Introduction

The Study of Prosody and Intonation

The study of prosody in general, and of tonal aspects in particular, has occupied the human mind for centuries in the attempt to elucidate their contributions to meaning over and above that conveyed by lexical fields and syntactic structures. Systematic analysis of intonation has been carried out with increasing breadth and depth over the past century. In the London School of Phonetics, the auditory description of contrastive tonal patterns formed part of detailed overall phonetic accounts of languages, first and foremost of English, but also of other European and non-European languages, with practical application to foreign language teaching (Cruttenden 1986 (2nd edn 1997), Crystal 1969a, O'Connor and Arnold 1961). This descriptive framework was also applied to the study of hitherto unwritten languages within the British Empire, going back to the late nineteenth century. To establish the formal prosodic carriers of meaning was the primary concern in this approach, only followed, in a second step, by an analysis of the meanings carried. The descriptions were largely given in relation to syntactic structure, and *ad hoc* pragmatic minutiae were provided without a systematic semantic theory behind them. The same methodology applies to Pike's analysis of American English intonation (Pike 1945).

The division of speech science into phonetics and phonology eventually also shaped the study of prosody, resulting in the systemic and structural phonological accounts of intonation by Halliday in Britain and by Liberman, Pierrehumbert and others in Autosegmental Metrical (AM) Phonology in North America. The latter approach set a pattern for research in Europe and around the world, now moving over from auditory description to experimental and instrumental, mainly acoustic, analysis. The phonological perspective, more particularly in the frame of Laboratory Phonology, became all-pervasive in Robert Ladd's textbook (1996 (2nd edn 2008)). In a phonological paradigm, phonetic substantiation takes second place to underlying phonological form

in the linguistic modelling of a language. To arrive at this form it is necessary to send speech signals through a linguistic filter that eliminates the exponents of the speaker's expressiveness and the attitudes to the listener, restricting the speech signal to a representational core. But, as was pointed out by Bolinger (1986), this deprives intonation of a substantial part of its specific signalling power in everyday speech communication, and reduces it to the status of the maid of syntactic structure.

As a logical corollary, meaning is subordinated to form in the resulting prosodic systems. The question is not 'What are the communicative functions in a network of human interactions, and how are they manifested by formal means – lexical, syntactic, prosodic, including gesture and facial expression – in speech acts in the languages of the world?' Rather, meaning is grafted onto formal linguistic structures, which Ladd (1996 (2nd edn 2008)) termed 'The Linguist's Theory of Intonational Meaning', focusing on the representational meaning of linguistic structures. Affect and attitude are brought in again *post hoc* by introducing 'intonological' choices, such as more or different pitch accents, as an overlay to linguistic intonational meaning.

Such form-oriented descriptive accounts of the prosodic phonology of a language are very useful as the first step of analysis, particularly when nothing or very little is known about the contrastive patterns in a language. But they do not give us enough insight into how speech communication works in all its facets of meaning transmission. Speech scientists should now prepare to take the second step and move from function to form in those languages that have been thoroughly investigated formally: the Germanic, Romance and Slavonic languages (especially Dutch, English, German, Danish, Swedish, French, Spanish, Italian, Russian and Czech), but also Hindi, Arabic, Japanese and Chinese, and develop an interlanguage network of communicative functions with their language-specific prosodic and linguistic formal exponents.

To make this successful, the prevalent paradigm of dealing with phonology in general and with prosody in particular will need adjusting. In recent decades, there has been an increased focus of structural linguistics on form through the introduction of speech signal analysis, especially in Laboratory Phonology, with its insistence on signal measures to substantiate phonological form. This may be subsumed under what the *gestalt* psychologist Karl Bühler termed *Stoffentgleisung*, i.e. 'the material fallacy' (Bühler 1934, p. 46), with pointed reference to Skinner-type behaviourist psychology, and to pre-phonology experimental phonetics of the early twentieth century (Panconcelli-Calzia 1948; Rousselot 1892, 1897–1901; Scripture 1902, 1935). The phonetics-in-phonology research paradigm of Laboratory Phonology (Ladd 2011; cf. Kohler 2013a) can even

leave out the initial auditory observation stage and go straight into experimental testing.

