ACTA UNIVERSITATIS CAROLINAE
PHILOLOGICA 2/2019

ACTA UNIVERSITATIS CAROLINAE

PHILOLOGICA 2/2019

Editor
RADEK SKARNITZL

CHARLES UNIVERSITY
KAROLINUM PRESS
2019

Editor: doc. Mgr. Radek Skarnitzl, Ph.D.

http://www.karolinum.cz/journals/philologica

ISSN 0567-8269 (Print)
ISSN 2464-6830 (Online)

OBSAH

EDITORIAL

This *Phonetica Pragensia* issue of *Acta Universitatis Carolinae – Philologica* involves several landmarks for the Institute of Phonetics in Prague, as well as for the journal itself. It is the 15th issue of the journal which focuses on spoken communication and it is published at the occasion of the 100th anniversary of the Institute. It brings me great pleasure to be able to say that one hundred years after its foundation, our Institute is thriving, with a remarkable team of researchers active in a number of subdisciplines within the speech sciences.

The centenary of phonetics in Prague is commemorated with the opening paper of this issue. Written by Pavel Šturm, it examines the Institute's beginnings under its founder, Josef Chlumský, and his student and successor, Bohuslav Hála.

In the past, *Phonetica Pragensia* was regarded mostly as an in-house journal of our Institute. It is my hope that this issue marks a new beginning in this respect. We are honoured to present three papers by contributors from abroad. The first two are authored by top researchers in their fields, which enabled them to take a broader perspective at their respective topics. The paper by Jan Michalsky and Oliver Niebuhr considers the myths associated with charisma and charismatic speech and confronts them with current research. The paper by Volker Dellwo and his team examines indexical information and specifically how speakers are able to control it to be more recognizable.

The following papers by members of the Prague team address two large areas. It is not surprising that our team continues to examine the sound patterns of the Czech language. Nikola Paillereau and Kateřina Chládková study spectral and temporal characteristics of Czech vowels in spontaneous speech. Eliška Churaňová explores the relationship between the phonotactic structure of Czech stress groups and perceived rhythm. The second area concerns foreign-accentedness in Czech speakers of English, as well as Anglophone speakers of Czech. The study by Radek Skarnitzl and Jana Rumlová examines multiple segmental and prosodic features in the pronunciation of speakers with a strong Czech accent in their English. Pavel Šturm and Lea Tylečková look at one specific aspect, assimilation of voicing, in Czech speakers of Moravian origin. Finally, Jan Volín studies Anglophone speakers of Czech and compares their prosodic phrasing with native Czech and English speakers.

The issue is concluded by a paper by Maral Asiaee, Mandana Nourbakhsh and Saeed Rahandaz from Iran. Their study investigates linguo-palatal patterns of coronal stops in Persian.

https://doi.org/10.14712/24646830.2019.25

I sincerely hope that this issue of *Phonetica Pragensia* will be successful and will contribute to the rising profile of *Acta Universitatis Carolinae – Philologica*. Most importantly, I hope readers will find the results interesting and the discussions stimulating!

Radek Skarnitzl

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 9–26
PHILOLOGICA 2/ PHONETICA PRAGENSIA

THE BIRTH OF AN INSTITUTE: A CENTENNIAL JUBILEE OF PRAGUE'S INSTITUTE OF PHONETICS

PAVEL ŠTURM

ABSTRACT

The current issue of Phonetica Pragensia is published at the occasion of the Institute of Phonetics celebrating the 100th anniversary of its official foundation. The aim of this paper is to provide background to the contributions that follow, namely a historical perspective to the tradition of long-term experimental research set up in Prague by the early phonetic pioneers and continued until today. Drawing primarily on archival materials, the article brings a more detailed account of the constitutive years in comparison to the reviews published so far. It reveals the complexity that is involved in the process of establishing and sustaining a new (phonetics) institute, which might be informative to wider audiences as well.

Key words: history of phonetics, Czech phonetics, Institute of Phonetics, Josef Chlumský, Bohuslav Hála, phonetic laboratory

1. Introduction

It is unrealistic to expect that a journal article of standard length will provide an in-depth account of the history of an institute stretching 100 years to the past. Nevertheless, the scope allows for an in-depth look at a single period in such a narrative. Obviously, it will still be limited: an external history reconstructed from archival materials cannot be presented along with shifts in scientific thought that occurred in a selected timespan or with a detailed analysis of some representative corpus of phonetic research from the period. Given that another article exists – written in English and thus available to a wide audience – in which different approaches to studying speech sounds are discussed in the context of Prague researchers (Volín, 2014), a natural choice is to turn attention to the former, i.e., to the question of how the institute was established, in what background, who were the key players and how it developed during its initial phase. Hopefully, exploring the birth of an institute will be of interest to members of both local and international audiences. A more extensive, book-length treatment of this and other topics is currently in preparation.

When examining the constitutive years of the Prague phonetics institute, institutional and personal history is often difficult to distinguish. The foundation of the Institute of

https://doi.org/10.14712/24646830.2019.16

Phonetics (henceforth "IPh") is to a great degree bound up with the person of Josef Chlumský (1871–1939). It must be noted at the same time that Chlumský did not appear from thin air. He was influenced by various persons both before and after establishing a phonetic laboratory in Prague. With very few exceptions, scientific life is and must be a collaborative (ad)venture if it is to survive, adapt and develop. Therefore, Chlumský's mentors and close collaborators will be part of the narrative as well.

