the texts in advance; allow them to adjust the texts slightly to their own way of expression by introducing, omitting or replacing words and phrases. Conceding this flexibility to speakers has proven effective to increase the comfort of speakers, which then positively affected the expressiveness and informality of their way of speaking (Kohler and Niebuhr 2007; Niebuhr 2010, 2012). A further means to control the way of speaking is the font type of the written prompts. According to the experience of the first author, expressiveness and informality are best elicited with font types other than the businesslike Times, Arial, Calibri and Tahoma fonts.

Be careful with translations. Translating experimental materials and instructions requires all the precautions usually associated with translation, which is a profession on its own. For example, the second author was asked on several occasions times to read question-answer pairs in French such as "Qui est allé au restaurant? – Jean est allé au restaurant", which had obviously been translated from English ("Who went to the restaurant? – John went to the restaurant"). Question-answer pairs like the one above aim at eliciting narrow focus. They aim to elicit a realization in which the name "Jean" stands out as the informative part (emphasis) whereas the rest of the sentence is backgrounded (post-focus compression). Whatever the validity of the original English, its translation sounds decidedly weird to native speakers of French: more appropriate answers would be simply "Jean", or "c'est Jean" ("it is Jean"), or – at a push – the cleft

sentence “C’est Jean qui est allé au restaurant”. Bad translation also ignores cultural factors. For example, the sentence “He decided to move house, but not to leave the town” may make good sense to speakers of American English, but it becomes odd when translated and used for French speakers, since the population density and urban structure of France is very different from that of the United States. Practical and detailed instructions on the use of translations in fieldwork are provided by Mosel (2011) and references therein.

Furthermore, monotonous tasks are to be avoided when using written prompts. In order to be comparable, utterances do not only need to be identical in their written form: *“most importantly, they have to be performed with the intention of achieving the same illocutionary act”* (Himmelmann 2006:168). Especially in experiments with a larger number of cross-combined independent variables it is often necessary to elicit numerous targets (e.g., words) with specific phonetic properties in prosodically controlled environments. In addition to artefacts in the form of ‘list intonation’, readers may spontaneously establish semantic/pragmatic relations between individual sentences, interpreting them as successive episodes within a single narrative, as it were – even if these sentences are presented on separate sheets of paper, or on separate slides shown on a computer screen. For example, the sequence “Peter came by car” - “Meghan came by bus” - “Steve came by boat” may cause “bus” and “boat” to be realized with prosodies of contrastive topic. In the sequence “The plate is on the table” - “The glass is on the table”, “table” becomes given information, and “glass” is likely to be realized in contrast to “plate”. Unless this is controlled for at the stage of data collection, wavering interpretations of the recording task will be treated as random variance at the stage of statistical analysis.

As the largest prosodic changes seem to occur after the fiftieth sentence (see 3.3), it seems safe to use lists of less than fifty sentences per session. Repetitions of the same sentence within the same session is to be avoided – or at least the investigators should be aware of the potential bias introduced by this repetition.

Isolated syllables should be carefully randomized, to avoid contrast effects similar to those described above. A sequence of syllables arranged by vowel, such as “ta, ma, ba, da, na, pa, ra…” will lead to more attention being focused on the realization of the consonant than on that of the vowel; and the opposite bias will be present for sequences arranged by consonant: “ta, tu, to, ti…”.

One way to limit such prosodic artefacts consists in interspersing dummy sentences in the sequence of sentences to be recorded, so that successive sentences will be as unrelated as possible. But this increases the length of the task, and hence the ever-present risk of fatigue, without providing any guarantee that the speaker will not invent links between successive sentences. Using dialogues appears a more powerful solution.

4.4 Training, dummy-runs, and debriefing techniques

Training of the subjects needs to be handled with care. A widely-cited article about “Intonational invariance under changes in pitch range and length” is based on an experiment in which *“...the pitch range instruction was varied in 10 steps, and six to eight repetitions of each pattern in each pitch range were recorded. In both of the experiments to be described, ‘degree of overall emphasis or excitement’ was the term*

used in the subjects' instructions, and the kind of variation desired was illustrated by example" (Liberman and Pierrehumbert 1984:169). Four speakers were recorded, including the two coauthors. *"For subjects other than the authors, the desired intonation patterns were demonstrated by example before the experiment, and the ability of the subjects to produce them naturally was checked"* (p. 172). This formulation exemplifies the problems mentioned above (cf. 3.1) concerning the notion of naturalness. Additionally, one may have a couple of minor quibbles about data collection here, concerning (i) the use of a metalinguistic indication, "degree of overall emphasis or excitement", allowing for a broad range of interpretations, and (ii) the example set by the co-authors for the other two speakers. These go a long way towards explaining, why the admirably clear-cut result obtained in that study – namely, that the terminal point of the F0 curve remains almost unchanged, whereas the highest F0 value is proportional to the degree of emphasis – could not be replicated in a later study (Nolan 1995).

Such salient artefacts are sometimes identified clearly by the community of phoneticians, leading to the adoption of new principles: it would not currently be considered good practice for linguists to report analyses based on their own speech data. This general principle is useful, but should be complemented by investigators as befits each specific experimental setup. An ideal towards which it may be useful to tend would consist in making laboratory experiments a mutual learning and teaching process for all people involved – like linguistic fieldwork. Dummy-runs (or a warm-up time for "spontaneous" – i.e. unscripted – conversations) and training can allow for this communicative process to take place. It is for the investigator to make adjustments and preparations carefully before the recording starts, in order not to distract the subjects' attention during the experiment. For instance, if the data are to be used in fine-grained acoustic analysis, it may be useful to instruct the speakers to avoid shuffling their feet or rubbing their hands on their clothes too vigorously during recording; but if these gestures are habitual for the speaker, making conscious attempts to suppress them takes up part of the speaker's attention, and may have consequences such as an increased amount of disfluencies. *"The importance of good recording needs to be balanced against the importance of keeping the participants at ease"* (Souag 2011:66).

Debriefing is a useful (though currently nonstandard) way of finding out more about the subjects' interpretation of the task and the evolution of their behaviour in the course of the session. Participants may sometimes point out accidental omissions: a speaker of French recruited for a recording of nasal vowels became aware of the absence of any example of efg/, which in his speech contrasted with /hgeijekgejlmnjeog/. This would not have been detectable on the basis of the recordings, and inclusion of data from this speaker would have detracted from the reliability of the results, since the study assumed a three-way contrast between /hgeijekgejlmnjeog/. In retrospect, this aspect of the phonological system should have been examined individually for each potential participant before they were selected to conduct recordings.

