LEAVING THE IVORY TOWER: HOW REAL-LIFE EVENTS HAVE IMPACTED PHONETICS THROUGHOUT HISTORY

Angelika Braun
University of Trier, Germany
brauna@uni-trier.de

Abstract: Academic research has a reputation of being somewhat detached from reality. The metaphor of researchers being confined to the ivory tower, unable to leave it and largely unable to relate to the real world is indicative of this view. However, this view might not always be adequate. This contribution provides a number of counter-examples from the field of phonetics. Several real-world events are discussed which have had a very direct and measurable impact on phonetic research. The first example goes back to the time before phonetics was known as “phonetics”. It concerns the change in attitude towards autopsies that unleashed the curiosity about human anatomy, which permitted the evolution of phonetics in the first place.

The other examples are more recent. They are often related to criminal cases in which phoneticians have acted as experts. One example is research on memory for voices, which was prompted by the kidnapping of the Lindbergh baby in 1932. A further example concerns the development of the spectrograph, which was part of the American WW II wartime strategy. The next major event affecting phonetic research was the impeachment trial of Richard M. Nixon, in the course of which a clandestine recording made in the White House had to be authenticated. In 1989, the supertanker Exxon Valdez grounded in Prince William Sound, Alaska. In this case as well as a similar one in Germany, the question of whether the captain was drunk at the time of the incident played a major role in the court cases. Finally, the various attempts at establishing a German pronunciation standard are reviewed in light of political developments in recent history.

1 Introduction

When academics try to explain to ordinary people what they are doing, they tend to meet with limited understanding. It is not easy to communicate why for instance proving a certain mathematical formula or studying deep-sea marine life should be allotted large grants. Compared to these topics, the phonetic sciences have always had closer ties to the real world. Early examples are the teaching of the deaf, the teaching of pronunciation to L2 learners or, more recently, the fitting of cochlear implants. In some instances, the approach has turned the default case of research planning upside down: instead of generating a theory and testing it against observation [58: 390], there can be no doubt that a whole wave of phonetic research on particular topics was prompted by real-world events. Some selected examples will be covered in this contribution. The aim of this paper is thus to demonstrate that sometimes new steps in phonetic research were brought about by specific real-world events as opposed to being theory-driven.

2 An early example – autopsies as a prerequisite for phonetic research

2.1 Some facts

Some very early developments in phonetics were influenced by historical events. Strictly speaking, what we now consider the beginnings of phonetics in the 16th century would not have been
possible without a change of mind with regard to autopsies. In third-century B.C. Egypt, autopsies, sometimes even vivisections, were used to gain insight into human anatomy. Herophilos of Chalcedon (335–280 B.C.) and Erasistratos of Ceos (310–250 B.C.) were the most famous representatives of that tradition [17: 93]. Unfortunately, their knowledge was lost to future generations in a large fire which destroyed the famous library of Alexandria in 47 B.C. [2: 4].

Claudius Galenus (129–199), who is considered one of the major anatomists in European antiquity, gained much of his knowledge about human anatomy in his capacity as “team physician” of the Roman gladiators and by dissecting animals. However, he did not venture to touch human corpses [2: 5]. In the Middle Ages, all this knowledge was lost. Autopsies were considered unethical, and especially the Church was strictly opposed to them. One of the most outspoken representatives of the Church was Augustinus (354–430) who harshly condemned autopsies. A ruling by Pope Bonifaz VIII in 1299 was interpreted along those same lines [17: 93-94].

The Renaissance brought a change in attitude towards autopsies. They became popular at Italian universities in particular. While they were carried out in secrecy at first at the University of Salerno, they soon developed into spectacula where chairs were set up in a manner reminiscent of a theater.¹ While a “dissector”, often a barber, did the actual opening up of the body, the professor was seated in an elevated chair – removed from the smell – and recited Galenus’ works. An assistant was supposed to point to the structures which were mentioned by the professor [17: 94].

![Figure 1. Typical scenario of an autopsy in Renaissance times: Mondinus (1275–1326), a well-known anatomist of his time, teaching anatomy. Wood carving in “Anathomia Mundini”, Geneva 1519 (reprinted with permission from [57: 425]).](image)

¹ Please note that, till today, the operating room is often called the “operating theater”.

