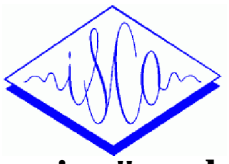
ISCA Archive
http://www.isca-speech.org/archive



First International Workshop on the History of Speech Communication Research (HSCR 2015)
Dresden, Germany
September 4-5, 2015

The "Bonn Connection" and its consequences: Paul Menzerath and Werner Meyer-Eppler's reunification of phonetics and phonology and the emergence of a new phonetic speech science based on Shannon's Mathematical Theory of Communication

Hans G. Tillmann, Jessica Siddins
LMU München, IPS

tillmann@phonetik.uni-muenchen.de
jessica@phonetik.uni-muenchen.de

Abstract: 1950, only two years after it was founded, the MIT's Electronic Research Laboratory (ERL) organised the first „Speech Communication Conference". Surprisingly, among the fourteen very prominent participants there were merely two representatives from the field of phonetic speech research: *Paul Menzerath* and *Werner Meyer-Eppler* from Germany; and it was on this very trip to the USA that they first met Claude Shannon of Bell Labs. Upon their return to the University of Bonn they founded the ***Institut für Phonetik und Kommunikations-forschung***[1] which completely changed the direction of phonetics into speech communication research by forming a unique collaboration originally based on Shannon's *Mathematical Theory of Communication*. We begin with a discussion of how it took another eighty years following the great early successes in speech physiology and subsequent school of early instrumental phonetics from 1850 onwards for the classic phonetic theory of vowels and consonants to finally come crashing down. We then outline how this crisis was overcome by the paradigm shift at the institute in Bonn and describe the consequences for the subsequent development of phonetic theory and its applications. Finally, we demonstrate that – looking back – the eventual collapse of early instrumental phonetics was virtually pre-programmed and could be expected because of the initially successful but in hindsight far to optimistic "simplifying assumptions" made by speech physiologists such as Brücke (1849/1856) and Bell (1865).

N.B.: An extended version of this paper will be prepaired for the Conference and made available on the internet for downloading.

1. Early modern phonetics since 1850 - from the initially great successes to the final disaster of instrumental work eighty years later

A full description of "Early Modern Phonetics, especially Experimental and Instrumental Work" has been given by the first author 1994 in EmP[2]. With respect to the paradigm shift described in section 2, it should be pointed out that, quite different from the situation in the US (and the UK), in the years after 1920 the academic development of speech sound research in central Europe had spiralled further and further into a deep crisis which finally led to

[1] It should be noted that to L1-speakers of German the term *Kommunikation* at that time was a strange foreign word and totally incomprehensible; it took a few years before e.g. the verb *kommunizieren* ('to communicate') became a familiar German loan-word.
[2] ENCOCLOPEDIA OF LANGUAGE AND LINGUISTICS (p. 3082 – 3095), reprinted in Koerner et. al (1995)

Trubetzkoy's strict partitioning of the scientific study of phonetics from purely linguistic phonology.

In the mid19th century, Brücke and with him the field of phonetic science split from the philological disciplines so that two different fields went their own ways. For Brücke "the time has come" to present the 'natural value' of speech sounds and to explain the 'natural connection between the sounds and their signs' (1856, p.1).

The *value* of any alphabetically given sound is determined by its properties which were called attributes, and Brücke distinguishes **two different ways** of finding these attributes:

> „Man kann die Art und Weise untersuchen, wie sie Nachbarlaute affizieren und von ihnen affiziert werden […], um hieraus ihre Attribute herzuleiten. **Dies ist der Weg des Sprachforschers**." [3]
>
> „Andererseits kann man directe Beobachtungen und Versuche über die Art und die Bedingungen ihrer Entstehung anstellen und hieraus eine Einsicht in ihre Natur und ihre Eigenschaften gewinnen. **Dies ist der Weg des Physiologen**."[4]

Brücke (1849, 1856) in Austria and - only a few years later - Bell (1863, 1867) in England had learned[5] that certain articulatory configurations could be attributed to each vowel and consonant: for each vowel, there is a specific tongue height and position and lip aperture: rounded or unrounded. Other articulatory configurations are characteristic of consonants, in that a certain articulating organ may be associated with a certain place of articulation, and the degree of constriction determines the manner of articulation[6].

Those who referred to themselves as "sound physiologists" at that time were not true physiologists, but rather early behaviourists with one eye on the letters of the alphabet and the other on the visual or tactilely perceivable speech movements which could be looked at when spelling out the alphabetic sounds in isolation.

In view of the phonetic revolution in Bonn (and the development from articulatory phonetics to experimental and instrumental phonetics described in EmP), we can identify three important stages:

(1) Instrumental phonetics with its graphical representations of the 'speech curves' can be seen as an early form of speech signal processing. (We return to this point in Section 3).