Debriefing plays a central role in the 'Kieler Sammlung Expressiver Lesesprache' (KIESEL Corpus, i.e. 'Kiel Collection of Expressive Read Speech', presented in Niebuhr 2010). The aim is to achieve a high degree of expressiveness in read speech. Read speech allows for segmental and prosodic control of the data; expressiveness is approached by instructing the speakers to judge each other's production performances, and repeating the dialogue until they are both satisfied and agree that they have

produced a dialogue that resembles their colloquial, everyday speaking style. This setup yields encouraging results, despite its limitations.

4.5 Minimizing the recorder's paradox

Every recording situation will inevitably raise the speakers' awareness about the way they speak. This, in turn, can make speakers change their speech behaviour, with the consequence that the analyzable object diverges from the actual research object. Our goal is to understand speech communication, and speech recordings make it at the same time easier and harder to reach this goal. Xu (2012) calls this the "recorder's paradox". In order to reduce influences of the recording situation, you may use headmounted microphones – which, unlike table microphones, are not present in the speaker's field of vision –, select experienced speakers, use dialogue rather then monologue scenarios (pairs of speakers distract each other more easily from the recording situation), hide or store away technical equipment, and avoid dark or darkcoloured recording rooms. For research questions that require a high degree of selfrevelation from the speakers, as in the case of expressive or dialectal speech or when small children are to be recorded, it can sometimes be even better to conduct the recordings in silent rooms of the speakers' own homes, as can be seen in Figure 5. Suitable rooms have as few plane and sound-reflecting surfaces as possible (e.g., bed rooms or living rooms). The remaining plane and sound-reflecting surfaces can be covered with bed sheets, large towels or similar pieces of household linen (cf. Fig.7). Bookcases and bookshelves also improve a room's acoustic qualities.

Figure 5. Speakers of the endangered North Frisian dialect Fering produce isolated sentences and scripted dialogues; recordings were conducted in the speakers' homes on the small island Föhr in the Northern Sea off the German coastline, cf. Niebuhr and Hoekstra (2014).

In fieldwork in small villages, in typically less-developed area, it is uncommon to have access to a room that is padded with bookshelves or soft furniture. Bare rooms with cement or tile flooring are common, as exemplified in Figure 6(a) by the kitchen of a Naxi farm in Yunnan, China. In such a room, there is an amount of reverberation that makes it difficult to make out on a spectrogram to what extent an intervocalic stop is voiced: the complete silence during the closed phase of an unvoiced stop is masked by reverberation, which results in noise on spectrograms even at points where one would expect complete silence. Doing a recording out in the open is hardly an option: a field or a pasture are acoustically fine, as there is close to zero reverberation, but the speaker and the investigator are then exposed to the elements – including the wind, which can ruin a recording if no windscreen is available. The comical scene of speech data acquisition is also mercilessly exposed to the gaze of passers-by and the curiosity of wandering animals, making it pretty hopeless to achieve the required degree of concentration on the part of all concerned.

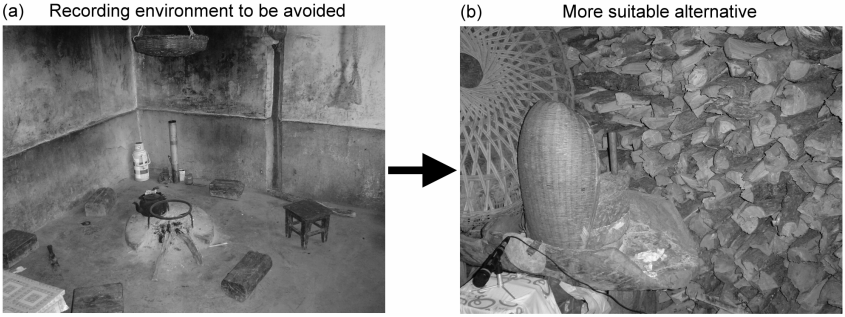

Figure 6. (a) shows an unsuitable recording room, a kitchen of a Naxi farm in Yunnan, China; (b) shows the more suitable recording environment on the farm, which was chosen instead.

Figure 6(b) shows the location that was chosen for recording inside the Yunnanese farm: in the courtyard, behind a thick stack of firewood which absorbs reverberation. This spot is located under a porch-roof, whose uneven tiling provides protection from the rain without creating strong reverberation. Large utensils of wood and stone partly cover the cement floor (a grindstone to make bean curd is seen on the photo, behind the microphone). When doing a recording, a piece of thick, rough cloth is propped across the open end of this makeshift recording booth, both contributing (at least minimal) acoustic improvement and providing a signal to the family members that noises are to be kept to a minimum, cf. Figure 7. The environment noises (chirping of birds and occasionally distant sounds of dogs barking, people shouting, or the rumble of an engine) are not a real issue for acoustic analysis, whereas reverberation is a major problem. Needless to say, any such changes in layout at someone's home needs to be discussed with one's hosts; good communication with one's consultants is absolutely essential.

Figure 7. Further optimation of the recording environment. A piece of thick, rough cloth is used to cover sound-reflecting surfaces and open space, which also dampens background noise.

4.6 Placing one's data within a broader context

Constant carrier sentences like "I don't know the word ___"; "The next word is ___"; or "I have seen ___ on the table" allow for a maximum degree of control, necessary for experimental and statistical purposes. It should be borne in mind that they represent a strong abstraction from everyday communication, however, so that one should be careful when attempting extrapolations. No set of data is complete in itself; no experiment provides a knock-out argument. Each phonetic data set represents a compromise between the competing demands of symmetry, on the one hand, and breadth of scope, on the other. One researcher, or even one team of researchers, cannot hope to gather an all-encompassing set of data.

One's data should therefore be placed within a broader context, and seen as a contribution to a broader set, whose gradual constitution can only be a collaborative endeavour. This constitutes a vision for the future of phonetics in the digital era, where the issue of data storage and accessibility looks very different from the predigital era.

Initial promising steps to a collaborative creation and use of annotated speech corpora and databases at an international level have already been made. Prominent representatives of this endeavour are the EU-funded 'CLARIN' network (Common Language Resources and Technology Infrastructure, Váradi et al. 2008), the 'Reciprosody', i.e. a repository for prosodically oriented and annotated speech corpora (founded by A. Rosenberg, http://jaguar.cs.qc.cuny.edu/), or archives for endangered languages data such as the 'DoBeS Archive' hosted by the Max-Planck-Institute for Psycholinguistics in Nijmegen (Drude et al. 2012) and the 'Pangloss Collection' at CNRS (Michailovsky et al. 2014). The advice provided in this paper, particularly concerning the need to use professional recording equipment and to collect detailed metadata, must

also be seen in the light of such international collaborations, which all put minimum (and in tendency increasing) demands on hosted corpora.