Most of the facts mentioned in the current article are based on primary sources and documents from the archives of Charles University and the Czech Academy of Sciences (referenced as appropriate). This approach is seen as preferable to compiling a text from secondary sources, which indeed do abound for this topic (but are also selective, differ in reliability and are written with varying intents). A substantial number of documents pertaining to Chlumský and the IPh were located in the archives and analysed, along with other internal materials procured from the IPh itself. The following is the picture that emerges out of reports, requests, personal letters, ministerial decrees – and well, secondary sources, too.

2. Josef Chlumský: Birth of a phonetician

Chlumský's phonetic education started in 1893, when he enrolled to study modern philology at the Faculty of Arts of the Czech university.[1] Although there was no phonetics study programme as such, it would be inaccurate to say that phonetics was not included in the philology section at all, as several lecture cycles were scattered in the curricula concerning the phonetics of individual languages. Chlumský explicitly mentions Zubatý's lectures on the Indo-European sound system as the inspiration which led him to dedicate himself to phonetics (Chlumský, 1928: 6). He might have chosen to attend phonetic lectures on Lithuanian and Latin, too. In later years, there were also lectures on the phonetics of French, German and various dead languages, and comparative and diachronic treatments of Slavic, Germanic or Romance languages (generally at least one such lecture cycle per semester). The first phonetic lecture on Czech appeared in 1904/05 ("The Speech Sound Structure of the Czech Language" by Emil Smetánka). Moreover, the German university offered regular phonetic lectures and practical phonetic exercises in French, German and English between 1889 and 1939.[2] Phonetics was thus by no means an exotic topic at the beginning of the twentieth century.

Chlumský nevertheless channelled his efforts into French philology and, with the intention of gaining more experience, studied abroad in 1895/96 and 1896/97, visiting several European universities in Berlin, Strasbourg and Paris. In the French capital,

[1] We must specify this, because since 1882 in fact two universities bore the emperor's name: the "German Charles-Ferdinand University" and the "Czech Charles-Ferdinand University" were two independent, parallel and presumably equal bodies until 1920, when the Czech university became the only rightful successor to the original medieval university (Štemberková, 2011).

[2] The most prominent teacher was Gustav Rolin (1863–1937), lecturing on phonetics for nearly 40 years (especially the phonetics of French, both lectures and seminars). After the first war, Percival Butler took care of English phonetic exercises, followed later by Cecil Wilkins. German phonetics appeared latest, starting with the Germanist Ernst Schwarz (1895–1983) in 1928. Interestingly, there were also several lecture cycles on public speaking and speech in the theatre.

Chlumský attended lectures by the Romance philologist Gaston Paris and the experimental phonetician Abbé Rousselot, among others.[3] Chlumský graduated in 1898[4] and assumed the career of a grammar-school teacher, a typical employment for Faculty of Arts graduates in the Austro-Hungarian Empire.[5] Secondary sources claim that Chlumský was interested in pursuing an academic career in French philology but that his efforts in this direction were to no avail because he soon noticed, with bitterness, that he was not taken into serious consideration (Miletić, 1930; Janko, 1931). I could not find any direct evidence supporting these claims, but several hints seem to be in tune with such an interpretation. Chlumský genuinely adored the French language and literature and was quite capable in this regard. But at around that time, another young Czech researcher, Maxmilián Křepinský (1875–1971), began lecturing at the philology department, and it might have been him who filled the vacant post.[6] As we shall see below, Chlumský left for France a year later.

Chlumský remembered his experience from Paris, and moreover, a brand-new, well-equipped phonetic laboratory had been launched in 1897 by Rousselot at Collège de France (Chlumský, 1920). So in 1910, after several years of full-time teaching, Chlumský commenced training with Rousselot once more. He probably did not expect that his second stay in Paris would turn out to be four years in duration.[7] Yet it certainly was the very place to be for an aspiring phonetician. Rousselot's laboratory did not have equals (except for the Hamburg one) and Chlumský could benefit greatly from the cooperation with Rousselot. Chlumský worked with instruments used for articulatory and acoustic measurements, learned the finesses of the trade and published articles in scientific journals mainly about the new and ever-changing experimental methods (e.g., Chlumský, 1911a, 1912, 1913, 1914). In October 1911 Chlumský writes from Paris to Prague, applying for a *venia docendi* in the field of experimental phonetics.[8] The application was accepted by the body of professors in January 1912. One of the reasons was that the latest philology programmes also included phonetics, and a new specialist in this field would be most

[3] Commission report from 13.1.1912 on Chlumský's habilitation application. ACU, the FA CU fonds, box 27, inv. n. 318.

[4] Chlumský's entry in the registry of doctors at the Czech university is available at https://is.cuni.cz/webapps/archiv/public/book/bo/1889173198808193/544. Chlumský's thesis dealt with the aesthetics of French and German verse.

[5] In the 19th century, the arts faculties were mainly schools for state officials, especially grammar-school teachers, who were allowed to teach after passing a state exam. Launching an academic career necessitated also the doctor's degree, which was conditioned – since the reform in 1872 – by further exams and most importantly by successfully defending a dissertation thesis. See Petráň (1997a: 155, 176) for details.

[6] Křepinský became a distinguished philologist, mostly interested in the diachrony of French (Ducháček, 1966). He graduated in 1902 in French and German philology, having spent the year 1898/99 at several French universities. Since 1908 he taught French philology at the faculty and was habilitated in 1909, becoming professor in 1919.