48
Tickets for these shows were extremely popular even though the narrations were usually unrelated to what was actually happening on the dissection table. This “hoax” was exposed in 1535 by a young student in Paris by the name of Andreas Vesalius (1514–1564), who ventured to step up against the famous anatomist Jacobus Sylvius\(^2\). Vesalius is considered to be the father of modern anatomy. He was born in Flanders and became an outspoken critic of Galenus, rather because he realized that the latter had not gained his knowledge from dissecting humans but animals instead. He received a professorship in Padova at the age of 23 and was very successful, because he did both the dissecting and the narrating, with these two corresponding for the first time ([2: 9], [17: 94-95]).

### 2.2 How phonetics comes into play
Vesalius is considered the founder of scientific anatomical studies [2: 9] and is thus at the same time a pioneer in phonetics in the sense that he expanded the knowledge about the anatomy of the organs involved in speaking and listening. He overcame the Galenic tradition and relied on data gained from dissecting human bodies. This brought about essential knowledge gain in the anatomical domain. The physiology of speaking and hearing, however, remained largely in the dark – the exact mechanism of vocal fold vibration, let alone the physiology of hearing, was not described until well into the 20\(^{th}\) century [53; 55; 59].

### 2.3 Conclusion
Anatomical studies are a prerequisite for phonetic research. Autopsies have always been the prime source of knowledge about human anatomy including the vocal organs. Not even the development of MRI has been able to replace the immediate inspection of the structures involved. Thus, the findings of the early anatomists were indispensable for any phonetic description of sound production and perception.

### 3 The kidnapping of the Lindbergh baby

#### 3.1 Some facts
On 1 March 1932, the 20-month-old son of Charles and Anne Morrow Lindbergh disappeared from his bed. Charles Lindbergh was a world-famous aviator known for having crossed the Atlantic Ocean solo in a small passenger plane in 1927. In the boy’s room, a note was found which was written in poor English and demanded $ 50,000 in ransom. In the course of the next five weeks, the parents received another twelve ransom notes. Some sources also mention phone calls made by the kidnapper and answered by Charles Lindbergh [58]. On 2 April 1932, ransom money was handed over by John F. Condon, who acted as an intermediary for the Lindberghs. On that occasion, Charles Lindbergh had a chance to listen to the kidnapper’s voice from a distance. However, the whole transaction turned out to be a hoax. The child was not found where the man who received the ransom said he was. Instead, the corpse of Charles Augustus Lindbergh, Jr. was found close to the Lindbergh home on 12 May 1932. It turned out that he had been killed shortly after the kidnapping [61]. Bruno Richard Hauptmann, a carpenter with German roots, was charged with the kidnapping and murder of the child. The evidence against him was mainly circumstantial. Part of it consisted of Charles Lindbergh’s testimony on the kidnapper’s voice [46]. Hauptmann denied any involvement in the case, and yet he was

\(^2\) The Sylvian fissure, which separates the frontal from the temporal lobe in the brain, was named after Sylvius.
convicted and – after several unsuccessful attempts to appeal the verdict – eventually electrocuted on 3 April 1936.

### 3.2 How phonetics comes into play

The conviction of Bruno Hauptmann rested, to a large extent, on the testimony by Charles Lindbergh. He testified to having recognized Hauptmann’s voice as that of the man whom he had heard on the occasion of the money handover when he was given a chance to listen to him shortly before taking the stand. His identification was based on nine syllables of speech (“Hey Doctor, over here, over here”) with a time lapse of close to three years between familiarization and recognition. Still, the jury went along with his statement and handed down a guilty verdict. The defense questioned the recognition procedure. It was pointed out that there was no scientific evidence to the effect that recognition was possible under the given circumstances. Prompted by this case, the psychologist Frances McGehee of Johns Hopkins University, Baltimore, MD carried out pioneering work on memory for voices [38], [39], [56]. She conducted a series of experiments some of which simulated the setting of the Lindbergh case. In doing so, she addressed most of the principal questions which are still at issue today when a so-called voice line-up is to be carried out: the a priori possibility of choosing a number between one and five (position effect), the effective reduction of the line-up size by including a voice which is very dissimilar to the others, the issue of memory type (incidental vs. intentional) and presentation format (live or taped), the effect of the number of voices present in the line-up, the effect of a target-absent scenario on the recognition rate, the effect of voice disguise, the role of race and gender in recognition, the performance of musically and linguistically trained listeners, and, last but not least, the time lapse.