5. Summary

Recordings have to be well planned, tested, and should not be conducted with the first available equipment. In short: Do not underestimate the challenge of speech data acquisition. Do not take recordings lightly! Many potential issues can be anticipated, managed, and controlled. This final section summarizes tips and recommendations on how to meet the demands of specific research questions and achieve results of lasting value for the scientific community.

5.1 The speaker

- (a) Do not select just any available speaker: screen your speakers carefully with respect to how they fit in with your research question (e.g., language skills, expressiveness), and in order to create a homogeneous sample (e.g., age, gender, smoking habits, musical experience)

- (b) If you record dialogues, either control phonetic entrainment or exploit it to implicitly change the speech patterns you get into a certain direction.

- (c) Collect a comprehensive set of metadata from your speakers; put special emphasis on linguistic experience and personality characteristics/habits.

- (d) Your sample size should be large enough to allow looking for individual preferences and between-group differences, for example, with respect to gender, dialectal background, and daytime. This probably means recruiting at least 10 speakers.

- (e) When you prepare the recording environment, the elicitation materials, and the task instructions, keep in mind that speakers are no “vending machines”, which produce representative speech by paying and pressing a button, and that speech is essentially a social phenomenon. Treat your speakers with respect and bond with them.

- (f) For practical reasons, phoneticians often record speakers living away from the area where the target language (or dialect) is spoken. The consequences of language contact can be reduced by selecting people who have recently arrived from their homeplace, but you should remain on the watchout for effects of language contact nonetheless.

5.2 The task

- (a) Be as detailed as you can in the instructions, and use the same instructions for all speakers. So, prepare them as a sound file or in written form.

- (b) Make sure that the semantic-pragmatic context in which you embed the elicitation task is as rich and specific as possible. Don't be afraid to be creative and integrate multi-medial resources and/or aspects of everyday life. The richer, clearer, and more specific the elicitation context is, the more homogeneous, replicable, and valid your speech material will be.

- (c) Include a debriefing during or after the recording session and take the feedback of your speakers seriously. In terms of everyday communication, they are no less an expert than you, so do not force them to produce utterances they reject, for example, because of their wording or grammar.

- (d) Be extremely careful when translating speech material or instructions from a different study.

- (e) Conduct pilot recordings to test your material and instructions, and don't use the same speakers for the following main recording.

- (f) If you need initial data to study the basic characteristics of a new field of segmental or prosodic phenomena, it is advisable to make use of the high event density and control offered by read speech, i.e. elicit isolated sentences or read monologues.

- (g) If, on the other hand, you would like to go into the details of a wellunderstood segmental or prosodic phenomenon, or if you are interested in the phonetic exponents of discourse functions, make use of functionally rich and ecologically more valid dialogue tasks, i.e. elicit scripted or unscripted dialogues.

- (h) If you are not sure whether (g) applies to your research topic, the best solution is always to elicit different types of speech materials, for example a combination of read sentences and unscripted dialogues.

- (i) Role-play, quiz, and instruction-giving tasks allow eliciting target words with reasonably high frequencies even in unscripted speech. Similarly, read dialogues also strike a reasonable compromise between event density, experimental control, expressiveness, and ecological validity.

- (j) In dialogues, you can also use a privy dialogue partner to channel the speech behaviour of your speakers towards a certain direction.

- (k) If you elicit monologues – be they read or spontaneous – give your speakers somebody to address. Even if s/he does not say anything, it will make your speaker feel more comfortable and increase the quality, diversity, and validity of your speech materials.

- (l) If your recording session involves two subjects, use good friends in order to create a more informal atmosphere during the recording session.

- (m) Avoid monotonous tasks. For example, limit sentence lists to no more than 50 tokens; randomize sentences. Avoid repetitions of the same sentence (structure); and if this is not possible, insert dummy sentences that clearly deviate from your target sentences.

5.3 The recording

- (a) Take time to look for a suitable recording environment, if you make recordings outside the laboratory, and further optimize the environment by reducing open space, background noise, and reverberant surfaces, if possible.

- (b) Use professional recording equipment, and digitalise your speech with 44.1 kHz and 16-bit or higher recording settings. Monitor your recording carefully, especially if you are not yet thoroughly familiar with the equipment.

- (c) Allow speakers to familiarize themselves with the recording situation. That is, include a warm-up task – for example, let the speaker summarize the previous weekend/dinner or ask his/her hobbies – before you start with the actual recordings. Use the warm-up phase for adjusting the recording level in order to avoid that sections of the actual recording are distorted by clipping, cf. (b) above.

- (d) Use a head-mounted microphone or a microphone with cardioid or shotgun characteristic to reduce environmental noise in your recordings. Head-mounted microphones have the further advantage that the distance between speaker/mouth and the microphone remains constant. Hence the intensity level is independent of head or body movements and becomes an analyzable acoustic parameter.

- (e) Hide away technical equipment and other things that have the potential to intimidate or distract your speakers.

- (f) Do not record “disposable data”. Make your data available to the scientific community by depositing them in institutional repositories that ensure their long-term preservation and access. This requires prior design of forms to be signed by the speakers, to indicate their informed consent to participate in the experiment and to give copyrights (in certain fieldwork settings, oral consent can be substituted as appropriate). ‘CreativeCommons’ licences have many advantages for data sharing in scientific research.

6. Conclusion

As a speech scientist, you record data as you think fit, in view of your immediate research purposes. The above review suggests that you stand to gain a lot by considering a range of options at each of the three main stages of speech data collection: different types of procedures offer different insights, and their combination yields an in-depth, well-rounded view of speech. Time constraints and tight deadlines make it appear unreasonable to lavish your time on seemingly preliminary tasks such as contextualizing data and exchanging at leisure with your consultants before and after experiments. But the time spent on preparing data collection is in fact wellinvested, yielding considerable benefits for research. You will get a handle on major sources of variability, instead of unwittingly leaving important parameters uncontrolled and treating the ensuing variability as random. Painstaking data collection makes for reliable and enduring documents, which can profitably be shared – not only re-used, but also enriched collaboratively. In this optimistic perspective, data collection (language documentation) and research can progress hand in hand, allowing for a cumulative approach to research in the phonetic sciences.

We are well aware that the present article, which is essentially intended to provide some practical suggestions, only scratches the surface of data collection methodology. Further work would require scrutinizing and comparing the full range of recording tasks, conditions, and instructions on the basis of systematic experimental studies. Pending such in-depth work, our provisional morality is that perfecting elicitation methods requires keeping a constant eye on function, meaning, and the individual; this holds true of all types of research in phonetics/phonology, over and above the great diversity of research goals and methods.