[7] Zubatý's report from 6.6.1911 on Chlumský's vacation leave application. ACU, the FA CU fonds, box 27, inv. n. 318. There was a substitute teacher for Chlumský at the grammar school, paid for by the state. Chlumský had to prolong the stay after each semester by presenting a new application. In the report Prof Zubatý praises Chlumský for his work up to then. He stresses that Chlumský needs to stay longer in Rousselot's laboratory so that he can use Rousselot's equipment, unavailable in Prague, for his research.

[8] Chlumský's habilitation application from 9.10.1911. ACU, the FA CU fonds, box 27, inv. n. 318. Chlumský submitted an experimental work, *An Attempt at Measuring Czech Speech Sounds and Syllables in Connected Speech* (Chlumský, 1911b), as his habilitation thesis.

welcome.[9] Chlumský travelled briefly to Prague in June for his colloquium in front of the professors, passed with success and returned back to Paris, where his career flourished. Rousselot's pupil at first, Chlumský had already become his assistant and collaborator, and was even appointed custodian of the laboratory.[10] Secondary sources suggest that their relationship was so close that, eventually, Chlumský was meant to be Rousselot's successor as director of the Collège de France department, had Rousselot moved to Sorbonne, which never happened (Hála, 1940; Ohnesorg, 1973).

In April 1914, the First International Congress of Experimental Phonetics was held in Hamburg (Mehnert, Pétursson & Hoffmann, 2016). It is not clear whether Chlumský participated in the congress. The professors of the Czech arts faculty received an invitation and recommended that a member should be present.[11] However, according to the list of abstracts (Panconcelli-Calzia, 1914), Chlumský did not have a presentation at the congress, nor did anybody else from the university. Nevertheless, judging from a photo taken at the congress, 200 people were present (Mehnert et al., 2006: 49), so it is still possible that the participation was passive. On the other hand, it might simply have been difficult to travel abroad with the impending Great War; in fact, the proceedings from the congress, prepared, never appeared in print due to the war (Neppert & Pétursson, 2006).

Chlumský returned to Prague in 1914 and commenced a new phase in his life. Requests for vacation leave from his grammar-school occupation were regularly sanctioned[12] so he could start giving lectures on phonetics at the Czech university. "Introduction to Phonetics" was attended by 18 students in the winter term and 13 in the summer term of 1914/15.[13] In the next year he also spoke about "Phonetic Methods"; in 1916/17 and 1918/19 he turned to the "Phonetics of the French Language". However, a major problem was that a phonetic laboratory was necessary if Chlumský wanted to conduct any research of the type he learned in France. He could work like Antonín Frinta (1884–1975), another phonetician at the university, who based all his phonetic research on direct auditory observation of speech.[14] But this is not what he preferred (Chlumský, 1920). Chlumský had already suggested in the habilitation application three years earlier that he "will establish a phonetic laboratory" and "supply the instruments necessary".[15] The task proved to be difficult. Already before the war, there was a shortage of rooms and buildings for the faculty departments, and this urgent need for adequate premises remained a long-term obstacle. Chlumský's lectures took place mainly in the Klementinum (the building also functioned as a military hospital during the war). Chlumský managed to assemble several instruments which he borrowed from other departments (e.g., a laryngoscope), and arranged with two professors at the Institute of Physics that he could

[9] Commission report from 13.1.1912 on Chlumský's habilitation application. ACU, the FA CU fonds, box 27, inv. n. 318.
[10] Commission report from 20.2.1916. ACU, the FA CU fonds, box 27, inv. n. 318.
[11] Minutes from the professors' meeting held on 19.2.1914. ACU, the FA CU fonds, box 2, inv. n. 32.
[12] Vicegerency decree from 24.7.1914. ACU, the FA CU fonds, box 183, inv. n. 1416.
[13] Commission report from 20.2.1916. ACU, the FA CU fonds, box 27, inv. n. 318.
[14] Frinta was mainly a Slavic philologist, especially later in his life. Nevertheless, he wrote two important phonetic books, *Modern Czech Pronunciation* (Frinta, 1909) and *The Phonetic Nature and Historical Development of the Consonant "v" in Slavonic* (Frinta, 1916), and as a member of the International Phonetic Association he advocated – to no avail – the use of the IPA alphabet in Czech linguistics. For more details, see e.g. Kurz (1959) and Ohnesorg (1959).
[15] Chlumský's habilitation application from 9.10.1911. ACU, the FA CU fonds, box 27, inv. n. 318.

use one of their rooms as a provisional laboratory.[16] There were six interested individuals who studied phonetic methods and participated in laboratory work in 1914/15, and five in the winter term of 1915/16.[17] Everything and everybody had to fit into the single room (Chlumský, 1920).