(2) The interest of research underwent a major transformation. While the strict identity of articulatorily defined vowels and consonants as constant units were the deciding factor for Brücke, Bell and their successors, the new perspective was that the instrumentally measured time functions showed a large amount of variability in the execution of speech movements. It were 1891 the "modifications phonétiques du langage" in the title of Rousselot's dissertation that anchored experimentally reproducible systematic variability of speech sounds at the centre of scientific interest. A summary of the most important results from today's perspective can be found in Scripture (1902)[7].

(3) As stated in EmP: „The paradoxical situation was that instruments had been introduced to replace subjective hearing by objective measurements, but the resulting picture did not show

[3] One can look at the sounds (represented by letters) to see how they affect their neighboring sounds, and how they are themselves affected by these, [and one can also follow the changes they suffer during the course of time and through their going over from one language into another] in order to determine their attributes. **This is the way of the linguist**.

[4] A different possibility is that direct observations and experiments concerning the manner and condition of their formation lead to the desired attributes. **This is the way of the physiologist**.

[5] Mainly from observations made during educational instruction for the deaf; cf. EmP (p. 3084).

[6] Why is manner of articulation not designated in Jespersen's analphabetic system? Look for the answer in EmP.

[7] For example, it included VOT, which was re-discovered only much later, as well as pre-final lengthening and the algorithm for manually calculating FFT for the spectral analysis of a vocal period drawn with a pencil from a gramophone disk.

what was originally looked for: visible speech sounds. These had disappeared and had to be reinvented by the new phonologists of the Prague School Phonology." (p. 3092)
The situation around 1930 is well preserved in the following quote from Scripture (1932), who reports that the audience was shocked when the first cineradiographic film of running speech was shown at a meeting of the International Society of Experimental Phonetics organized by Menzerath 1930 in Bonn:

> „Die kleinsten Bewegungen der Lippen, der Zunge, des Gaumensegels, des Zungenbeins, der Kehlkopfknorpel usw. spielen sich vor dem Auge ab". Scripture nennt den Eindruck eines solchen Films „überwältigend." Denn da sehe man einen „Schattenmenschen […], wie er spricht, atmet und schluckt. Die Sprechwerkzeuge stehen nicht für einen Augenblick still, jeder Sprechakt ist die Summe der Bewegungen aller Organe des Mundes, des Rachens, des Kehlkopfes usw., und diese Summe spielt sich in der Zeit ab. Lautstellungen gibt es überhaupt nicht: es kommt alles auf Lautbewegungen hinaus. Man begreift sofort, daß die bisherige Lautphysiologie nur eine Irrlehre sein kann, und wartet gespannt auf neues."(S.. 173)[8]

Another prominent researcher, Panconcelli-Calzia, argued that sound segments or syllables were not a phonetic reality at all. He considered such units pure fiction and an invention of linguists.

2. The Bonn Connection's way out of the cul-de-sac - a paradigm shift in phonetic speech science

If - according to Eli Fischer-Jörgensen - "all phoneticians are something else" (p.c.), then this clearly applies to the two Protagonists that caused such an innovative change of phonetic speech research.
Paul Menzerath (1883-1954), a Professor of Psychology at the University of Gent, left his chair when the University of Bonn offered him the possibility to work in the field of experimental phonetics. He installed a small Lab for instrumental and experimental phonetics which since 1928 published a whole series of excellent results (cf. EmP p. 3091 ff). For our context here the 1933 study on "Koartikulation, Steuerung und Lautabgrenzung" is of outstanding importance. It opened the door for a new phonetic theory of speech sounds in terms of vowels and consonants and explained their specific role in the articulatory production of syllabically well-formed natural speech utterances. When Menzerath and Lacerda discovered the phenomenon of coarticulation they were able to propose a *phonetic* way out of the dilemma described above. They used the kymographic method to show that the speech sounds at the beginning of the syllable are coarticulated, i.e., produced at the same time, and that the vowel in a syllable rhyme is stopped ('controlled') by the following consonant if there is one. Menzerath and de Lacerda observed various coarticulatory phenomena; however, in modern terms, *Koartikulation* for them mainly seems to refer to the anticipation of consonantal and vocalic features in syllabic initial position, whereas carry-over effects cannot easily be related to their concept of *Steuerung*. They formulated the new theory that the complex combination of simultaneous movements of the speech organs, which they called 'Synkinese', serves only for the purpose of producing 'acoustically' a clear sequence of separated and therefore directly segmentable sound units.[9] In any case they were able to

[8] The impression of such a film is overwhelming. The organs of speech do not remain still for an instant, every speech act is the combination of movements of all organs of the mouth, the throat, the larynx etc., and this combination is deployed over time. Sound positions simply do not exist. One understands at once that the sound physiology up to now has been based on an illusion and one awaits new explanations.
[9] In EmP it was pointed out, that Menzerath and Lacerda "failed to make any distinctions between purely acoustic sound segments in the physical world and the auditorily perceived sounds which again were taken as a common reality, an assumption that would later lead to a repetition of the first crisis" (p. 3093)

answer the open questions of "Wohlartikuliertheit"[10] that Techmer had asked in his inaugural lecture 1871 in Leipzig: "Naturwissenschaftliche Analyse und Synthese der hörbaren Sprache". An elaborated version was published in Techmer (1874); more about this is presented in EmP (p. 3087).