McGehee found that recognition accuracy deteriorated from a recognition rate of 83% one day after familiarization with the voice to a mere 13% five months later. She also found that recognition rates go down if the number of voices in the line-up is increased. While disguise presented a problem to listeners, a foreign accent did not. Men generally outperformed women at recognition tasks [38].

Her research has been criticized for reverting in part to a within-subjects design by assigning some listeners to more than one group [58: 391]. However, even though the research paradigm may not comply with today’s standards in all respects, she has to be given credit for identifying the relevant issues and for addressing them in a large number of experiments.

### 3.3 Conclusion

To this day, there is hardly any publication on the subject of speaker identification by lay listeners that does not pay tribute to McGehee’s pioneering work and the ties to the Lindbergh case (cf. e.g. Clifford [13: 375]; Deffenbacher et al. [15] and especially Yarmey et al. [58]). To this date, the issue of victims and witnesses recognizing a perpetrator’s voice remains largely unresolved. The existing studies are largely “academic”, and there is a paucity of projects with a clearly defined forensic perspective. Specifically, it would be highly desirable to develop an independent test which could provide a measure of a given witness’s general ability to recognize voices. That would solve the principal question behind any line-up in which an identification is made – to what extent can the witness be relied on? In the meantime, the issues raised by McGehee in the 1930s should be addressed and re-investigated in terms of time lapse, memory type and voice characteristics in particular.
4 The Sound Spectrograph

4.1 Some facts
During World War II, the American military worked closely together with the nation’s leading think tanks, among them Bell Labs at Mahwah, N.J. After Japan entered the war with the attack on Pearl Harbor on 7 December 1941, Japanese and German submarines posed a continuous threat to the US navy. It was of prime strategic importance to be able to trace the movements of the submarines in order to know what the enemy forces were up to. There was nothing to go by except the radio communication between the vessels, which was intercepted by the US forces on a regular basis.

4.2 How phonetics comes into play
Until the 1930s, the kymograph and the cathode ray oscilloscope were the main instruments that were available for visualizing different aspects of speech behavior. This changed in the early 1940s, when Ralph K. Potter developed the Sound Spectrograph. According to Koenig [31: 1], a prototype was finished in the spring of 1941. Potter filed the patent application on 14 April 1942. In the course of testing the new device it must have occurred to Potter that the spectrograms did not only show properties of different sounds, but they could also be used to depict reflections of different vocal tracts, i.e., differences between speakers. This is hinted at, but never elaborated in various post-war papers by Potter and his colleagues (cf. Potter [43], [44], and Kopp & Green [34]). The spectrograph gained strategic importance and was developed as a secret device which could be used to produce “voice prints” of the German and Japanese radio operators and thus to identify them (cf. Braun [7], [8]). A task force was formed that developed a framework for implementing the spectrograph as a speaker identification device. Figure 1 shows an excerpt from a secret memorandum entitled “Field Uses of Voice Identification” of 31 May 1944 which was authored by Ralph Potter. It evidently was part of a preliminary version of the final report submitted to the military on 18 August 1944 [21].

The voices of individuals differ. These differences in some degree are familiar to the ear and can be recorded and classified by the sound spectrograph. Such records and classifications may be used as a basis for voice identification which, with radio direction finding methods, would make it possible to follow movements of enemy personnel using radio telephone facilities and hence to trace changes in disposition of enemy units. The use of voice for this purpose is analogous to that which has already been developed for manual telegraph.

Figure 2. Facsimile from a memo of 31 May 1944, prepared by Ralph Potter [42: 1].

In the final report that passage reads as follows (cf. Figure 3):
By detecting the characteristic differences in the way individual radio telegraph operators send telegraph signals and correlating these with direction finding data, it is possible to follow the movements of these operators from place to place and therefore determine indirectly the movements of enemy groups with which they are associated. Similarly, by identifying the speech of individuals talking over radio telephone circuits, it should be possible to associate these identifications with direction finding data and to determine the movements of enemy troops or other mobile groups.