3. Laboratory of Experimental Phonetics and the Phonographic Archive

The route towards the establishment of a fully-fledged laboratory was not straightforward. The story that emerges from archival materials is as follows. In the early months of 1918, when Prague was still part of the Austro-Hungarian Empire, Chlumský sent a request to Vienna asking for subsidy to establish a phonetics institute at the arts faculty; it was left unanswered.[18] Fortune smiled on Chlumský when the new Czechoslovak regime was established. A proposal concerning Chlumský's professorship was put forward in October 1918, and the report of the responsible commission was endorsed in January 1919 by the body of professors.[19] On March 31, 1919, Chlumský presented another proposal for establishing a phonetics institute within the linguistic department of the arts faculty, consisting of a laboratory and a phonographic archive.[20] A commission was designated to process the proposal, but before a conclusion could be drawn in the May session, the Czechoslovak Ministry of Education and National Enlightenment had in the meantime dealt with the old request and complied, contributing 10.000 K for initial arrangements.[21] As a result, Chlumský asked for more money to finance a journey to Paris, where he would buy instruments and enter into agreements with French companies.[22] Simultaneously, in May 1919, Chlumský was appointed professor of phonetics, "with a special regard to experimental phonetics".[23] This also ended Chlumský's official duties at the grammar school: he was no longer a teacher (1898–1919) but a university

---

[16] There are many very interesting parallels between Chlumský and professors Čeněk Strouhal and Bohumil Kučera. They all studied at the Czech arts faculty, they all had substantial experience from abroad and they all founded a new institute out of scratch. The Institute of Physics was launched provisionally in 1883 in the Klementinum, functioning in fairly insufficient conditions, and moved as late as in 1908 to the new building in Karlov (Petráň, 1997b: 287–288). Moreover, Strouhal wrote a book on acoustics (Strouhal, 1902). When Chlumský approached the professors in 1914, it might have been this experience and sympathy to Chlumský's intentions – in addition to the fact that Strouhal was an acquiescence of professors Mareš and Král, important figures at the university and supporters of Chlumský – that contributed to the arrangement for the provisional phonetic laboratory.

[17] Commission report from 20.2.1916. ACU, the FA CU fonds, box 27, inv. n. 318.

[18] Commission report from 7.5.1919 on Chlumský's application for a journey to Paris. ACU, the FA CU fonds, box 27, inv. n. 318. See also Chlumský (1920).

[19] Minutes from the professors' meeting held on 24.10.1918 and 23.1.1919. ACU, the FA CU fonds, box 3, inv. n. 37.

[20] Commission report from 7.5.1919 on Chlumský's application for a journey to Paris. ACU, the FA CU fonds, box 27, inv. n. 318. See also minutes from the professors' meeting held on 3.4.1919, ACU, the FA CU fonds, box 3, inv. n. 37.

[21] Commission report from 7.5.1919 on Chlumský's application for a journey to Paris. ACU, the FA CU fonds, box 27, inv. n. 318.

[22] *Ibid.*

[23] Ministerial order from 7.7.1919 and the dean's office letter to the ministry of education from 18.6.1921. ACU, the FA CU fonds, box 27, inv. n. 318.

professor (1919–1939) with an appropriate salary. Professor Chlumský was then named director of the *Laboratory of Experimental Phonetics* and of the *Phonographic Archive* (hereafter Laboratory and Archive).

Both the Laboratory and the Archive were still situated in the single room belonging to the Institute of Physics. The new status resided rather in the official recognition of the Laboratory as a core part of the philology sciences of the faculty (the so-called "seminar"), listed as "Laboratory" in the curricula from the summer term of 1919/20 onwards. Moreover, an assistant was appointed to the Laboratory from January 1, 1920. Another important change was that Chlumský could apply for subsidies to procure the equipment he needed. A buying spree ensued over the following years during which Chlumský oscillated between Prague and Paris. Several professors at the faculty were enthusiastic about this "new institution", a "novelty that did not and does not exist at any of the former universities, and thus not even at the Vienna university".[24] Especially the Archive was seen as an expression of the patriotic spirit of the new republic, being envisioned as a saviour of the gradually vanishing dialects and thus of national importance.[25] Also, the practical use of phonetics was highlighted, for instance in language teaching or speech elocution.

The first new piece of equipment[26] was a kymograph, a machine used for recording speech (i.e., variations in sound pressure) graphically. It used up the entire ministerial subsidy mentioned above. A variety of tools was acquired for operating, maintaining and cleaning the machine. Another large subsidy was necessary for the Lioret machine, which could transcribe phonographic cylinders to kymographic curves, allowing these records to be analyzed visually as well. A microscope was shipped from Paris so that the tiny kymographic curves could be properly investigated (e.g. for measurements of F0). The most expensive purchase was a set of tuning forks,[27] which had to be imported – one by one, or several at once – over the years. Chlumský personally went to Paris in order to save some money as he helped with their construction (namely, with fine-tuning the pitch). Thirteen French tuning forks cost 49.000 K, equalling the total of special subsidies allocated for the preceding three years. Gramophonic records for the Archive were either bought or obtained as gifts. Moreover, phonetic journals and books were regularly ordered from abroad. In 1927, ten more tuning forks were bought. All in all, it took over ten years before one could finally say that the Laboratory was equipped properly.

This coincided with the year 1931, when the whole laboratory was moved to the recently constructed building of the arts faculty, something that had eagerly been expected for years. After some negotiations, the Laboratory was allocated five rooms on the ground floor.[28] There was a machinery room, a workroom, a microscopy room (assistant's room),

---

[24] Commission report from 7.5.1919 on Chlumský's application for a journey to Paris. ACU, the FA CU fonds, box 27, inv. n. 318, p. 1.

[25] *Ibid*.

[26] All items, down to the smallest pieces (like a pen, a knife, a bottle), were carefully logged in an accounting book along with their price. The book covers the years 1919–1950 (internal archive of IPh).

[27] The Prague collection of tuning forks was modelled on Rousselot's laboratory with Rudolph Koenig's *grand tonomètre universal* comprising more than a hundred tuning forks. The Czech tonometre consisted of a smaller number of forks, and thus a smaller range of frequencies. For a detailed discussion of tuning forks in phonetic research see Šturm (2015).