7. Acknowledgments

Many thanks to Jacqueline Vaissière for useful comments. Further thanks are due to Gu Wentao and Klaus Kohler, and two anonymous reviewers for their constructive comments on an earlier draft of this paper. Needless to say, they are not to be held responsible for the views expressed here. We are also indebted to Sarah Buchberger and Jana Bahrens for helping us with formatting the paper and corss-checking the references. Support from Agence Nationale de la Recherche (HimalCo project, ANR-12-CORP-0006) and from LabEx “Empirical Foundations of Linguistics” (EFL) is gratefully acknowledged.

8. References

ABBERTON, E. / A.J. FOURCIN. 1984. Electroglottography. Experimental Clinical Phonetics, 62–78.

ABRAMSON, A.S. 1972. Tonal Experiments with Whispered Thai. In: A. Valdman (Ed.), Papers on Linguistics and Phonetics in Memory of Pierre Delattre (pp. 31–44). The Hague: Mouton.

AMBRAZAITIS, G. 2005. Between Fall and Fall-Rise: Substance-function Relations in German Phrase-final Intonation Contours. Phonetica 62 (2-4), 196–214.

AMDAL, I. / T. SVENDSEN. FonDat1: A Speech Synthesis Corpus for Norwegian. Proc. 5th International Conference on Language Resources and Evaluation, Genova, Italy, 2096-2101.

AMDAL, I. / O. STRAND / J. ALMBERG / T. SVENDSEN. 2008. RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus. Proc. 5th International Conference on Language Resources and Evaluation, Marrakech, Morocco, 1907-1913.

ANDERSON, A.H. / M. BADER / E. GURNAN BARD / E. BOYLE / G. DOHERTY / S. GARROD / S. ISARD / et al. 1991. The HCRC Map Task Corpus. Language and Speech 34, 351–366.

ANDERWALD, L. 2014. You Just Don't Understand – Nichtverstehen zwischen Männern und Frauen. In: O. Niebuhr (Ed.), Formen des Nicht-Verstehens (pp.113-128). Frankfurt: Peter Lang.

ANDREEVA, B. / W. BARRY. 2012. Fine phonetic detail in prosody. Cross-language differences need not inhibit communication. In: O. Niebuhr (Ed.), Prosodies - context, function, and communication (pp. 259-288). Berlin/New York: de Gruyter.

BARBOSA, P.A. 2012. Panorama of Experimental Prosody Research. Proc. Gruppo di Studi sulla Comunicazione Parlata Workshop, Belo Horizonte, Brazil, 33-42.

BARNLUND, D.C. 2008. A transactional model of communication. In: C. D. Mortensen (Ed.), Communication theory (pp. 47-57). New Brunswick, New Jersey: Transaction.

BARNES, J. / A. BRUGOS / E. ROSENSETIN / S. SHATTUCK-HUFNAGEL / N. VEILLEUX. 2013. Segmental sources of variation in the timing of American English pitch accents. Paper presented at the annual meeting of the Linguistic Society of America, Boston, USA.

BERTRAND, R. / P. BLACHE / R. ESPESSER / G. FERRE / C. MEUNIER / B. PRIEGO-VALVERDE / S. RAUZY. 2008. Le cid-corpus of interactional data-annotation et exploitation multimodale de parole conversationnelle. Traitement Automatique des Langues 49, 1–30.

BRAUN, B. / D.R. LADD. 2003. Prosodic correlates of contrastive and non-contrastive themes in German. Proc. 8th Eurospeech Conference, Geneva, Switzerland, 789-792.

BRUNELLE, M. 2012. Dialect Experience and Perceptual Integrality in Phonological Registers: Fundamental Frequency, Voice Quality and the First Formant in Cham. Journal of the Acoustical Society of America 131, 3088–3102.

BÜHLER, K. 1934. Sprachtheorie. Die Darstellungsfunktion Der Sprache. Jena: Gustav Fischer.

BURKARD, T. 2014. Mythen und freie Erfindungen in der lateinischen Grammatik – Das Nicht-Verstehen einer toten Sprache und seine Konsequenzen. In: O. Niebuhr (Ed.), Formen des Nicht-Verstehens (pp.45-76). Frankfurt: Peter Lang.

CANGEMI, F. 2013. Listener-specific perception of speaker-specific productions? Evidence from intonation and supralaryngeal articulation across focus structures in German. Proc. 4th Summerschool on Speech Production and Perception: Speaker-Specific Behaviour, Aix-en-Provence, France, 16-17.

CAMPBELL, N. / P. MOKHTARI. 2003. Voice Quality: The 4th Prosodic Dimension. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2417–2420.

CLARKE, C.M. / M.F. GARRETT. 2004. Rapid Adaptation to Foreign-accented English. Journal of the Acoustical Society of America 116, 3647–3658.

COOKE, M. / O. SCHARENBORG. 2008. The Interspeech 2008 Consonant Challenge. Proc. 7th Interspeech Conference, Brisbane, Australia, 1-4.

CRAWFORD, M.D. / G.J. BROWN / M.P. COOKE / P.D. GREEN. 1994. Design, collection and analysis of a multisimultaneous- speaker corpus. Inst. Acoustics 16, 183-190.

CULIOLI, A. 1995. Cognition and Representation in Linguistic Theory. Current Issues in Linguistic Theory. Amsterdam: John Benjamins.

DELVAUX, V. / A. SOQUET. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 64, 145-173.

DUNG, D.T / T.H. TRÂN / G. BOULAKIA. 1998. Intonation in Vietnamese. In: D. Hirst & Albert Di Cristo (Eds), Intonation Systems: A Survey of Twenty Languages (pp. 295-416). Cambridge: Cambridge University Press

DOMBROWSKI, E. 2003. Semantic features of accent contours: effects of F0 peak position and F0 time shape. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 1217-1220.

DIMMENDAAL, G. J. 2001. Places and People: Field Sites and Informants. In: P. Newman & M. Ratliff (Eds.), Linguistic Fieldwork (pp. 55-75). Cambridge: Cambridge University Press.

DRUDE, S. / P. TRILSBEEK / D. BROEDER. 2012. Language Documentation and Digital Humanities: The (DoBeS) Language Archive. Proc. International Conference of Digital Humanities, Hamburg, Germany, 169-173.

EDMONDSON, J.A. / K.J. GREGERSON. 1993. Western Austronesian Languages. In: J.A. Edmondson & K.J. Gregerson (Eds), Tonality in Austronesian Languages (pp. 61-74). Honolulu: University of Hawai'i Press.