The detachment from auditory observation and understanding of speech events has also removed the need for the analyst to be proficient in the language to be analysed; analysis software like Praat is supposed to do the work, and statistics provides the significance test. However, taking subjects to the laboratory and putting them in various experimental stylisations may not be the appropriate method for investigating exponents of communicative functions in natural speech. Therefore, rethinking the place of speech signal analysis in a function-form framework of language theory is becoming a pressing need.

A Change of Perspective

This monograph aims to develop 'A Speech Scientist's Theory of Intonational Meaning', in place of 'The Linguist's Theory of Intonational Meaning', by putting a network of communicative functions first and then relating formal exponents to them. Although the main focus will be on prosody, lexical and syntactic forms need to be taken into account as well, whenever they are part of the formal manifestations of particular functions. Therefore, the title refers to the wider *Linguistic Forms* rather than the narrower *Prosodic Forms*. The functional approach allows us to replace the linguistic and phonological filtering of speech by an allocation of signals in individual communicative acts to types of interrelated functions in human interaction. These speech act signals are, in turn, reduced to significantly distinctive property categories in relation to the system of communicative functions. This is 'phonology coming out of phonetics', rather than 'phonetics going into phonology'. The principal goal in writing this monograph is theoretical, to present a (partial) network of communicative functions in human language and to relate intra- and interlanguage speech forms to these functions. Prosodic form will be taken in a very broad sense, including the phenomena of *elaboration* in functional highlighting, as well as of *reduction* in functional attenuation, which both involve segmentals alongside prosodies in the narrow sense.

In developing a functional framework, the monograph takes its point of departure from Karl Bühler's *Sprachtheorie* (1934), and from an evaluative review of the relevant phonetic, phonological and linguistic literature against such a functional background. It includes a discussion on a methodology of data acquisition that is based on contextualisation adapted to the function-form paradigm. The presentation of a network of communicative functions is built around a comparison of formal exponents in German and English, with

extensions across a wider array of European languages, including the more distant Romance family, especially French, a prosodically unique language within Europe. The theoretical framework is proposed as a powerful tool in comparative prosodic research into the world's languages. The potential of this approach is shown by the analysis of data from a tone language, Mandarin Chinese, which were collected in functionally contextualised scenarios, and are compared with functionally corresponding data from German and English.

The function-form paradigm is based on the axiomatic postulate that a core network of communicative functions is inherent in human speech interaction, irrespective of any particular language. For example, to signal authority and dominance, or subordination and compliance, to highlight or attenuate messages, or to stimulate dialogue partners into action through questions, commands or requests can, among others, be assumed to belong to such a communicative core in human interaction. However, the association of form at various levels of description from lexicon and syntax to prosody varies between languages. Yet there are also very clear cases of universally used forms for specific communicative functions, e.g. high pitch in certain types of question, or phonation features in negative intensification, or low- versus high-pitch register for the expression of authority versus subordination. In other cases, languages form typological groups, genetically related or not, using the same forms for particular functions, and, finally, formal features may be individual-language specific. Thus, the formal mapping of communicative functions across the world's languages becomes an exciting field of research in language comparison and typology.

Principles of a Communicative Phonetic Science

Doing phonetic analysis in a function-form framework follows general scientific principles. In all sciences dealing with experience of the world and of actions within it, scientific questions start with individual observations *hic et nunc*: this may be the legendary apple falling on Newton in physics, or the sound of an utterance impinging on the eardrum, and being understood by the brain, of a phonetician, who picks it up, e.g. during a bus ride on a particular day in a particular area. Both the physicist and the phonetician then start asking how they can explain these events beyond the *hic et nunc*. They have an idea that generalises to other occurrences of such events observable on other occasions. They then formulate first principled statements which they incorporate into the theory they already have of the universe they are working on, the physical world or the communication between speakers and listeners. On the

basis of these principled statements incorporated into the theory, the physicist and the phonetician derive hypotheses 'If A holds, then B must be true.' They then take these hypotheses into the laboratory for experimental validation or rejection, representing the results in numbers.