Werner Meyer-Eppler (1913-1960) studied mathematics, physics, and chemistry, and when he had presented his postdoctoral thesis on "Periodenforschung" (the title of his *Habilitationsschrift*[11]) he proved to be the world's leading expert in applying new mathematical and statistical methods to electronic speech signal processing.

It was an incredible stroke of luck that in the years after World War II the world's leading specialist of experimental phonetics (with his new acoustically based theory of alphabetically referable units of speech in a given language) came together with the much younger physicist Meyer-Eppler, the leading specialist in the field of acoustic speech signal production and transmission. Inviting him to become a researching member of his laboratory was the very beginning of the paradigmatic shift which in the first author's contribution to the Heike-Festschrift has been called "Bonner Wende" (much more details in Heike-FS (2013)). Already in 1949 Meyer-Eppler's published his very influential monograph "***Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache***"[12]. And it was due to Meyer-Eppler's initiative and his connections to Bell Labs and to the MIT that they both attended the first Conference on Speech Communication as mentioned in the Abstract. After their return they founded the new *Institut für Phonetik und Kommunikationsforschung* at the University of Bonn (belonging to the philosophical faculty of this University).

Four years after its formation, from 1954 to 1960, Meyer-Eppler took over the sole leadership of the institute. During these years he carried out a whole set of far-reaching activities[13]. The two most important ones must be mentioned here explicitly[14].

The first most important scientific activity of Meyer-Eppler was to develop - on the bases of Shannon's Mathematical Theory of Communication - his own completely new version of ***Information Theory*** which then appeared 1959 as the first Volume in the Series „**Kommunikation und Kybernetik in Einzeldarstellungen**" (edited by himself).

The second type of activity consisted in creating a whole set of research projects which were all financed externally. So Meyer-Eppler was able to offer research positions to a group of highly qualified young scientists which he attracted to Bonn.

Georg Heike was interested in Electronic Music[15], but Meyer-Eppler convinced him to investigate the acoustic features of the phonemes of the dialect of the city of Cologne.

Gerold Ungeheuer came from a technical university with degrees in electronics and communication engineering. In his dissertation supervised by Meyer-Eppler he used physics to solve the problem of computing the resonance frequencies of the vocal tract given the area-

[10] In section 3 below we will return to the concept of „well-articulatedness" which has been extensively dealt with in TmM (1980) the first author's book on "Phonetics: Spoken Signs, Speech Signals, and Speech Communication.

[11] By which he received 1942 the qualification as a university lecturer in physics at the *Faculty for Mathematics and Natural Sciences* of the Bonn University.

[12] Electric sound-generation: Electronic Music and Synthetic Speech

[13] In the more extensive version of this paper we will present a summary of the many diverse research projects initiated and conducted by Meyer-Eppler during the six years until his early death.

[14] More details are given in Heike-FS including the fact that Meyer-Eppler caused Manfred Schröder's move from the 3rd Physical Institute in Hanover to Bell Labs in the US (pc).

[15] In Elena Ungeheuer (1992) the complete story is given about the role, Meyer-Eppler played in the creation of the "Studio für Elektronische Musik" at the broadcastingstation "Kölner Rundfunk". The leading composer of this studio, Karlheinz Stockhausen, became a PhD-student of Meyer-Eppler and visited his lectures on information theory for several semesters.

function of the vocal tract, which determines the formant frequencies of the respective vowel. In 1957, he had already discovered what later has been called the *quantal nature of acoustic speech production*, but this has been widely ignored up until now in the English written literature (Jim Flanagan is one of the exceptions[16]).
Helmut Schnelle was a nuclear physicist and became a specialist of symbolic logic, artificial intelligence, and computer linguistics. He started the DAWID-Project with Heike and Rupprath before he left the group in Bonn to take a chair in Berlin.
The "Bonn-Connection's" idea of reunifying instrumental phonetics and phonology can be very easily depicted by Meyer-Eppler's schematic presentation of the Verbal Communication Chain. Fig. 1 is taken from page 2 of his Book "Grundlagen und Anwendungen der Informationstheorie" (Principles and Applications of Information Theory):

Die sprachliche Kommunikationskette

Sprachliche[2] Kommunikation setzt als Expedienten in der Regel ein menschliches Individuum voraus. Das von ihm dem Perzipienten übermittelte Signal ist als Träger sprachlicher Funktionen *Zeichen* kraft seiner Zuordnung zu geistig erfaßten Gegenständen und Sachverhalten[3]. Die Zuordnung selbst ist *beliebig*[4] und Ergebnis einer *Setzung* oder einer besonderen *Übereinkunft* zwischen dem Expedienten und dem Perzipienten; der Zeichencharakter wird dem Signal *verliehen*, er haftet ihm nicht an wie das Symptom.