[28] Chlumský originally requested nine rooms (Hála's letter to the material commission from 9.1.1947; ACU, the FA CU fonds, box 22, inv. n. 256).

the director's room, and the Phonographic Archive. The location was ideal because of the substantial load due to heavy machinery: the kymograph, the Lioret machine and the phonograph totalled 350 kg, while other smaller pieces of equipment summed up to 360 kg, not counting the library and the Archive.[29] Not everything went well, however. For instance, in September 1932 Chlumský reported problems about the battery room, as he needed direct current power for the seminars; no repairs had been done as of March 1933, when he was urging the matter further.[30] Furthermore, several letters and phone calls were exchanged concerning the construction of window shades capable of complete room darkening.[31] This was necessary for examining and photographing sound waves using manometric flames, for experiments with tuning forks, and especially for capturing the vocal folds on film. Also, new equipment was being purchased from time to time, and the rooms were quite soon full. Last but not least, the body of professors had to debate over relatively unimportant issues, such as the change of door labels from "Phonetic seminar" to "Laboratory of Experimental Phonetics", which, according to Chlumský, more precisely reflected the type of work done at the institute.[32]

4. Gramophonic archives at the Academy

The idea of a national sound archive with recordings of dialects originated in the early part of the twentieth century, inspired by other such archives abroad. However, the budget of the Laboratory and the Archive was markedly insufficient for such a kind of venture, so in October 1928 a Phonographic Commission was established at the Czech Academy of Sciences and Arts, taking over the management and especially the recording of material (Chlumský, 1930). Unfathomably, the word "phonographic" appeared everywhere, from the Phonographic Commission and the Archive to the distribution materials (Kratochvíl, 2010: 19). However, phonographs had in fact been replaced by gramophones long before that, and the media were thus gramophonic records and not phonographic cylinders.

Without going into details (see Gössel, 2006: 113–120 and Kratochvíl, 2010), let's focus on the interconnection of the institution and the person. Chlumský chaired the commission at first, and was indeed the propelling force of the whole undertaking, investing much of his time in it. On November 16, the Commission accepted to enter into business negotiations with the French company Pathé after favourable recommendation by Chlumský following his past good experience with the company.[33] The Academy extended the scope of recording from dialects to records of poets and proficient public speak-

[29] Chlumský's report to the construction department from 15.12.1924. ACU, the FA CU fonds, box 116, inv. n. 1304.
[30] Chlumský's letter to the dean's office from 11.3.1933. ACU, the FA CU fonds, box 116, inv. n. 1304.
[31] A letter of the state construction administration to the dean's office from 6.3.1931. Chlumský's letter to the ministry of education from 7.3.1931. Chlumský's letter to the dean's office from 14.3.1931. ACU, the FA CU fonds, box 115, inv. n. 1298.
[32] Chlumský's letter to the body of professors from 11.3.1931. ACU, the FA CU fonds, box 115, inv. n. 1298.
[33] Report of the Phonographic Commission from 16.11.1928. MIA CAS, the CASA fonds, box 233, inv. n. 483.

ers to folk songs and other aspects of the vernacular culture. High financial demands were apparent to everyone from the very start, and it was anticipated that budgets would undoubtedly be exceeded. It was Chlumský's job to supervise the project and bring it to a successful conclusion.

It was incredibly difficult to record new material. The first session stretched over two months in 1929, under the supervision of the French phonetician Hubert Pernot. The wax discs were unreliable and much of the recorded material had to be thrown away (Gössel, 2006). A large number of people from all parts of the country were moved to Prague at great financial costs. The participants were chosen on the basis of fieldwork, favouring people with well-preserved dialect markers (see Suchý, 1934 for a description of a few participants). The person was seated in front of a microphone and was asked to speak on the prepared topic, usually several times in order to get a clean recording. The environment outside of the building was also controlled, with policemen patrolling, heavy horse carts forbidden access (Kratochvíl, 2010: 26–27).

The following development turned into a nightmare for Chlumský. First, the financial situation was hopeless. Aid was sought from all sides, including the ministry, banks, and various affluent individuals (Kratochvíl, 2010: 32–36). Chlumský was both the financier, begging for money, and the salesman, offering records to schools and public institutions. His advantage was that he had a wide net of contacts. Second, the French company turned out to be unresponsive and unreliable. Chlumský had to solve defective and delayed deliveries, as well as numerous problems with the distribution, facing obstacles from the side of Czech companies as well (consult Gössel, 2006 for details). The popularity likewise did not meet expectations, and the records did not sell well, despite an initial wave of sales at secondary schools; in 1937, the Commission reported a total of 525 sold records, a ridiculous accomplishment.[34] Finally, Chlumský was the chairman of the commission until 1932, when he resigned in protest to unconfirmed (but not recanted) allegations that the quality of the records was inadequate.[35] This dispute had occupied the Academy for several months.[36] Nevertheless, Chlumský remained an active member of the commission and participated in its running under the professors Josef Janko and Emil Smetánka, who in turn became the next chairmen.

It would be unfair to reproach the commission for not fulfilling the initial plans, which were simply too ambitious. However, a problem was the debatable usefulness of the recorded material. Several thousand records were available by the 1940s. Yet although a number of persons were cataloguing and transcribing some of the records, the archive was never put to serious scientific use (with the exception of Mazlová, 1942). Only recently was it analyzed by ethnologists focusing on the musical part (Kratochvíl, 2009, 2010).