ELLIS, L. / W.J. HARDCASTLE. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. Journal of Phonetics 30, 373- 396.

ERNESTUS, M. 2000. Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface. Utrecht: LOT.

FABRE, P. 1957. Un Procédé Électrique Percutané D'inscription De L'accolement Glottique Au Cours De La Phonation: Glottographie De Haute Fréquence. Bulletin De l'Académie Nationale De Médecine 141, 66–69.

FEDERMAN, J. 2011. Effects of musical training on speech understanding in noise. PhD Dissertation, Vanderbilt University, Nashville, USA.

FERLUS, M. 1979. Formation Des Registres Et Mutations Consonantiques Dans Les Langues Mon-khmer. Mon- Khmer Studies 8, 1–76.

FITZPATRICK, M. / J. KIM / C. DAVIS. 2011. The effect of seeing the interlocutor on auditory and visual speech production in noise. Proc. 11th International Conference on Auditory-Visual Speech Processing, Volterra, Italy, 31–35.

FON, J. 2006. Shape Display: Task Design and Corpus Collection. Proc. 3rd International Conference of Speech Prosody, Dresden, Germany, 181-184.

FÓNAGY, I. 2001. Languages Within Language: An Evolutive Approach. Foundations of Semiotics 13. Amsterdam/Philadelphia: Benjamins.

FOURAKIS, M. / G.K. IVERSON. 1984. On the 'Incomplete Neutralization' of German Final Obstruents. Phonetica 41, 140–149.

FRANCOIS, A. 2012. The Dynamics of Linguistic Diversity: Egalitarian Multilingualism and Power Imbalance Among Northern Vanuatu Languages. International Journal of the Sociology of Language 214, 85–110.

GARTENBERG, R. / C. PANZLAFF-REUTER. 1991. Production and Perception of F0 Peak Patterns in German. Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung der Universität Kiel (AIPUK) 25, 29–113.

GENDROT, C. / M. ADDA-DECKER. 2007. Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. Proc. 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 1417–1420.

GENZEL, S. / F. KÜGLER. 2010. The prosodic expression of contrast in Hindi. Proc. 5th International Conference of Speech Prosody, Chicago, USA, 1-4.

GILES, H. / N. COUPLAND. 1991. Language: Contexts and Consequences. Mapping Social Psychology. Belmont: Thomson Brooks/Cole Publishing Co.

GODIN, K.W. / J.H.L. HANSEN. 2008. Analysis and perception of speech under physical task stress. Proc. 8th Interspeech Conference, Brisbane, Australia, 1674-1677.

GÖRS, K. 2011. Von früh bis spät - Phonetische Veränderungen der Sprechstimme im Tagesverlauf. BA thesis, Kiel University, Germany.

GÖRS, K. / O. NIEBUHR. 2012. Hocus Focus - What the Eliciation Method Tells Us About Types and Exponents of Contrastive Focus. Proc. 6th International Conference of Speech Prosody, Shanghai, China, 262–265.

GOW, D. 2003. Feature parsing: Feature cue mapping in spoken word recognition. Perception and Psychophysics 65, 575-590.

GRAUPE, E. 2014. Zusammenhänge von Gesangserfahrung und Stimmklang - physiologische, akustische und perzeptorische Analysen. Kalipho 2, this volume.

GRINEVALD, C. 2006. Worrying about ethics and wondering about “informed consent”: Fieldwork from an Americanist perspective. In: A. Saxena & L. Borin (Eds), Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology (pp. 339-370). Berlin: De Gruyter.

GROSS, J.J. / O.P. JOHN. 1997. Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. Journal of Personality and Social Psychology 72, 435-448.

GUENTHER, F.H. / C.Y. EPSY-WILSON / S.E. BOYCE / M.L. MATTHIES / M. ZANDIPOUR / J.S. PERKELL. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. Journal of the Acoustical Society of America 105, 2854-2865.

GUSSENHOVEN, C. 2004. The Phonology of Tone and Intonation. Cambridge: CUP.

HAAS, M. 1944. Men’s and Women’s Speech in Koasati. Language 20 (3): 142 149.

HAUDRICOURT, A.-G. 1961. Richesse En Phonèmes Et Richesse En Locuteurs. L’Homme 1, 5–10.

HENDERSON, E.J.A. 1952. The Main Features of Cambodian Pronunciation. Bulletin of the School of Oriental and African Studies 14, 149–174.

HELDNER, M. / J. EDLUND / J. HIRSCHBERG. 2010. Pitch similarity in the vicinity of backchannels. Proc. 11th Interspeech Conferece, Makuhari, Japan, 3054-3057.

HENRICH, N., C. D'ALESSANDRO, M. CASTELLENGO, and B. DOVAL. 2004. On the Use of the Derivative of Electroglottographic Signals for Characterization of Non-pathological Voice Phonation. Journal of the Acoustical Society of America 115, 1321–1332.

HERMES, A. / J. BECKER / D. MÜCKE / S. BAUMANN / M. GRICE. 2008. Articulatory gestures and focus marking in German. Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 457-460.

HIMMELMANN, N. 2006. Prosody in Language Documentation. In: J. Gippert, N.P. Himmelmann & U. Mosel (Eds), Essentials of Language Documentation (pp. 163-181). Berlin/New York: de Gruyter.

HIRSCHBERG, J. 2011. Speaking More Like You: Entrainment in Conversational Speech. Proc. 12th Interspeech Conference, Florence, Italy.

HIRST, D. / A. DI CRISTO. 1998. A Survey of Intonation Systems. In: D. Hirst & Albert Di Cristo (Eds), Intonation Systems: A Survey of Twenty Languages (pp. 1–43). Cambridge: Cambridge University Press.

HOOLE, P. 1999. On the lingual organization of the German vowel system. Journal of the Acoustical Society of America 106, 1020–1032.

HOUSE, J. 1989. Syllable structure constraints on F0 timing. Poster presented at LabPhon II, Edinburgh, Scotland.

HUALDE, J.I. 2003. Remarks on the diachronic reconstruction of intonational patterns in Romance with special attention to Occitan as a bridge language. Catalan Journal of Linguistics 2, 181–205.

HUFFMAN, F.E. 1976. The Register Problem in Fifteen Mon-Khmer Languages. Austroasiatic Studies. In: P.N. Jenner, L.C. Thompson & S. Starosta (Eds), Oceanic Linguistics Special Publication No 13 (pp. 575-589). Honolulu: Hawaii University Press.

ITO, K. / S.R. SPEER. 2006. Using interactive tasks to elicit natural dialogue. In: S. Sudhoff et al. (Eds), Methods in empirical prosody research (pp. 229-258). Berlin/ New York: de Gruyter.