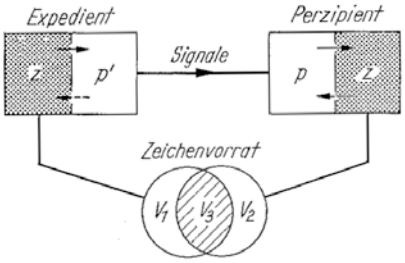


Abb. 1;3. Modell der einfachsten sprachlichen Kommunikationskette. V_1 aktiver Zeichenvorrat des Expedienten, V_2 passiver Zeichenvorrat des Perzipienten, V_3 gemeinsamer Zeichenvorrat

Die einfachste *sprachliche* Kommunikationskette (Abb. 1;3) weist also, im Gegensatz zur Beobachtungskette und diagnostischen Kette, eine *doppelte* Verbindung zwischen den beiden Kommunikationspartnern auf. Neben der realen, mit physikalischen Methoden nachweisbaren

[2] Das Wort *Sprache* (language) wird hier in dem allgemeinen Sinn verwendet, den ihm unter anderen CH. W. MORRIS beilegt (Foundations of the theory of signs. In: International encyclopedia of unified science, Bd. 1, S. 77—137. Chicago: Univ. of Chicago Press 1955).
[3] BÜHLER Sprth S. 28ff.
[4] DE SAUSSURE GSpr S. 79ff.

Fig. 1

The following statement is a citation from the editor's foreword:

> „Aufgabe der Informationstheorie ist es, die Kommunikation von Mensch zu Mensch, die sich als Zeichenverkehr manifestiert, oder die Kommunikation des Menschen mit der Welt, die auf eine Beobachtung hinausläuft, einer quantitativen und strukturellen Erfassung zugänglich zu machen, während die Kybernetik als „science of relations" (N.WIENER) die regulären Verhaltensweisen von hochkomplexen energetisierten „Systemen" (d.h. von informationsverarbeitenden „Maschinen", Lebewesen und Gruppen von Lebewesen) mit mathematischen Methoden studiert" (ohne Seitenzahl).[17]

[16] His wonderful book on "Speech Analysis, Synthesis, and Perception" appeared 1965 in Meyer-Eppler's Series mentioned above; cf. also Schroeder 1967.
[17] Information theory aims to make different types of communication quantitatively and structurally tangible. Communication of humans with each other is manifested in the exchange of signs, while the communication of

This quotation shows the broad communication-theoretical framework in which experimental phonetics and phonology come together. In particular, Trubetzkoy's distinction between relevant and irrelevant features played an important role in Meyer-Eppler's thinking (and teaching) and received a quite new interpretation.
Methodologically, Brücke's concept of 'direct observation' was replaced by the newly established 'authority' of an ***external observer*** ("externer Beobachter"):

> „Die in einer Kommunikationskette sich abspielenden Prozesse können nur von einem außerhalb der Kette stehenden *externen Beobachter* hinreichend exakt beschrieben werden, einem Beobachter, dem *sämtliche* Glieder der Kette zugänglich sind. Zur Beschreibung des Beobachteten bedient er sich einer wissenschaftlichen *Metasprache*, die nicht mit der zwischen dem Expedienten und Perzipienten vereinbarten *Objektsprache* übereinstimmt. Alle informationstheoretischen Ausführungen der folgenden Kapitel sind in der Metasprache des externen Beobachters formuliert" (Meyer-Eppler 1959, S.5f)[18].

We cannot deal here in detail with the very complex relations between the "gemeinsamer Symbolvorrat", which is to be categorically identified on all linguistically relevant levels on the one hand and, on the other hand, the production and reception of the phonetically produced utterance with which a speaker addresses himself to a listener. In Kohler-FS (1993) the first author has discussed these problems by asking the question which ***phonetic facts*** a speaker must create when conducting an act of speech.
There is another point of interest which can be inferred from Fig.1. Whereas the traditional 'narrower' Type of phonetics has clearly always been a small subpart of linguistics, in Meyer-Eppler's schema of the verbal communication chain linguistics becomes only a small, but necessary subpart of phonetic speech science. It is needed in order to be able to specify the *shared symbol inventory* ("Gemeinsamer Symbolvorrat").[19]
Following his untimely death at the age of 47, his work could be continued because of his ongoing, successfully running research projects. This is the main reason why Meyer-Eppler's research group stayed active and continued to grow. Ungeheuer, Schnelle, Heike and others not only continued their work in the following years, but also began working in relatively new areas. The faculty invited guest professors. Max Mangold came from Saarbrücken to brilliantly teach narrow phonetic transcription. Göran Hammarström gave an introduction to the traditions of diachronic and synchronic linguistics and explained his system of linguistic units which was also published in Meyer-Eppler's series as volume V (then edited by the computer scientist Steinbuch who later introduced the term "Informatik", the new German word for Computer Science).
In 1963, Ungeheuer became officially responsible at the institute for „Kommunikationsforschung" [20]. Under his guidance, the electronics lab established by Meyer-Eppler and run by

humans with the world can be reduced to observation. Instead, as the "science of relations" (N. WIENER), cybernetics is the study of regular behavioural patterns found in highly complex energetised "systems" (such as "machines" designed to process information; living creatures and groups of those) using mathematical methods.