Although there is in general perfect parallelism between the physicist's and the phonetician's investigation, a fundamental difference needs to be recognised between the 'events' the two deal with. The physicist simply observes events out there in the world; the phonetician observes and *understands* events in relation to a system of linguistic signs in linguistic structures used in action fields. This means that the phonetician must be able to understand the signals received for analysis, i.e. must have a sufficient proficiency in the language under investigation, or acquire it in the course of fieldwork. It also means that physical analysis of speech signals can never give the whole answer about the events the phonetician has observed. Contrary to quite common belief, the phonetician's analysis is not a discovery procedure for communicative functions and forms, either. It is a validation of hypotheses derived from a theory of *Communicative Phonetic Science*, which provides a scientific construct for analysing speech interaction in socio-cultural language settings by auditory evaluation and experimental measurement. Thus *Communicative Phonetic Science*, as conceived of here, combines principles of the natural sciences with the phenomenology of the humanities.

I see five essentials for this validation process in *Communicative Phonetic Science*:

(1) Phonetic science is built on a theory of speech communication in human interaction based on a small number of axioms.
(2) Specific research questions, raised by the phonetician, in observing *ad hoc* speech events, or in evaluating the state-of-the-art in the particular research field, are anchored in this theory.
(3) Hypotheses are derived for the validation of the specific questions within the theory, thus continually deepening the theoretical foundations.
(4) Appropriate methodologies are developed for the collection of communicatively valid data for the specific questions.
(5) In data analysis, auditory evaluation precedes measurement.

Here is an example to illustrate the progression from observation to scientific analysis. On a local bus in Kiel, I was sitting in a window seat beside a button for signalling to the driver that one wants to get off at the next stop. Diagonally opposite and facing me was a young woman who was playing with her

smartphone. When the bus was approaching the stop where she wanted to get off, she put her smartphone in her handbag, looked at me, stretched out her arm and, with her index finger pointing to the button, said, 'Drücken Sie bitte' [Press (the button) please]. 'Drücken Sie' was high level, then there was a small step of about two semitones down to another level tone. In this communicative speech-and-gesture action, the woman presupposed shared 'world' knowledge between us for me to interpret her deixis appeal correctly in the way she intended it. I had this common communicative ground and understood her utterance to mean 'I cannot reach the button, please press it for me because I want to get off.' It was a pure stimulant to act on her behalf, without any command or exuberant request appeal. And I was quite happy to comply. Afterwards it occurred to me that if she had used a falling pitch on 'Drücken', ending in low pitch on 'Sie bitte', I would have received it as an impolite command. I have since extrapolated this to other instances of stepping patterns I have collected, and I have studied the relevant literature on the subject. A clear picture is beginning to emerge, which I have formulated as a principled statement in the function-form framework of speech communication in 4.1, preparing the ground for further testing.

Experimental testing of such interactive speech production becomes a problem because the data are so context-dependent and require such a great deal of empathy on the part of informants that it requires a lot of ingenuity to devise situational dialogue interactions between subjects. Therefore, research into phonetic exponents of specific speech functions needs to draw on two data sources. The *first data source*, comprising corpus data of various forms of spontaneous speech, collected in a variety of scenarios, provides a rich documentation of segmental and prosodic variability in words and utterances. The *Call Home* corpus of American English telephone conversations is one form of a spontaneous speech corpus, which allows the *Communicative Phonetic Science* analysis of an array of interactional phenomena, especially in a Conversation Analysis framework (Local and Walker 2005; Ogden 2012), of course only within the frequency and signal/noise ratio limitations of telephone speech.

Many other corpora that are called spontaneous are not spontaneous in the defined sense of natural communicative interaction. They are unscripted at best, generated in a metalinguistically designed scenario in a recording studio, such as the appointment-making scenario of the *Kiel corpus of spontaneous speech*; (IPDS 1995–7). These dialogues should be called semi-spontaneous. Generally, they have the sound of pairs of speakers playing games according to instructions, but there are some where the listener gets the impression that the speakers are interacting in a natural communicative setting of arranging

mutually suitable days and times to meet. These dialogues contain typical non-lexical interactive sounds such as laughing. The *Video Task* scenario (Peters 2001) creates a communicative situation in the studio that moves one step further towards natural interaction. Similar but non-identical video clips from the well-known German television series *Lindenstrasse* are presented to two subjects sitting in separate rooms. After the presentation, the subjects discuss differences and similarities in what they have seen and heard.