Figure 3. Facsimile from Gray & Kopp ([21, p. 1])

This application of the sound spectrograph was not published until well after the war. In the meantime, the spectrograph was said to have been developed to enable the deaf to communicate over the telephone. This is completely unrealistic but was used as a cover for many years and even continued in the famous monograph by Potter, Kopp and Green [45].

4.3 Conclusion

In times of war, the ties between military goals and science were particularly close. This sped up the development of the spectrograph, and it prompted research into the potential of spectrograms displaying speaker specificity.

It seems that the end of World War II put a preliminary halt to the spectrograph as a tool in voice comparison. After all, the relevant documents were still classified. Soon after their declassification in 1960, Lawrence Kersta, an engineer who had also worked at Bell Labs, published the “voiceprint” method of speaker identification, thereby misrepresenting Potter and his colleagues [8], [29], [30]. This, in turn, brought about more studies on the “voiceprint” method of forensic voice comparison (cf. e.g. Braun [7]). To this day, the software tools which are based on the sound spectrograph are still the principal means of visualizing and analyzing human speech.

5 The Watergate Affair

5.1 Some Facts

On 17 June 1972, five men were caught breaking into the Democratic Party headquarters in a building complex called Watergate in Washington, D.C. It turned out that they had not only stolen sensitive documents but also planted surveillance devices in order to assure the re-election of President Richard Milhous Nixon, who was a Republican. The question then was if and when the President had learned about this criminal offense which had been masterminded by the so-called Committee to Re-Elect the President (C.R.P.). The investigation was led by John Sirica, then District Judge in Washington, D.C.. Congress was also looking into the affair: On 25 May 1973, the Senate Select Committee appointed a Special Prosecutor by the name of Archibald Cox. On 16 July 1973, the former Presidential Appointments Secretary Alexander Butterfield testified that President Nixon had had all conversations taped that had taken place in the Oval Office since 1970, including those conducted over the telephone [1: 857]. The controversy revolved around one conversation between the President and his then Chief of Staff, John Haldeman, on 20 June 1972. On 11 April 1974, the House Judiciary Committee subpoenaed 42 tapes, but Nixon refused to supply them, invoking Executive Privilege [1: 876]. In what

3 Note the reserve that is contained in the wording “should be possible” for speaker identification as compared to “is possible” with respect to identifying telegraph signals.
has come to be known as the Saturday night massacre, President Nixon fired his Attorney General Richardson for his refusal to fire Cox and had his successor fire Cox instead, who was replaced by Leon Jaworski [1: 866]. Judge Sirica then subpoenaed 64 tapes, and once again the White House did not comply. The issue of the tapes went all the way to the Supreme Court, which rejected Nixon’s reasoning by an 8:0 vote on 24 July 1974.

Eventually, the tapes, including the one of interest here, were served by the White House. On one tape, there was an 18 ½ minute gap which contained no speech but only low-frequency hum. The refusal to provide the tapes had in the meantime prompted an impeachment trial against Nixon. It was formally initiated on 6 February 1974 [1: 873]. Richard Milhous Nixon did not wait for the outcome. He resigned on 9 August 1974.

5.2 How phonetics comes into play

The million-dollar question in conjunction with the mysterious 18 ½ minute gap in the recording of 20 June 1972 was, of course, whether it was accidental or the result of a volitional attempt to obstruct justice. The President’s personal secretary, Ms. Rose Mary Woods, offered an explanation: On 27 November 1973, she testified that she might have inadvertently erased about five of the 18 ½ minutes in question when she was trying to transcribe the tape [1: 868].

Judge John Sirica proceeded to appoint a panel of six senior figures in audio engineering, physics, and also phonetics, who were to analyze the ominous 18 ½ minutes. They were Richard H. Bolt of Bolt, Beranek, and Newman, Inc., a former professor of Acoustics at M.I.T., James Flanagan, Head of the Acoustics Research Department at Bell Labs, John McKnight, a private consultant, formerly with Ampex Corporation, Thomas G. Stockham of the University of Utah, formerly at M.I.T., Mark R. Weiss, Vice President for Acoustics Research at the Federal Scientific Corporation (F.S.C.), and Franklin Cooper of Haskins Labs (cf. [3: C.1-C.7]). This panel started working on 30 November 1973 and produced a final report on 31 May 1974. The reason these senior figures in acoustics, phonetics, and audio engineering were assembled is that this evidently was the first time that an investigation of this nature was carried out in the US, and probably worldwide. Never before had this type of question arisen, at least not in such a high-profile case. The experts were supplied with a total of nine tape recorders (seven SONY 800B which had been used to record the conversations in the Oval Office, one UHER 5000 with a foot pedal from Ms. Woods’ office, and one UHER 5000 which was in use by the Secret Service) and the famous tape. The following questions were addressed in the investigation:

- Is the recording in question an original or an – edited – copy which contains erasures or splices?
- Where does the buzzing noise in that 18 ½ minute section originate from? Was it recorded continuously or in several chunks?
- Can it be determined whether there was speech “underneath” the buzzing sound? If so, can the speech be recovered? (cf. [3: 1]).

The authors of the report describe in detail the methods they applied (cf. [3: 7-20]). They range from critical listening to acoustic analyses of the on- and off-transients of the record head and

---

4 Incidentally, all their work was carried out under the supervision of US Marshalls, who never left the evidence unattended (cf. [3: B.1]).

5 The foot pedal is used to start and stop the machine when transcribing so that the transcriber does not have to take his/her hands off the typewriter. It cannot be used, however, to erase and record over.
the erase head of the tape recorders in question and “magnetic development”, i.e., the application of ferrofluid to the tape in order to render the magnetic imprints visible. These methodo-
logical elements have since become standard procedure for the authentication of analogue audio tapes (cf. [29], [30], [32]).

The panel stated its conclusions unanimously and with no lack of clarity. They concluded that the tape was an original which had, however, been tampered with. This conclusion was primarily based on acoustic phonetic analysis and the inspection of magnetic patterns on the tape. Between five and nine erasures had been willfully executed.

The buzzing noise was determined to originate from a power line leakage. This finding rested on precise frequency analysis. It was established that there must have been speech signal before the erasures, but the experts were not able to recover it. This could be concluded from an analysis of the short snippets which were located in the 55ms gap between the erase head of and erase head on marks (cf. [3: 15-16]). Short as they were, the acoustic structure within those snippets clearly showed that language had been contained on those parts of the tape before they were erased.

The experts also checked the possibility of accidental erasure, and they came to the conclusion that this was basically impossible because “manual operation of a keyboard control is required to put the recorder into erase-record mode” [3: 44]. On 18 February 1974, the Nixon camp supplied the panel with an unsolicited report by Allan D. Bell, president of Dektor, a small company specializing in surveillance equipment [47]. That report stated that it would have been possible for Ms. Woods to erase part of the tape accidentally by stretching to pick up the phone while her foot remained on the tape recorder foot pedal. The panel of experts refuted that, and it was not followed up in court.6

5.3 Conclusion

Authenticity analyses have since become an integral part of forensic phonetic work (cf. [5], [10], [20], [24], [32], [33], [37]). Most publications covering this aspect point to the Watergate affair as the starting point of forensic authenticity examination. The fact that this was the first time that the question of authenticity arose in a high-profile case has helped to establish the methodology used by, e.g., Koenig [32] and Boss et al. [4] in later cases.

In the meantime, methods have changed due to digitization [33]. Therefore, this is now history, unless an old case resurfaces, as just recently happened in Germany, and the reports from the 1990s are re-examined [60].

6 About drunk drivers and drunk seamen

Until the 1980s, studies on the effect of alcohol on speech were scarce. The bulk of research was medical by nature. Based on a thorough literature search, Chin and Pisoni [12] count a total of 22 papers on alcohol and speech between 1915 and 1982 (p. 167). There were two main reasons why that changed dramatically in the 1980s.

6.1 Drunk drivers

In the early 1980s, the US automobile industry was trying to install a safeguard against driving while intoxicated into its vehicles. They were hoping to develop a device which would determine whether or not the driver was sober, based on a short sequence of speech. If the algorithm

6 Decades later, in 1995, a memorandum was found in Ms. Woods’ papers which confirmed that the erasure was volitional [26].
detected effects of alcohol, the engine would not start. As is quite evident to any phonetician, such a device is not a realistic option for a number of reasons, among them that there is no simple one-to-one relationship between speech behavior and the (degree of) intoxication (see below). In fact, it was never implemented. However, General Motors Research Laboratories had commissioned researchers at Indiana University in Bloomington to determine the possibility of detecting whether a potential driver was intoxicated and, as a consequence, putting a lock on the ignition [12: 168]. This resulted in a certain expertise by those researchers on the subject of alcohol and speech when the next incident happened [40].