[18] The processes taking place in a communication chain can only be described in sufficient detail by an *external observer* who is able to access *all* the links in the chain. In order to describe what he observes he makes use of a scientific *meta-language* which does not correspond to the *object language* agreed upon by the sender ("Expedient") and the receiver ("Perzipient"). Thus, all information-theoretic explanations given in the following chapters are formulated in the meta-language of the external observer.

[19] This can also be inferred from the three footnotes of page 2 (in Fig. 1).

[20] He succeeded in getting Klaus Kohler from Edinburgh who confronted his new colleagues in Bonn with the English tradition of careful phonetic ear training; Helmut Richter came from Zwirner's institute and discussed with the present author the psychological aspects of speech communication; the linguist and specialist in foreign language teaching Franciszek Grucza stayed as a guest scientist from Poland in the institute until he received a leading position at the University of Warsaw (together with a professorship in language teaching).

R. Rupprath developed the first well functioning speech recognition system, the **Device for Automatic Word Identification through Discrimination**, **DAWID**. It was restricted to identify twenty Italian words, the numbers from "zero" to "nove" and command words such as "meno", "piu", "per", "diviso", "uguale", "dacapo" etc.[21]
The presentation of this great success in a small paper (by Tillmann, Heike, Schnelle, and Ungeheuer) at the International Acoustic Conference 1965 in Liège caused BMVtdg (the defence department of the German Government) to contact Schnelle and Ungeheuer offering them more than enough money to start a broad ensemble of application oriented phonetic speech research projects with the final aim of reliable automatic speech recognition of spoken German, but also projects were financed on Speaker Identification and Speaker Verification (cf. Tillmann 1973). This enabled Ungeheuer a tremendous extension of the number of his research projects. In the framework of this application-oriented work financed by the German Government, Meyer-Eppler's early ***electronic*** speech signal processing was, within a short time, replaced by the new much more powerful methods of ***digital*** signal processing. Thus the computer became the single most important instrument in phonetic speech research in Germany.
Soon after Bonn this was the case also in Köln, in Munich, and in Kiel.

3. Looking back on the first century of academic phonetic speech research

Phonetics as speech science in the sense of Menzerath and Meyer-Eppler has become an integral part of the large and complex research field of SLP (Spoken Language Processing)[22]. It therefore makes sense to consider the development of early modern phonetics as outlined above in Section 1 from today's SLP perspective[23]. In the following we consider a set of eight critical points which have been ignored during the first one hundred years of early modern phonetics.

(1) Early phoneticians presumed that in all spoken languages of the world there is no natural oral speech act without an actual utterance produced by a real speaker who utilizes his speech organs according to the pronunciation rules of his language, dialect, and sociolect. However, the simplified view that in an oral speech act one only has to distinguish between ***directly observable utterance*** and the ***verbal meaning*** semantically associated with it mentally was (as we will show) wrong[24]. This becomes evident as soon as we consider that the term 'utterance' is ambiguous at least in two different respects. While it was totally sufficient for linguists such as Bloomfield or Trubetzkoy to assume that "recurrent utterances are alike or partly alike", phonetic speech scientists were responsible for investigating the utterances produced by a speaker as phonetically given facts in order to find out why and how, in a given speech community, recurrent syllables, consonants, vowels, diphthongs etc are alike or different. However, there are two methodological mistakes which we would like to describe and clarify in (2) and (3) below.

[21] More detailed information in Heike-FS (p. 11ff).
[22] This widely accepted term was proposed by Hiroya Fujisaki in direct contrast to NLP (Natural Language Processing), which according to Adrian Fourcin could also be an acronym for "non-spoken language processsing" (pc).
[23] We will not comment on Techmer's thorough and comprehensive approach of a "naturwissenschaftliche Analyse und Synthese der hörbaren Sprache" from today's perspective, as we plan to discuss this in a further paper.
[24] We also find this misleading assumption in Bloomfield's famous definition D1: "An act of speech is an utterance".

(2) First of all, speech acts cannot be separated into isolated phonetic and semantic parts[25]. In reality, the concept of an utterance is ambiguous in a way that has been dealt with extensively and made made clear in TmM: each time a speech act occurs, the external observer registers the material processes taking place throughout the whole communication chain from brain to brain. These material processes can be measured in terms of analogue time functions f(t) which can be digitalised and further processed using the methods of SLP. This type of data remains absolutely transphenomenal for all subjects involved in the speech act, unless there is a microphone and an oscillograph available.

(3) In order to understand how it is possible to establish whether two utterances are alike or partly alike from a categorical perspective it is necessary to distinguish ***between two types of speech acts***. They differ *depending on the* ***intention*** *of the speaker*. In the first case, the speaker produces an utterance (e.g. "xyz") in order to confer the meaning xyz. The speaker and the listener transcend the perceived utterance mentally into the concrete or abstract situations in their worlds, which are semantically governed according to the linguistic structure of "xyz".