[34] Report of the Phonographic Commission from 28.4.1937. MIA CAS, the CASA fonds, box 233, inv. n. 486.

[35] Report of the Phonographic Commission from 27.1.1932. MIA CAS, the CASA fonds, box 233, inv. n. 486.

[36] Straka's letter from 20.3.1931. Chlumský's complaint from 29.4.1931. Report of the Phonographic Commission from 10.6.1931 and its annexe. MIA CAS, the CASA fonds, box 233, inv. n. 486.

5. Chlumský's academic and scientific influence

Chlumský was not only a good organizer and coordinator. He was endowed with other qualities academics need: pedagogical and scientific work. The teaching was all the more difficult because Chlumský did not have many predecessors to draw on, so he was forced to develop the lectures himself. The topics of his *lectures* from 1919 to 1939 are summarized in Table 1. Note especially that there are three types of topics: general phonetic lectures, linguistically oriented lectures about the pronunciation of languages, and methodologically oriented lectures about various scientific procedures and phonetic instruments. Unfortunately, there seems to be no extant written record of the specific contents of the lectures. Ohnesorg (1973) documents that Chlumský was very particular about the lectures and prepared them meticulously; in his (lost) diaries, Chlumský even noted the students' response to individual lectures. Another student, Miletić (1930), also stressed the clarity of Chlumský's presentation and his rhetorical talent. The accompanying *laboratory work* was practical and "experimental", thus often intriguing to new students, standing out from other philology courses. The participants learned about and practised various methods, conducted painstaking measurements and were encouraged by Chlumský to carry out independent research. They also received auditory and transcription training. Colleagues from the university who knew Chlumský well frequently mentioned his genuine, almost obsessive interest in the academic work and an impeccable character, modest, strict yet gracious (Miletić, 1930; Janko, 1931; Hála, 1940; Ohnesorg, 1973).

Table 1: A summary of Chlumský's lectures since 1919. The "occasional" lecture cycles were given less than five times during that period. The titles have been simplified and unified.

Regular lecture cycles	Occasional lecture cycles
Introduction to phonetics	Melody of the French language
Physiology of speech	On French stress/accent
Acoustics of speech	On French 'e muet'
Comparative phonetics of Czech, French, English and German	On French nasals
French phonetics (practical)	On French liaison
Use of phonetics in teaching French	Diachronic sound change
Czech quantity (based on measurements)	How to read the curves of speech
Czech stress (based on measurements)	Experimental and auditory phonetics
Quantity, melody and stress in European languages	Discussion of Grammont's book
History of (Czech) phonetics	On the methods in phonetics

The early part of the twentieth century was marked by debates between advocates of the auditory approach to phonetics, which relies on the skilled phonetician's capacity to differentiate speech sounds by ear, and the instrumental approach, which places more weight on the measurements of speech events registered for instance by the kymograph or the artificial palate (Mehnert et al., 2016). Chlumský, a disciple of Rousselot, took an

active part in the debate, as evidenced by the published polemic contributions in various journals (*Listy filologické*, *Naše věda*). He fervently defended the experimental method, arguing that whenever subjective evaluations are not clear-cut, as is the case with stress or quantity, some "objective tools" must be employed, and even minute, "microscopic" data might be perceptually relevant (Chlumský, 1926, 1927). He did not wish to simply substitute listening with machines, though; the latter was a useful and sometimes necessary extension and verification of the former. Note that he taught both about auditory analysis and instrumental work. Moreover, Chlumský was quite aware of the limitations of the devices, and many of his works were methodological in nature (a series of articles in *Revue de phonétique* or his 1911 dissertation thesis mentioned earlier). He also repeatedly stressed that linguistics is important for the phonetic endeavour, and was concerned with communicative meaning.[37]

It is easy to see that Chlumský did not have much spare time. He effectively divided his activities in the early 1920s between furnishing the laboratory and teaching, and in the 1930s between work for the Academy and the Laboratory. Chlumský did not publish anything substantial until 1924, when a series of experimental works appeared (on English and French consonants, and several articles on Czech prosody). In 1928, Chlumský's seminal work *Czech Quantity, Melody and Accent* followed (Chlumský, 1928), for which he assembled an unprecedented amount of material. It is no wonder that he frequently applied for exemption from lecturing so that he could finish laboratory work.[38] The use of instruments allowed him to note important acoustic details, but an inseparable part of the book concerns the innumerable examples Chlumský gathered by observing casual speech around him, specifying who said what, how, where and in what circumstances. Chlumský provides several important findings, either experimental verifications of previous auditory impressions of Czech or entirely original in the Czech context (such as the temporal compression of consonants in complex clusters or vowels in closed syllables, or the effect of phrase final lengthening).

Chlumský's most renowned publication was *Radiography of French Vowels and Semi-Vowels* (Chlumský, Pauphilet & Polland, 1938). The book was commissioned from abroad[39] and includes an extended French résumé. The quality of the X-ray images was indeed outstanding, and it was the first such description of the French language. It cannot be stressed how demanding the job was. For illustration, the 145 X-ray images were procured from a French speaker during the eight years preceding the publication. The radiography took place in one of the university hospitals under the supervision of Chlumský and the technical supervision of the radiologist Bohumír Polland. The results had to be checked multiple times in order to (1) detect errors in the choice of articulatory phase (to be selected from whole isolated words) and (2) ensure no motion blur due to the subject's

[37] Chlumský's approach is not that far from structuralism. For instance, despite tracing the durations of vowels, he stresses the distinctiveness of length in the system and the importance of relative rather than absolute values (Chlumský, 1928: 26, 101, 108). He also investigates shifts in meanings associated with different forms used in comparable environments (Chlumský, 1928: 218).