6.2 A drunken captain - the grounding of the Exxon Valdez

6.2.1 Some facts
On 24 March 1989, an American oil tanker, the Exxon Valdez, grounded on Bligh Reef in Prince William Sound, Alaska. Some 10,836,000 US gallons of crude oil leaked into the environment [12: 285]. About 2,100 kilometers of coastline were polluted; hundreds of thousands of birds were killed. It is considered the worst oil spill worldwide to date. The damage amounted to billions of dollars. The ensuing investigation revealed that the captain, Joseph Hazelwood, had not been on the bridge during the critical hours. Instead, he had withdrawn to his quarters, leaving the vessel to the third mate Gregory Cousins to maneuver through the difficult territory. Further, it turned out that Captain Hazelwood had a history of alcohol abuse. He had lost his driver’s license because of repeated DUI (“driving under the influence”) offenses and admitted to having had a couple of drinks earlier that evening. It wasn’t until nearly 10 hours after the incident that blood and urine samples could be taken from the captain. Hazelwood, it turned out, had a blood alcohol concentration (BAC) of .061% at that point in time. Witnesses testified that his breath had smelled of alcohol that night, but they also said that they had not observed any signs of intoxication in his behavior although his eyes had been watery [12: 286].

Joseph Hazelwood was fired by Exxon and charged with three counts of misdemeanor (reckless endangerment, operating a vessel while intoxicated, and negligent discharge of oil) and one felony (criminal mischief; cf. [12: 292-293]). He denied all charges and went to trial in February 1990. In the criminal trial, he was acquitted of all charges but negligent discharge of oil, but he was also facing a civil trial because local fishermen and property owners had filed suit against him personally as well as against Exxon, asking for damages.

Incidentally, a similar accident had taken place in Germany three years prior to the Exxon Valdez disaster: on 3 January 1986, the British MS Waylink collided with the Panamanian tanker Brady Maria on the Elbe River. About 260 tons of residual fuel oil were spilled into the water and caused major damage to the marine wildlife. It turned out that the pilot on the Waylink, who was known for his alcohol addiction, was completely drunk when tested shortly after the incident (BAC > .2%). Speech samples from the pilot around the time of the collision were available for analysis and prompted a research project [36].

6.2.2 How phonetics comes into play
In the course of the Exxon Valdez investigation, it occurred to the authorities that there were a number of recordings of Captain Hazelwood’s speech on marine radio communication tapes at various points in time: He had talked to the Coast Guard 33 hours and then again one hour prior to the grounding, a few minutes after the incident, and one hour as well as nine hours after the grounding. The question arose whether it was possible to establish whether he was inebriated on any of these occasions based on these samples of his speech, and if the level of intoxication could be determined. The National Transportation Safety Board (NTSB) investigated the case. Three groups of experts were consulted: The NTSB’s own audio experts Brenner and Cash [9], Mark B and Linda C. Sobell of the Addiction Research Foundation in Toronto, Canada, and a
group of researchers from the Speech Research Laboratory at Indiana University ([27], [41], [49]).

The first and the last group of experts published their results. They concluded that “the speech changes exhibited by the master of the Exxon Valdez were consistent with changes observed when a talker is impaired by alcohol” [12: 230].

A controversy arose about the admissibility of this type of evidence under US law. Tanford et al. [49] argued in favor, stating “that the scientific evidence of intoxication in the Exxon Valdez case was sound and reliable” [12: 297], whereas Hollien [25] concluded “that the speech and voice data did not support the findings of the NTSB and its consultants that Hazelwood was intoxicated around the time of the grounding” [12: 300].

That same question of whether (the degree of) inebriation could be established from speech samples had come up in a German court two years earlier in conjunction with the case mentioned above. However, there was not much previous research that the German experts in this case could rely on ([25], [27], [36]). Both these cases prompted a major research project by voice experts from the Federal Criminal Police Office (BKA) on the effects of alcohol on human speech.