In the second case, *the speaker's intention* is to *mean the utterance itself*. This is the case when, for example, Daniel Jones speaks a series of isolated speech sounds into a microphone to record onto audiotape (in multiple *categorically identical reproductions*) in order to demonstrate the category of his cardinal vowels. The IPA vowel symbols are thus ostensively defined by the presentation of these clearly perceivable citation forms.

From a logical viewpoint, the utterance is intended to be **autonymic** (meaning itself). On the other hand, in the first case, we are dealing semantically with a **heteronymic** utterance[26]. In everyday speech autonymically and heteronymically produced utterances differ not only in their semantics, but also in their phonetic forms[27]. Normally, we are immediately able to detect with which of the two intentions a speaker produces his speech utterance.

(4) The term "utterance" is also ambiguous in a second way. We need to distinguish between the time functions registered by the external observer and the symbolic representation of the given sound categories. These are two fundamentally different empiricisms that are logically independent of each other, i.e. contingent (Feigl 1958). The relationship between the two types of speech data is purely empirical and can therefore - as in the case of categorical perception - be reproduced experimentally. Early phoneticians overlooked the difference between logical identity, which is expressed with the equal sign, and empirical identification which must be tested. They were naïve realists. Today, as critical realists, we know that the empirically given correlations between signals and symbols, i.e. time functions and categories, must be revealed by systematic research and have to be made explicit by empirically verifiable relating theories. It is thus no wonder that early instrumental phonetics with its naïve aims was doomed to fail.

[25] This mistake is also made by philosophical speech act theorists, who ask and try to establish how it can be possible that a speaker produces an utterance, "xyz", in order to express meaning xyz. Here the utterance of the speaker is given in quotation marks as if it could exist by itself. The answer that is given is that the semantic meaning of xyz (without quotation marks) arises out of the ***speaker's intention***, which the listeners must recognise in order to understand the speaker. We return to this in the next footnote.

[26] Here language philosophists should ask the question how it can be possible that a speaker produces "xyz" in order he means "xyz" as such itself. As the example of the cardinal vowels shows, it is really the case that the listener understands that Daniel Jones has the intention of demonstrating the phonetic category of a certain cardinal vowel by repeating categorically identical repetitions of that certain vowel.

[27] This can be demonstrated by comparing the words from spontaneous speech with their citation forms.

(5) Today, we can view the early phoneticians as naïve realists in yet another light. They were not aware of the importance of distinguishing between the bottom-up and top-down processes which always play a role during the cognition of autonymically or heteronymically produced utterances. By ignoring top-down processes the early phoneticians believed they were able to directly observe in a bottom-up way the alphabetic form of orally produced utterances in the physical reality. Today, we know that the listener's cognitive system needs to use top-down processes in order to interpret information transmitted in the acoustic signal by empirically verifying the L1 categories acquired during child language acquisition. The same is true for any phonetically acquired L2 knowledge.

(6) With the arrival of the DAWID project, the word as a category in the mental lexicon (Meyer-Eppler's "Symbolvorrat") suddenly became a central phonetic unit[28], because lexical units play a crucial role not only in the listener's perception of a natural speech act, but also in automatic speech recognition. A very good example of this is the **M**unich **Au**tomatic **S**egmentation tool, **MAUS**, developed at the IPS[29].

(7) It is only possible to investigate the actual realisations of the phonetic form of the words of an utterance (compared with their citation forms) if we know the canonical form of the given utterance. For example, in addition to the acoustic signal of an utterance, Munich MAUS requires a canonical form or an orthographic transliteration plus the name of the language in order to produce a fine phonetic transcription and acoustic segmentation.

(8) Our example of MAUS-annotated sound segments can be used to demonstrate our final point. This relates to the fact, neglected by the early speech physiologists, that natural speech articulation is typically produced far too quickly to be processed adequately by bottom-up processes, as coarticulated CvCvC sequences (C standing for consonants, including none, v for a vowel or diphthong), are characterised by their C prosody (see TmM's distinction between slow A, rhythmic B and fast C prosody (p. 39 ff))[30].

Brücke treated the letters of the alphabet as isolated lexical units and subjected them to direct observation in their citation forms. This explains how he came up with the absurd idea of improving the orthography by adapting it to actual pronunciation. The actual phonetic notation of such real pronunciations, such as those computed by MAUS, cannot possibly be used to deliver a useful orthographic representation. The cognitive systems of speakers and listeners have learned to process what from a lexical point of view could well be called "modifications phonétiques du langage". In order to develop empirically verifiable theories explicating the functioning of all the processes that must run effectively in an act of successful speech communication in the interdisciplinary field of SLP it becomes necessary (in the words of another title of Rousselot) to use more than one "méthode graphique appliqué à la recherche des transformations inconscientes du langage".