IWASHITA, N. / T. MCNAMARA / C. ELDER. 2001. Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. Language Learning 21, 401-436.

JANSE, E. / P. ADANK. 2012. Predicting Foreign-accent Adaptation in Older Adults. Quarterly Journal of Experimental Psychology 65, 1563–1585.

JASSEM, W. / L. RICHTER. 1989. Neutralization of Voicing in Polish Obstruents. Journal of Phonetics 17, 317–325.

JOHANNES, B. / P. WITTELS / R. ENNE / G. EISINGER / C.A. CASTRO / J.L. THOMAS / A.B. ADLER / R. GERZER. 2007. Non-linear function model of voice pitch dependency on physical and mental load. European Journal of Applied Physiology 101, 267–276.

JOHNSON, K. / P. LADEFOGED / M. LINDAU. 1993. Individual differences in vowel production. Jorunal of the Acoustical Society of America 94, 701-714.

JOHNSTONE, T. / C.M. VAN REEKUM / K. HIRD / K. KISNER / K. SCHERER. 2005. Affective speech elicited with a computer game. Emotion 5, 513-518.

JUN, S-A. / J. FLETCHER. 2014. Methodology of Studying Intonation: From Data Collection to Data Analysis. In: S.-A. Jun (Ed.), Prosodic Typology II: New Developments in the Phonology of Intonation and Phrasing (pp. 493-519). Oxford: Oxford University Press.

KEATING, P. / C. ESPOSITO / M. GARELLEK / S. UD DOWLA KHAN / J. KUANG. 2010. Phonation Contrasts Across Languages. UCLA Working Papers in Phonetics 108, 188–202.

KIRBY, J. 2014. Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. Journal of Phonetics 43. 69–85.

KIM, M. 2012. Phonetic Accommodation after Auditory Exposure to Native and Nonnative Speech. PhD thesis, Northwestern University, IL, USA.

KLEBER, F. / T. JOHN / J. HARRINGTON. 2010. The Implications for Speech Perception of Incomplete Neutralization of Final Devoicing in German. Journal of Phonetics 38, 185–196.

KOHLER, K.J. 1983. F0 in speech timing. AIPUK 20, 55-97.

KOHLER, K.J. 1990. Macro and micro F0 in the synthesis of intonation. In: J. Kingston & M.E. Beckman (Eds), Papers in Laboratory Phonology I (pp. 115-138). Cambridge: Cambridge University Press.

KOHLER, K.J. 1991. A model of German intonation. AIPUK 25, 295-360.

KOHLER, K.J. 1995. Einführung in die Phonetik des Deutschen. Berlin: Erich Schmidt.

KOHLER, K.J. 2004. Categorical speech perception revisited. Proc. of the Conference "From Sound to Sense: 50+ years of discoveries in speech communication", MIT Cambridge, USA, 1-6.

KOHLER, K.J. 2006. Paradigms in experimental prosodic analysis: From measurement to function. In S. Sudhoff et al. (Eds.), Methods in empirical prosody research (pp. 123-152). Berin/New York: de Gruyter.

KOHLER, K.J. / O. NIEBUHR. 2007. The phonetics of emphasis. Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken, Germany, 2145-2148.

KOHLER, K.J. / O. NIEBUHR. 2011. On the Role of Articulatory Prosodies in German Message Decoding. Phonetica 68, 1–31.

KOHTZ, L.-S. 2012. Datenerhebung mittels Leselisten - Eine kritische phonetische Evaluation. BA thesis, Kiel University, Germany.

KÜHNERT, B. / P. HOOLE. 2004. Speaker-specific kinematic properties of alveolar reductions in English and German. Clinical Linguistics and Phonetics 18, 559-575.

LADD, D.R. 2003. Phonological conditioning of F0 target alignment. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 249-252.

LANDGRAF, R. 2014. Linguistische Charakteristika von Dialogen im fahrenden Auto und wie sie sich im Labor simulieren lassen. MA thesis, Kiel University, Germany.

LAVER, J. 1994. Principles of Phonetics. Cambridge: Cambridge University Press.

LEE, C. / M. BLACK / A. KATSAMANIS / A. LAMMERT / B. BAUCOM / A. CHRISTENSEN / P. GEORGIOU / S. NARAYANAN. 2010. Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples. Proc. 11th Interspeech Conference, Makuhari, Japan, 793–796.

LEVITAN, R. / J. HIRSCHBERG. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. Proc. 12th Interspeech Conference, Florence, Italy, 1-4.

LIBERMAN, M. / J. PIERREHUMBERT. 1984. Intonational Invariance Under Changes in Pitch Range and Length. In R.T. Oehrle & M. Aronoff (Eds), Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students (pp. 157–233). Cambridge: MIT Press.

MAFFIA, M. / E. PELLEGRINO / M. PETTORINO. 2014. Labeling expressive speech in L2 Italian: the role of prosody in auto-and external annotation. Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 81-85.

MANSON, J.H. / A. GREGORY / A. BRYANT / M.M. GERVAIS / M.A. KLINE. 2013. Convergence of speech rate in conversation predicts cooperation. Evolution and Human Behavior 34, 419–426.

MARCHAL, A. 2009. From Speech Physiology to Linguistic Phonetics. London: Wiley.

MICHAILOVSKY, B. / M. MAZAUDON / A. MICHAUD / S. GUILLAUME / A. FRANCOIS / E. ADAMOU. 2014. Documenting and researching endangered languages: the Pangloss Collection. Language Documentation and Conservation 8, 119–135.

MICHAUD, A. 2004a. Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. Phonetica 61, 119–146.

MICHAUD, A. 2004b. A Measurement from Electroglottography: DECPA, and Its Application in Prosody. Proc. 2nd International Conference of Speech Prosody, Nara, Japan, 633–636.

MICHAUD, A. 2012. Monosyllabicization: Patterns of Evolution in Asian Languages. In: N. Nau, T. Stolz & C. Stroh (Eds), Monosyllables: From Phonology to Typology (pp. 115–130). Berlin: Akademie Verlag.

MICHAUD, A. / T. VU-NGOC / A. AMELOT / B. ROUBEAU. 2006. Nasal Release, Nasal Finals and Tonal Contrasts in Hanoi Vietnamese: An Aerodynamic Experiment. Mon-Khmer Studies 36, 121–137.

MITHUN, M. 2001. Who Shapes the Record: The Speaker and the Linguist. In: P. Newman & M. Ratliff (Eds), Linguistic Fieldwork (pp. 34–54). Cambridge: Cambridge University Press.