In the German case, the pilot’s license was revoked for five years whereas Hazelwood’s was suspended for nine months only. However, he was sentenced to $ 50,000 in damages and to 1,000 hours of community service [12].

6.3 Conclusion

Even though the voice evidence did not play a major role in either the criminal or the civil trial, mainly because the prosecution stated that it was made aware of the findings too late [12: 292]. Chin and Pisoni [12] summarize that “the audiotape recordings of radio transmissions between Captain Hazelwood and the Coast Guard Vessel Traffic Center were the sole objective physical evidence from the time around that of the grounding” (p. 312). Even though the voice evidence did not decide the case, it did set off a series of scientific experiments which have put researchers in a position to argue more reliably in court. After having carried out extensive research in this domain, the BKA group concluded that speech is a sensitive indicator of intoxication. In fact, it proved more sensitive than some of the common neurological tests such as walking a straight line. On the other hand, results show clearly that speech behavior allows no estimate of the degree of intoxication due to confounding factors such as individual tolerance, food intake or fatigue. If all of these are known, it would be conceivable to replicate the circumstances, though. While there may not be an easy solution to the basic question, the Exxon Valdez case “does represent the movement of basic science into a fairly well-known real-world situation” [12: 312].

7 The standardization of German pronunciation

7.1 Some historical facts

It is virtually impossible to understand the process of standardizing German pronunciation without looking at the political developments in that country. As opposed to countries such as England and France which had been political entities for centuries, the area that Germany now covers was split up into literally dozens of kingdoms, principalities etc. It was Bismarck who managed to unite Germany and form an empire following the Franco-Prussian war in 1871.

---

7 As is customary in adversarial systems like for instance in the US, there were opposing experts, in this case from the University of Florida.
This empire broke up as a consequence of WW I. After WW II, Germany was divided up among the allied powers, and in 1949 the groundwork was laid out for the – western – Federal Republic of Germany (FRG) and the – eastern – German Democratic Republic (GDR) as independent states. This historical rift ended with German reunification in 1989, when the GDR effectively ceased to exist and joined the FRG.

### 7.2 How Germany ended up with five pronouncing dictionaries

Establishing a pronunciation standard for German has proven to be a tedious and highly politicized endeavor. It is worth taking a moment to consider how other countries went about defining their standard variety. In centralist France, the Académie Française was quick to determine that the pronunciation by the educated speakers in the greater Paris area, was to be the standard [11]. The Comédie-Française served as the model for stage pronunciation. In England, the socially but not regionally marked pronunciation in the expensive private boarding schools and colleges was accepted as Received Pronunciation in the 19th century. Until fairly recently, it was totally inconceivable to hear anything but RP on, say, the BBC.⁸ In Germany, the situation was different, mainly because there had been no Germany until 1871. The newly united country needed to shape its identity, and one of the measures to that effect was to standardize orthography and pronunciation ([19]; [35: 8]). Those in charge went about it in a very “German” manner – i.e. a task force was formed that was to establish the “proper” pronunciation. That group consisted of six representatives from academia and theater, among them Theodor Siebs, who was also responsible for the publication of the dictionary. Most of them came from the north of Germany, and that group decided to use stage talk as a model. To cite an example, this meant that it was mandatory to pronounce the /r/ as an apical trill in all positions, i.e. also postvocically [48].⁹ There was also a strong bias toward “northern” characteristics, although no clearly defined region served as a model. A phonetician by the name of Wilhelm Viëtor opposed this and other prescriptive elements and published a competing pronouncing dictionary in 1890 based on a survey among educated speakers [54]. This publication never achieved a relevant degree of public attention, not least because the Siebs, as it had come to be known by that time, was recommended for use in schools. It appeared essentially unchanged in 18 editions, and its role as the “standard” throughout the first half of the 20th century was strengthened. Only the 19th edition [16], besides removing the term “stage pronunciation” (“Bühnenaussprache”) from the title, adopted a “moderated standard pronunciation” (“gemäßigte Hochlautung”) as an acceptable alternative to the stage norm.