Indeed, Francis Nolan and the first author, both quite well trained in narrow phonetic transcription, had no idea what their tongues where doing when producing the combined EPG and EMMA data in fast spoken German or English, respectively, for Barbara Kühnert's PhD thesis on alveolar-velar [t-k] assimilation (Kühnert 1994). Thus, they were both quite surprised when they saw the very complex systematic picture presented by their student after she

[28] Cf. Kohler (1998) and Tillmann (1998)'s contributions to the conference "The word as a phonetic unit" organised by Pompino-Marschall in Berlin.
[29] https://clarin.phonetik.uni-muenchen.de/BASWebServices/#/services
[30] Cf. Phil Hoole's demonstration of A, B and C prosody:
http://www.phonetik.uni-muenchen.de/~hoole/kurse/artikul/abc.pdf

had looked through all the data as a scientific external observer in order to analyse the tongue behaviour[31] of the German and English subjects.

PS:

The idea to write this paper was born two years ago in Venice when the first author had given, invited by Jonathan Harrington, one of his penultimate looking-back talks at VIU, and in the audience the second author proposed to compare the early results (such as intrinsic pitch of vowels, the P-center, or those other ones mentioned above in footnote 7) with the results of today's research in more detail. As one can see this idea is still reflected in our section 3 above.

Here, the first author would like to add that he is going to prepare a fourth section of the paper (to be presented in Dresden) in order to discuss the question which role some really innovative ideas in the recent past of phonetic speech research could or should play in future SLP work.

He believes that there are at least four important topics that should not be neglected in future ***application oriented*** SLP-research:

(i) Is the proposal to develop a **CPT** (***C***omplete ***P***honetic ***T***heory) of spoken German[32] - making use of the huge available databases collected in Verbmobil and BAS[33] - definitely out of date?

(ii) Why has Hartmut Pfitzinger's amazing success in measuring the variation of local speech tempo and to depict it graphically like a smoothed F0-contour[34] been more or less ignored in SLP (up to now)?

(iii) What is the future of speech ***synthesis-by-analysis***[35]?

(iv) How could in the near future phonetic speech signal processing contribute to improve the pronunciation of L2-speakers in Chinese[36], in German[37] or other foreign languages[38]?

[31] While in the EPG-Data there **is** either an alveolar [t]-contact **or is not** before the alveolar contact of the following [k]-word, the EMMA-date show all degrees of tongue tip movements from zero to full contact (as in the [k-k] control items).

[32] Tillmann, H.G. & B. Pompino-Marschall: Theoretical Principles Concerning Segmentation, Labelling, and Levels of Categorical Annotation for Spoken Language Database Systems, EUROSPEECH 1993, pp. 1691ff.

[33] Hess, W., Kohler, K., Tillmann, H. G. "The PhonDat-Verbmobil Speech Corpus", Proceedings of EUROSPEECH, pp. 863-866, Madrid 1995

[34] Pfitzinger, H.R : Phonetische Analyse der Sprechgeschwindigkeit. , FIPKM 38, pp. 117-264, 2001. (This outstanding milestone of phonetic speech research should be printed as a book and also translated into English.) (one can download it from http://www.ipds.uni-kiel.de/hpt/)

[35] Tillmann, H. G.; Pfitzinger, H. R.: Parametric High Definition (PHD) Speech Synthesis-by-Analysis. Proc. ICSLP 2000, vol. III, pp. 295-297. Beijing 2000

[36] Tillmann, H.G.; Pfitzinger, H.R. (2004): Applying the Munich Parametric High Definition (PHD) Speech Synthesis System to the Problem of Teaching Chinese Tones to L1-Speakers of German. Proc. of the Int. Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages (TAL), pp. 185-188. Beijing 2004

[37] Bissiri, M.P.; Pfitzinger, H.R.; Tillmann, H.G. (2006) Lexical Stress Training of German Compounds for Italian Speakers by means of Resynthesis and Emphasis. Proc. of the 11th Australasian Int. Conf. on Speech Science and Technology (SST 2006), pp. 24-29, 2006

[38] Tillmann, H.G.; Pfitzinger, H.R. (2004): The Development of an Advanced SLP-based System for the Individual Learning and Fast Training of Speaking Skills in a New Foreign Language. Proc. of the InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, pp. 17-20. Venice 2004

References

Bell, A. M.: Visible Speech. Universal alphabetics or self-interpreting physiological letters for the writing of all languages in one alphabet. London/New York 1867

Bissiri, M.P.; Pfitzinger, H.R.; Tillmann, H.G. (2006) Lexical Stress Training of German Compounds for Italian Speakers by means of Resynthesis and Emphasis. Proc. of the 11th Australasian Int. Conf. on Speech Science and Technology (SST 2006), pp. 24-29, 2006

Brücke, E. W. v.: Untersuchungen über die Lautbildung und das natürliche System der Sprache. Sitzungsber. der königl. Akad. der Wissenschaften. Mathem.-Naturwiss. Classe II, 182-208, Wien 1849

Brücke, E. W. v.: Grundzüge der Physiologie und Systematik der Sprachlaute für Linguisten und Taubstummenlehrer. Wien !956

Feigl, H.: The „"Mental" and the „"Physical", in Volume II of *Minnesota Studies in the Philosophy of Science: Concepts, Theories, and the Mind-Body Problem, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell*, Univ. Minnes. Press, 1958.