MIXDORFF, H. 2004. Qualitative analysis of prosody in task-oriented dialogs, Proc. 2nd International Conference on Speech Prosody, Nara, Japan, 283-286.

MIXDORFF, H. / H.R. PFITZINGER. 2005. Analysing fundamental frequency contours and local speech rate in map task dialogs. Speech Communication 46, 310-325.

MIXDORFF, H. / U. PECH / C. DAVIS / J. KIM. 2007. Map Task Dialogs in Noise - a Paradigm for Examining Lombard speech. Proc. 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 1329-1332.

MIXDORFF, H. / A. HÖNEMANN / G. ZELIC / J. KIM / C. DAVIS. 2014. The Cartoon Task – Exploring Auditory-Visual Prosody in Dialogs. Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 1067-1071.

MOSEL, U. 2006. Field Work and Community Language Work. In: J. Gippert, N.P. Himmelmann & U. Mosel (Eds), Essentials of Language Documentation (pp. 67–83). Berlin/New York: de Gruyter.

MOSEL, U. 2011. Morphosyntactic analysis in the field - a guide to the guides. In: N. Tieberger (Ed.), The Oxford handbook of linguistic fieldwork (pp. 72-89). Oxford: OUP.

NENKOVA, A. / A. GRAVANO / J. HIRSCHBERG. 2008. High frequency word entrainment in spoken dialogue. Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL) with the Human Language Technology Conference, Columbus, USA, 169-172.

NEWMAN, P. / M. RATLIFF. 2001. Linguistic Fieldwork. Cambridge: Cambridge University Press.

NIEBUHR, O. 2007a. The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. Phonetica 64, 174-193.

NIEBUHR, O. 2007b. Perzeption und kognitive Verarbeitung der Sprechmelodie – Theoretische Grundlagen und empirische Untersuchungen. Berlin/New York: de Gruyter.

NIEBUHR, O. 2008. Coding of Intonational Meanings Beyond F0: Evidence from Utterance-final /t/ Aspiration in German. Journal of the Acoustical Society of America 142, 1252–1263.

NIEBUHR, O. 2009. Intonation Segments and Segmental Intonations. Proc. 10th Interspeech Conference, Brighton, UK, 2435–2438.

NIEBUHR, O. 2010. On the Phonetics of Intensifying Emphasis in German. Phonetica, 170–198.

NIEBUHR, O. / J. BERGHERR / S. HUTH / C. LILL / J. NEUSCHULZ. 2010. Intonationsfragen hinterfragt - Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. Zeitschrift für Dialektologie und Linguistik 77, 304-346.

NIEBUHR, O. / M. D'IMPERIO / B. GILI FIVELA / F. CANGEMI. 2011a. Are There 'Shapers' and 'Aligners'? Individual Differences in Signalling Pitch Accent Category. Proc. 17th International Congress of Phonetic Sciences, Hong Kong, China, 120–123.

NIEBUHR, O. / M. CLAYARDS / CH. MEUNIER / L. LANCIA. 2011b. On Place Assimilation in Sibilant Sequences - Comparing French and English. Journal of Phonetics 39, 429–451.

NIEBUHR, O. / CH. MEUNIER. 2011. The phonetic manifestation of French /s#S/ and /S#s/ sequences in different vowel contexts - On the occurrence and the domain of sibilant assimilation. Phonetica 68, 133-160.

NIEBUHR, O. 2012. At the edge of intonation - The interplay of utterance-final F0 movements and voiceless fricative sounds. Phonetica 69, 7-27.

NIEBUHR, O. 2013. On the acoustic complexity of intonation. In: E.-L. Asu & P. Lippus (Eds), Nordic Prosody XI (pp. 15-29). Frankfurt: Peter Lang.

NIEBUHR, O. / J. HOEKSTRA. 2014. Pointed and plateau-shaped pitch accents in North Frisian dialects. Proc. 14th International Conference on Laboratory Phonology, Tokyo, Japan.

NOLAN, F. 1992. The descriptive role of segments: evidence from assimilation. In D.R. Ladd & G.J. Docherty (Eds), Papers in Laboratory Phonology 2 (pp. 261–280). Cambridge: CUP.

NOLAN, F. 1995. The Effect of Emphasis on Declination in English Intonation. In: J.W. Lewis (Ed.), Studies in General and English Phonetics. Essays in Honour of Professor J.D. O'Connor (pp. 241–254). London & New York: Routledge.

ODGEN, R. 2006. Phonetics and social action in agreements and disagreements. Journal of Pragmatics 38, 1752–1775.

OHL, C.K. / H.R. PFITZINGER. 2009. Compression and Truncation Revisited. Proc. 10th Interspeech Conference, Brighton, UK, 2451–2454.

PARBERY-CLARK, A. / E. SKOE / N. KRAUS. 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. Journal of Neuroscience 25, 14100–14107.

PARDO, J.S. 2006. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 119, 2382-2393.

PASTEUR, L. 1939. OEuvres, tome VII: mélanges scientifiques et littéraires. Paris: Masson. http://catalogue.bnf.fr/ark:/12148/cb37416454q.

PETERS, B. 1999. Prototypische Intonationsmuster in Deutscher Lese- Und Spontansprache. AIPUK 34, 1–177.

PETERS, B. 2005. The Database 'The Kiel Corpus of Spontaneous Speech'. AIPUK 35a, 1–6.

PETERS, J. 2006. Intonation deutscher Regionalsprachen (Linguistik - Impulse & Tendenzen, Vol. 21). Berlin/New York: de Gruyter.

PETTORINO, M. / E. PELLEGRINO / M. MAFFIA. 2014. "Young" and "Old" Voice: the prosodic auto-transplantation technique for speaker's age recognition. Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 135-139.

PICKERING, M.J. / S. GARROD. 2004. Towards a mechanistic psychology of dialogue. Behavioral and Brain Sciences 27, 169–226.

PIERREHUMBERT, J.B. / J. HIRSCHBERG. 1990. The meaning of intonation contours in the interpretation of discourse. In: P.R. Cohen, J. Morgan, and M.E. Pollack (Eds), Intentions in communication (pp. 271-311). Cambridge, Mass.: MIT Press.

PITT, M.A. 1994. Perception of pitch and timbre by musically trained and untrained listeners. Journal of Experimental Psychology: Human Perception & Performance 20, 976-986.

PORT, R. / M. O'DELL. 1985. Neutralization of Syllable-final Voicing in German. Journal of Phonetics 13: 455–471.

PRIETO, P. 2012. Experimental methods and paradigms for prosodic analysis. In: A.C. Cohn, C. Fougeron & M.K. Huffman (Eds), Handbook in Laboratory Phonology (pp. 527-547). Oxford: Oxford University Press.