The (semi-)spontaneous corpora lack the systematicity of experimental speech designs, but they allow the formulation of tentative hypotheses for further systematic data collection. These hypotheses can then be tested with a *second data source* which is constructed in the laboratory, taking care to achieve the greatest possible naturalness with data collection scenarios and data types that are communicatively plausible and meaningful. This rules out data samples of the type 'Die Nonne und der Lehrer wollen der Lola in Murnau eine Warnung geben, und die Hanne will im November ein Lama malen' [The nun and the teacher want to give a warning to Lola from Murnau, and Hanna wants to paint a lama in November], which are detached from communicative functions and serve a metalinguistic principle, i.e. to generate continuous stretches of voicing for F0 analysis (Truckenbrodt 2002).

The lab analysis of communicative functions, e.g. of focus, increases the insight into these functions when meaningful sentences are contextualised in plausible interaction scenarios. But question–answer paradigms of the type

Prompt:	Target:
Who may know your niece?	Lee may know my niece.
What may Lee do to your niece?	Lee may lure my niece.
Who may Lee know?	Lee may know my niece.
What did you say?	Lee may know my niece.

(Xu and Xu 2005), frequently used in the study of contrastive focus, do not meet this criterion. Quite apart from the questions being communicatively odd, their nominal and verbal elements would not all be repeated in natural dialogue answers, which would rather be given as 'Lee may', 'He may lure her', '(He may know) my niece.' The reason all elements are to be repeated by the informants again derives from a metalinguistic principle, this time to provide a homogeneous frame for comparing narrow focus realisation phrase-initially, -medially and -finally in relation to the broad focus of the fourth target. The analysis of such data sets can only make statements about focus exponents in this metalinguistic data generation and should not be generalised to focus in speech communication. We must not take subjects to the laboratory and put

them through highly stylised procedures only to obtain numerical data under controlled conditions. The recordings may be substantially removed from natural talk in interaction and greatly limited in what they can tell us about the exponents of communicative functions.

Contextualisation requires a great deal of refinement to meet the requirements of *Communicative Phonetic Science*. To tease apart representational, attitudinal and expressive meanings, scenarios need to be devised in which competent speakers (whose proficiency in the use of their language is tested beforehand) enact communicative functions that are the goal of the analysis. Kohler and Niebuhr (2007) and Niebuhr (2010) use such a methodology of data acquisition for negative or positive intensification. But even the most sophisticated contextualisation procedures in the laboratory cannot guarantee natural interaction, crucial in spontaneous communication. For certain phenomena of spontaneous speech interaction, for example the use of pitch stepping, it is exceedingly difficult to obtain recordings of talk in interaction. In such cases, the speech scientist needs to accept the trained native language expert's auditory and visual observation of ongoing speech communication as a valid *third data source* beside the systematic analysis of recorded (semi-)spontaneous and lab-generated corpora.

This is Bertrand Russell's 'Knowledge by Acquaintance' versus 'Knowledge by Description', i.e. *ad hoc* native expert observation, and competent reproduction, of talk in interaction unfolding *in situ* as against extra-communicative formal and numerical analysis of recorded linguistic objects. This kind of *ad hoc* observation of meaning transmission through speech is even further removed from systematicity than phenomena in collected speech corpora, but it constitutes an essential and continually necessary step in the acquisition of knowledge about speech and language, and reinstates the very successful methodology of the traditional 'ear phonetician' in the London School of Phonetics. It may be extremely difficult to evaluate sensory observation of communicative action by subsequent measurement of its physical parameters because eliciting and recording communicatively relevant data is problematic. In such cases, numerical validation can be obtained through suitably devised speech perception and understanding experiments.

Methodologies of data acquisition in speech perception and understanding have already been practised for some considerable time, but they, too, need further refinement. When function-form structures have been clarified in speech production, listening tests can be developed that systematically present natural speech stimuli for judgement, either varied in natural production by a competent native speaker, or in systematic parameter manipulation

(e.g. F0) of one or more natural base stimuli in such analysis tools as Praat. Judgement paradigms may be the Semantic Differential (Ambrazaitis 2005; Dombrowski 2003, 2013; Dombrowski and Niebuhr 2010; Kohler 2005, 2011b; Osgood, Suci and Tannenbaum 1957; Uldall 1960, 1964) or context matching of test stimuli (Kleber 2006; Kohler 1987b; Niebuhr 2007a, b; Niebuhr and Kohler 2004). The former are more powerful if the test stimuli are contextually embedded (Kohler 2011b). In the cited papers, semantic scales were constructed for the particular questions in hand. The strength of the Semantic Differential Technique in phonetic research will be increased in future when it is developed into a methodology of standardised sets of scales.