[38] The ministry of education approvals from 12.6.1924 and 6.2.1925; from 3.2.1932, 13.12.1936 and 16.9.1937 for later works. ACU, the FA CU fonds, box 27, inv. n. 318.

[39] Chlumský's request for exemption from lecturing from 25.11.1931. ACU, the FA CU fonds, box 27, inv. n. 318. Compare also Hála (1939: 252).

movement during exposition. Chlumský's advantage was that, as we shall see below, he built on the previous experience of his assistant in this field.

5.1. Bohuslav Hála

Given the laboriousness and enormous time demands of laboratory work in those days, Chlumský naturally did not work alone. As mentioned above, an assistant was allocated to him right from the establishment of the Laboratory. Bohuslav Hála (1894–1970) knew Chlumský early on because Chlumský, along with other faculty members, taught at the grammar school which Hála attended. Unfortunately, his education and early life were severely affected by the Great War. In the winter term of 1913/14 Hála enrolled to study classical philology, but soon revised his course to Czech and French philology (Ohnesorg, 1954). He signed for Chlumský's classes, and Hála's name stands out on top of the list of phonetics students. He completed two terms before he joined the Austro-Hungarian army on the Eastern front and then on the South Western Front, spending three years in the fights, becoming an officer (Lieutenant) and receiving a Golden Medal for Bravery in 1918. After a severe injury, he served for another few months in the rear before the end of the war. [40] Afterwards, Hála resumed his studies. He spent the summer term of 1920 in France, studying at Strasbourg University to "improve his qualification".[41] He was already Chlumský's assistant, and it was viewed as part of his phonetic growth. Hála attended seven phonetic and linguistic lectures, and five more lectures which he used as preparation for the upcoming state exams.[42] He passed the exams in 1921, and received the doctor's degree in 1927.[43] Hála's first publication, a book concerning articulatory description of Czech sounds, was also delayed by the war and appeared in 1923, although he started the research in 1914 and continued during his military leave in 1918 (Hála, 1923: 3). Later, Hála argued in an application related to employment perquisites that, had it not been for the war, he would have become an assistant already in 1917 or 1918, i.e., three years earlier.[44]

The job of an assistant was determined by university regulations, but it was department specific, too. Laboratory work was too strenuous for one person, so Hála was often assigned the task of measuring and drawing diagrams for Chlumský's publications. Between 1925 and 1928 Hála prepared a total of 191 diagrams.[45] Hála often aided Chlumský in the phonetic seminars, and he also showed phonetic instruments to visiting students and guests during tours of the Laboratory.[46] Similarly, Chlumský conferred some teaching duties on Hála in light of his deteriorating health condition and preoccupation

[40] Hála wrote outstanding memoirs depicting his war experiences, discovered and published posthumously (Hála, 2018). Hála furthermore spent three more summers (1919, 1922 and 1926) in military service exercises (personnel sheet, ACU, the FA CU fonds, box 22, inv. n. 256).

[41] Annexe to the recommendation for conferring the degree of Doctor of Science on Hála from 23.6.1955 (Hála's CV). ACU, the FA CU fonds, box 22, inv. n. 256.

[42] Report on the progress of Hála's studies in Strasbourg from 1.7.1920. ACU, the FA CU fonds, box 115, inv. n. 1298.

[43] https://is.cuni.cz/webapps/archiv/public/book/bo/1391711927350373/452.

[44] Hála's letter to the ministry of education from 24.4.1946. ACU, the FA CU fonds, box 22, inv. n. 256.

[45] Annexe to the recommendation for conferring the degree of Doctor of Science on Hála from 23.6.1955 (list of Hála's publications). ACU, the FA CU fonds, box 22, inv. n. 256.

[46] *Ibid*.

with research; Hála filled in for Chlumský for several semesters.[47] After the establishment of the Phonographic Commission, Chlumský collaborated with Hála on the work for the Academy as well. Hála was present during the first large recording session in 1929, and in the 1930s he for instance travelled around Moravia in order to acquire speakers from that region or was assisting during later recordings.[48]

However, in addition to assisting Chlumský, Hála conducted research of his own, heartily encouraged and supported by his former teacher. Hála's book on Czech articulation mentioned above (Hála, 1923) was in fact the very first fruit of the Prague laboratory, as Chlumský's work up to then was based on measurements acquired in Paris. Hála continued in articulatory research, following with a book presenting X-ray drawings of the tongue and other organs during the articulation of Czech sounds (Polland & Hála, 1926). This work was especially important because the data were the foundation on which many of Hála's later popularizing and teaching publications are based (Hála, 1941, 1942, 1960), as well as other publications on Czech phonetics (Romportl, 1985; Palková, 1994). Interestingly, the research took place at the Faculty of Arts, since the X-ray machine was owned privately by Hála's collaborator, Dr Polland, who operated the machine. Hála – himself the only subjec – was quite aware of the harmful effects of X-rays to the irradiated skin, which is why they needed to limit the amount of time operating and the total number of expositions. As a skilled phonetician, Hála was able to lock the articulators in the target position for the three to four seconds necessary for proper exposition.[49] The resulting images speak for themselves.