PROBERT, PH. 2009. Comparative philology and linguistics. In: B. Graziosi, Ph. Vasunia & G. Boys-Stones (Eds.), The Oxford Handbook of Hellenic Studies (pp. 697-708). Oxford: Oxford University Press.

REITTER, D. / J.D. MOORE. 2007. Predicting success in dialogue. Proc. 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 808–815.

RIETVELD, T. / C. GUSSENHOVEN. 1995. Aligning pitch targets in speech synthesis: Effects of syllable structure. Journal of Phonetics 23, 375-385.

SCHERER, K. / D. GRANDJEAN / T. JOHNSTONE / G. KLASMEYER / T. BÄNZIGER. 2002. Acoustic correlates of task load and stress. Proc. International Conference on Spoken Language Processing, Denver, USA, 2017-2020.

SCHÖN, D. / C. MAGNE / M. BESSON. 2004. The music of speech: music facilitates pitch processing in language. Psychophysiology 41, 341–349.

SCHÖTZ, S. 2006. Perception, Analysis and Synthesis of Speaker Age (Travaux de l'Institut de linguistique de Lund 47). Lund : Media-Tryck.

SCHWAB, S. 2011. Relationship between Speech Rate Perceived and Produced by the Listener. Phonetica 68, 243- 255.

SHEN, R. 2013. Tones and consonants in Shibei Min Chinese. Proceedings of International Conference on Phonetics of the Languages in China (ICPLC-2013), Hong Kong, China, 171–174.

SIMPSON, A.P. 2009. Phonetic differences between male and female speech. Language and Linguistics Compass 3, 621-640.

SIMPSON, A. 2012. The First and Second Harmonics Should Not Be Used to Measure Breathiness in Male and Female Voices. Journal of Phonetics 40, 477–490.

SIMPSON, A. / K.J. KOHLER / T. RETTSTADT. 1997. The Kiel Corpus of Read/Spontaneous Speech: Acoustic Data Base, Processing Tools, and Analysis Results. Kiel. AIPUK 32.

SKOPETEAS, S. / I. FIEDLER / S. HELLMUTH / A. SCHWARZ / R. STOEL / G. FANSELOW / C. FÉRY / M. KRIFKA. 2006. Questionnaire on Information Structure: Reference Manual. Interdisciplinary Studies on Information Structure 4. Potsdam: Potsdam University Press.

SOUAG, L. 2011. Review of: Claire Bowern. 2008. Linguistic Fieldwork: A Practical Guide. Language Documentation and Conservation 5, 66–68.

SPILKOVÁ, H. / D.S. BRENNER / A. ÖTTL / P. VONDRICKA / W. VAN DOMMELEN / M. ERNESTUS. 2010. The Kachna L1/L2 Picture Replication Corpus. Proc. 7th International Conference on Language Resources and Evaluation, Malta, Spain, 2432-2436.

STEGEMÖLLER, E. / E. SKOE / N. TRENT / C.M. WARRIER / N. KRAUS. 2008. Musical Training and Vocal Production of Speech and Song. Music Perception 25, 419-428.

STEPPLING, M.L. / A.A. MONTGOMERY. 2002. Perception and production of rise-fall intonation in American English. Perception and Psychophysics 64, 451-461.

SUNDBERG, J. 1979. Maximum Speed of Pitch Changes in Singers and Untrained Subjects. Journal of Phonetics 7, 71–79.

TITZE, I.R. 1989. Physiologic and Acoustic Differences Between Male and Female Voices. Journal of the Acoustical Society of America 85, 1699–1707.

TORREIRA, F. / M. ADDA-DECKER / M. ERNESTUS. 2010. The Nijmegen Corpus of Casual French. Speech Communication 52, 201-221.

TURCO, G. / M. GUBIAN / J. SCHERTZ. 2011. A quantitative investigation of the prosody of Verum Focus in Italian. Proc. 12th Interspeech Conference, Florence, Italy, 961-964.

TURK, A. / S. NAKAI / M. SUGAHARA. 2006. Acoustic segment durations in prosodic research: A practical guide. In: S. Sudhoff et al. (Eds.), Methods in empirical prosody research (pp. 1-28). Berlin/New York: de Gruyter.

VAISSIÈRE, J. 2004. The Perception of Intonation. In: D.B. Pisoni & R.E. Remez (Eds.), Handbook of Speech Perception (pp. 236–263). Oxford: Blackwell.

VAISSIÈRE, J. / K. HONDA / A. AMELOT / SH. MAEDA / L. CREVIER-BUCHMAN. 2010. Multisensor Platform for Speech Physiology Research in a Phonetics Laboratory. Journal of the Phonetic Society of Japan 14, 65–77.

VÁRADI, T. / P. WITTENBURG / S. KRAUWER / M. WYNNE / K. KOSKENNIEMI. 2008. CLARIN: Common language resources and technology infrastructure. Proc. 6th International Conference on Language Resources and Evaluation, Marrakech, Morocco, 1244-1248.

WAGENER, P. 1986. Sind Spracherhebungen paradox? Über die Möglichkeit, natürliches Sprachverhalten wissenschaftlich zu erfassen. In: A. Schöne (Ed.), Akten des VII. IVG-Kongresses, Vol. 4 (pp. 319-327). Tübingen: Niemeyer.

WARD, A. / D. LITMAN. 2007. Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. Proc. SLaTE Workshop on Speech and Language Technology in Education, Farmington, USA, 533-538.

WATSON, I. 2002. Convergence in the Brain; the Leakiness of Bilinguals' Sound Systems. In: M. Jones & E. Esch (Eds), Language Change: The Interplay of Internal, External and Extra-linguistic Factors (pp. 243–266). Berlin/New York: Mouton de Gruyter.

WELBY, P. / H. LOEVENBRUCK. 2006. Anchored down in Anchorage: Syllable structure and segmental anchoring in French. Italian Journal of Linguistics 18, 74-124.

XIA, Z. / R. LEVITAN / J. HIRSCHBERG. 2014. Prosodic Entrainment in Mandarin and English: A Cross-Linguistic Comparison. Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 65-69.

XU, J. 2012. Problems and coping strategies of speech data collection - Insights from a special-purpose corpus of situated adolescent speech. Manuscript, University of Science and Technology of China. http://fld.ustc.edu.cn/123/xujiajin/index.htm

XU, Y. 2011. Speech prosody: A methodological review. Journal of Speech Sciences 1, 85-115.

ZELLERS, M. / B. POST. 2012. Combining Formal and Functional Approaches to Discourse Structure. Language and Speech 55, 119-139.