In 1953, a meeting of delegations from both sides of the Iron Curtain took place in Frankfurt with the aim of developing a joint strategy for revising the outdated Siebs, but the delegations could not agree on a common set of guidelines to be followed. This is where politics comes in – the Western researchers insisted on a largely prescriptive process whereas the East German phoneticians from the University of Halle pursued a usage-based, descriptive approach which was more in line with the “democratic” self-perception of communist Germany. This split was now emphasized by the East German side because it regarded the emerging language gap between the two German states as a necessary consequence of the two ideologies on which they were based ([14], p. 120). This development increased the number of pronouncing dictionaries by two – one from Halle [22], and another authored by phoneticians dissenting from the firm

---

⁸ Various authors have pointed out that “RP” is not quite as uniform as the term suggests. Wells [56] describes no less than five sub-varieties of RP: Conservative RP or U-RP (as spoken by the “real” aristocrats and members of the Royal Family), mainstream RP (as then spoken in the BBC), advanced RP (as spoken by the younger London upper class), near-RP (as spoken by the ambitious middle class) and adoptive RP (as spoken by those aspiring to rise socially and trying a little bit too hard to conform to RP rules).

⁹ This single pronunciation feature may have contributed to Hitler’s speeches sounding so unpleasant irrespective of the content.
position held by the editors of the *Siebs*, Helmut de Boor and Paul Diels, in West Germany. The latter group was headed by Max Mangold from Saarland University, who – like the East German researchers – tried to introduce descriptive elements. The resulting dictionary was published as part of the DUDEN book series [18].

Following German reunification in 1989, the issue came up again. Finally, the West German delegates were ready for a usage-based approach. A monstrous pronouncing dictionary project was initiated which essentially was to follow the principles developed in Halle, but it was at the same time to include several stylistic levels and was to be accompanied by sound files representing different speaking situations of varying formality for each entry. Halle was in charge of the codification; the University of Cologne was responsible for producing the sound files. Briefly, the project proved to be much too big to be feasible. The publication that emerged was the *Deutsches Aussprachewörterbuch* [35], published by speech scientists from Halle. It turned out to be a much less comprehensive product than projected, which undoubtedly has its merits but has also met with criticism.\(^\text{10}\)

### 7.3 Conclusion

The codification of German pronunciation reflects some aspects of German history in a nutshell. This extends from the 19th century aspirations to unify language usage to the ramifications of the competing political systems in East and West Germany up to the developments that came with reunification. The different approaches to standardizing pronunciation in the two German states represent the respective political philosophy, and it is indicative of a rift within the West German society that there were two competing pronouncing dictionaries in the FRG, both publishing new editions in parallel.

### 8 Final remarks – phonetics and real life

Clearly, phonetics is not confined to the ivory tower. The main reason for this is the nature of the object of study, i.e., all aspects of spoken language including paralinguistic events. Since communicating is part of everyday life, the study of communication touches upon people’s everyday lives. In this contribution, it has been demonstrated how real-life events have fostered and sometimes even initiated phonetic research.

On the other hand, though, there are also examples of researchers turning to everyday events as objects of their study. One of the prime examples is Jürgen Trouvain. He has had the courage to address the phonetics of sports commentaries, laughter, and clicks [50], [51], [52]. Another example is the analysis of the Queen’s Christmas messages [23]. Research on these topics tends to be regarded as entertaining rather than scientific. Therefore, it has not always received the recognition it deserves from the scientific community because these topics may not be perceived as being at the center of interest to phoneticians.\(^\text{11}\) This contribution is intended to expressly encourage this kind of enterprise. It is this very ability to relate to everyday phenomena which will keep rare subjects such as phonetics alive and which will allow it to reach out to people beyond academia.

\(^{10}\) For instance, the most common pronunciation of the prevocalic and intervocalic /r/ is termed “velar” in the *Deutsches Aussprachewörterbuch* [35], which seems hardly tenable in view of the fact that the trill (and also the fricative) produced in the back of the oral cavity are clearly uvular.

\(^{11}\) Many years ago, this author studied grunting in tennis from a phonetics perspective [6], which has met with much more interest in the medical community than in the phonetics community.
Acknowledgements

A warm word of thanks goes to Katharina Zahner for valuable comments on a draft version of the paper, and to Anna-Maria Duplang for proofreading it and checking the English.

References


[61] https://www.fbi.gov/history/famous-cases/lindbergh-kidnapping