In any form of context matching, to be of use in a function-form framework, context and test stimuli should represent a natural communicative sequence, either in the same voice within a dialogue turn, or across two successive turns with a change of voices, preferably of gender as well. The test stimulus always follows the context setting to trigger an immediate response to it in its contextual embedding, which is not possible if the order of context and test stimulus is reversed. The context can either be given generally in the introduction to the whole listening test (Kohler 2011b) or, if it can be captured in a short enough phrasing within or across dialogue turns, it may be appended before each test stimulus (Kohler 1987b; Niebuhr 2007b; Niebuhr and Kohler 2004). The proper matching paradigm tests whether listeners do or do not apprehend the test stimulus as fitting into the preceding context, and the experimenter then interprets these responses as referring to the categories under investigation.

Content and Readership

This monograph develops a prosodic model within a function-form framework of human speech communication and then looks at prosodic, beside syntactic and lexical, manifestations of selected speech functions. The discussion starts with a focus on German and English. The variety of German is the Northern Standard, the variety of English the Southern British Standard, with occasional references to other varieties of English around the world, especially in North America. To keep function and form categories clearly distinct, the former are symbolised in small capitals, the latter in italics.

Chapter 1 **Speech Communication in Human Interaction** sets the theme by taking, as the point of departure, two central concepts of Karl Bühler's *Sprachtheorie* (Bühler 1934): (1) the *Organon Model*, which relates the linguistic sign to the Speaker, the Listener and the world of Objects and Factual

Relations in the three basic communicative functions of EXPRESSION, APPEAL and REPRESENTATION; (2) the two fundamentally different fields of speech communication, the *pointing* or *deictic* and the *naming* or *symbolic*. The chapter looks at the ways deictic communication is structured with reference to four pointing dimensions, the sender, the receiver, objects away from the sender and the receiver, and far-away objects. Illustrations are given from German and English.

The English translation of Bühler's *Sprachtheorie, Theory of Language* by Donald Fraser Goodwin, in collaboration with the sematologist Achim Eschbach, was first published by John Benjamins in 1990, and then in a new edition in 2011 (see Bühler 1934). It is a welcome production, with the Editor's (Achim Eschbach) *Introduction – Karl Bühler: Sematologist*, the translator's *Preface*, a modern-style bibliography of the works cited by Bühler and a glossary. The new edition also contains a *Postscript Twenty-Five Years after…* by Eschbach, and a paper by Abraham (2011). The pagination of the German book was inserted in the text of the English translation, so I will give page references to the German text, except in English quotations, where both page references are provided. The translation is on the whole good and in fluent English style, which is quite remarkable, considering the very complex academic diction of the original. However, a couple of key terms of Bühler's theory do not seem to me to be quite adequate:

- Bühler's 'Gegenstände und Sachverhalte' was translated as 'Objects and States of Affairs'. 'State of affairs' is 'Zustände', referring to something static and passive. 'Sachverhalte' is linked to the verb 'verhalten', and thus refers to the active relations that objects enter into. I therefore replace 'state of affairs' by 'factual relations'.
- Bühler's verb 'zuordnen' and noun 'Zuordnung' were rendered by 'coordinate' and 'coordination'. In walking and running the movements of the legs are coordinated, but what Bühler refers to is something different: the mapping of sound signs to objects and factual relations 'in terms of modern mathematics' (Bühler 1934, p. 29). I use the translation 'map'.
- Bühler distinguishes two types of deixis: (1) direct pointing in the actual situation in which the interaction between sender and receiver takes place, (2) mediated pointing in a situation constructed mentally in talk and displaced from the one of actual interaction. The second deictic situation is created 'im Bereich der ausgewachsenen *Erinnerungen* und der konstruktiven *Phantasie*' (Bühler 1934, p. 123),