Hammarström, G.: Linguistische Einheiten im Rahmen der modernen Sprachwissenschaft. Berlin-Göttingen-Heidelberg 1966

Heike, G.: Sprachliche Kommunikation und linguistische Analyse. Heidelberg 1969

Hess, W., Kohler, K., Tillmann, H. G. "The PhonDat-Verbmobil Speech Corpus", Proceedings of EUROSPEECH, pp. 863-866, Madrid 1995

Kemp, J. Alan. (1994). Phonetic transcription: History. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (Vol. 6, pp. 3040–3051). Oxford: Pergamon.

Koerner, E. F. K. and R. E. Asher (Eds.): Concise History of the Language Sciences from the Sumerians to theCognitivists. Cambridge UP, 1965

Kohler, K. J.: The disappearance of words in connected speech. ZAS Papers in Linguistics 11, 21-34, Berlin 1998

Kühnert, B. : Die alveolar-velare Assimilation bei Sprechern des Deutschen und des Englischen - kinematische und perzeptive Grundlagen. München 1964

Menzerath, Paul und A. de Lacerda: Koartikulation, Steuerung und Lautabgrenzung. Berlin-Bonn 1933

Meyer, E. A.: Beiträge zur deutschen Metrik. Die neueren Sprachen 6, 1-37, 122-40, 1897

Meyer-Eppler, W.: Elektrische Klangerzeugung: Elektronische Musik und synthetische Sprache. Bonn 1949

Meyer-Eppler, W.: Grundlagen und Anwendungen der Informationstheorie. Berlin-Göttingen-Heidelberg, 1959

Meyer-Eppler, W. und G. Ungeheuer: Die Vokalartikulation als Eigenwertproblem. Zeitschr. für Phonetik 10, 245-257, 1957

Pfitzinger, H. R : Phonetische Analyse der Sprechgeschwindigkeit. FIPKM 38, pp. 117-264, München 2001

Pompino-Marschall, B.: Einführung in die Phonetik. Berlin 2003

Rousselot, P.J.: Les modifications phonétiques du langage. Revue des patois gallo-romans 4, 65-208, 1981a

Rousselot, P.J.: La méthode graphique appliqué à la recherche des transformations inconscientes du langage. Revue des patois gallo-romans 4, 209-13,1981b

Schroeder, M. R.: Determination of the geometry of the human vocal tract by acoustic measurements. JASA 41, 1002-1010, 1967

Scripture, E. W.: The Elements of Experimental Phonetics, New York/ London 1902

Scripture, E. W.: Referate. Zeitschrift für Experimental-Phonetik I (3/4), 171-88, 1932

Techmer, F.: Naturwissenschaftliche Analyse und Synthese der hörbaren Sprache. Intern. Zeitschr. f. allgemeine Sprachwissenschaft I, 69-170, 1884

Tillmann, Hans G. Das Subjekt und seine individuelle Identität im phonetischen Kommunikationsprozess. IPK-FB Band 48, Buske, Hamburg 1973

-- (TmM (mit Phil Mansell)): Phonetik -- Lautsprachliche Zeichen, Sprachsignale und der lautsprachliche Kommunikationsprozess. Klett-Cotta, Stuttgart 1980

-- (EmP): Early modern phonetics, especially experimental and instrumental work. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (Vol. 6, pp. 3040–3051). Oxford: Pergamon. 1994/2005[39]

-- (Kohler_FS): Kleine Phonetik und Große Phonetik. Phonetica 52, 144-159, 1995

-- Why the word should become the central unit of phonetic speech research. ZAS Papers in Linguistics 11, 1-20, Berlin 1998

-- (Heike_FS): Von der kleinen zur großen Phonetik: Wie es an der Universität Bonn in der Mitte des letzten Jahrhunderts zu einem Paradigmenwechsel in der phonetischen Sprachforschung kam. Hallesche Schriften zur Sprechwissenschaft und Phonetik, Band 45, Frankfurt 2013

Tillmann, H. G., G. Heike, H. Schnelle und G. Ungeheuer: DAWID I – Ein Beitrag zur Automatischen Spracherkennung, Beitrag A12, 5. Intern. Akustikkongress, Lüttich 1965

Tillmann, H.G. & B. Pompino-Marschall: Theoretical Principles Concerning Segmentation, Labelling and Levels of Categorical Annotation for Spoken Language Database Systems, Proc. of EUROSPEECH 1993, pp. 1691ff, Berlin 1993

Ungeheuer, E.: Wie die elektronische Musik ´erfunden´ wurde..., B. Schott´s Söhne, Mainz 1992

Ungeheuer, G. : Elemente einer akustischen Theorie der Vokalartikulation. Berlin 1962

Vennemann, Th. und Joachim Jacobs: Sprache und Grammatik. Darmstadt 1982

Zierdt A., Hoole P., Honda M., Kaburagi T., Tillmann H.G.: Extracting tongues from moving heads, in Proc. of the 5th Seminar on Speech Production: Models and Data, pp 313-316, Kloster Seeon 2000

[39] Also p. 401-416 in Koerner et. al (1995)