Hála's next important project was the examination of the vocal folds during phonation. It turned out to be a unique accomplishment also in terms of international impact. Several attempts had been done at filming the vocal folds, but Hála offered a combination of high-speed cinematography and stroboscopy (see Šturm, 2019 for details). Both methods allowed him to directly observe phonatory cycles. Hála collaborated this time with Dr Honty, a specialist in scientific cinematography. They filmed Hála's vocal folds in 1928 and 1929, presenting the film to an international audience in 1930, when Prague hosted the Fourth International Congress of Logopedics and Phoniatrics (Hála, 1942: 14). The film was a huge success, and several copies were requested from England, France, Belgium, Germany and the USA.[50] The copy preserved at the IPh was digitized by the National Film Archive and is available online.[51] A report written in French (Hála & Honty, 1931) is still often cited in the literature on voice and phonation. The German phonetician (of Italian origin) Panconcelli-Calzia considers it the first instance of capturing the vocal folds with high-speed cinematography.[52] As a result, Hála was admitted in

[47] Hála's letter to the ministry of education from 24.4.1946. ACU, the FA CU fonds, box 22, inv. n. 256. Compare also note 39 concerning Chlumský's exemptions from lecturing.

[48] Annexe to the recommendation for conferring the degree of Doctor of Science on Hála from 23.6.1955 (list of Hála's publications). ACU, the FA CU fonds, box 22, inv. n. 256.

[49] These "long-window" images were identical to a few "momentary" control images developed with a short exposition time (only 50–100 ms). The authors concluded that Hála's sustained articulation was similar to his normal articulation, and could thus be used as the data.

[50] Hála's letter to the ministry of education from 24.4.1946. ACU, the FA CU fonds, box 22, inv. n. 256.

[51] https://fonetika.ff.cuni.cz/en/research/from-our-research/history/

[52] *VOX*, 1931, vol. 17(1), 77–78.

1932 to the prestigious Société francaise de phoniatrie, and regularly received specialized literature in return for a membership fee.[53]

This was not the only international activity of Hála. Although he did not attend the first two International Congresses of Phonetic Sciences (ICPhS), which were held in 1932 in Amsterdam and in 1935 in London, he participated at the third ICPhS held in 1938 in Ghent. He spoke on the acoustics of vowels, which was a topic that Chlumský assigned to him in 1929. The main product of this extended – ten-year! – research was *The Acoustic Nature of Vowels* (Hála, 1941). Hála used a variety of methods, from auditory analysis to experiments with resonators, tuning forks, oscillators and also direct mathematical computation of spectra from the waveform (see also Šturm, 2015). The primary objective was to derive formant values for the Czech vocalic system, which he described, with limited technical possibilities, quite accurately (compare Skarnitzl & Volín, 2012). Chlumský thus had a very talented and diligent person at hand at the institute.

5.2. Unpaid assistants

Hála continued to be Chlumský's assistant even after receiving *venia docendi* in experimental phonetics in 1930. However, three more assistants were successively associated with the institute. Chlumský's circle was first expanded in 1933 by Jiří Straka (1910–1993, later known as Georges Straka).[54] The son of a well-known philologist developed a passion for French at an early age. He attended Charles University between 1928 and 1934, graduating in Romance philology and phonetics.[55] Unfortunately, Straka's connection to the IPh became loose when he moved to Paris on a stipend by the French government in order to deepen his education. He visited several French universities between 1934 and 1937, including the Sorbonne and Collège de France, where he focused on French philology (under Meillet, Roques, Vendryes, Benveniste, among others). In 1936, Chlumský decided not to prolong Straka's contract.[56] Straka eventually returned to Prague, but he spent the pre-war years teaching at a grammar school before exiling to France in 1939.

The position of a second assistant was thus transferred to Karel Ohnesorg (1906–1976), a keen student who participated in the phonetic seminar during and even *after* his studies at the university between 1924 and 1928.[57] Ohnesorg taught at a grammar school until 1945, and his position at the university, since 1936, was unpaid until the war.[58] Ohnesorg's field of specialization was once again the French language (compare Chlumský, Hála, Straka), but he was also exceptionally interested in pedagogy and teaching methodology (Bartoš, 1966). Ohnesorg helped Chlumský and Hála with the seminars; the latter praised him for being very dutiful and capable of working inde-

[53] The dean's letter to the bank from 6.9.1951. ACU, the FA CU fonds, box 22, inv. n. 256.
[54] For more details on Straka's life and work, see Swiggers (1993, 1994) and Roques (1994).
[55] https://is.cuni.cz/webapps/archiv/public/book/bo/1513001005202135/492.
[56] Chlumský's proposal from 9.5.1936 to appoint Ohnesorg an assistant. Internal archive of the IPh.
[57] Hála's expert opinion on Ohnesorg from 13.11.1950. Internal archive of the IPh. As regards his education, Ohnesorg studied Latin and French, but his doctor's exams were in pedagogy, philosophy and aesthetics (https://is.cuni.cz/webapps/archiv/public/book/bo/1836656452491438/76). Ohnesorg's early publications include especially grammarbooks of Latin and French.
[58] Notice of Ohnesorg's appointment from 8.9.1936. ACU, the FA CU fonds, box 115, inv. n. 1298.

pendently.[59] His other duties included for instance neat drawing of articulatory sketches for Chlumský's X-ray images (Chlumský et al., 1938: 11) or conducting several smaller but not trivial experiments over the years on vowel acoustics for Hála (Hála, 1941: 43, 51–55, 60, 133, 137). Later, at Hála's direct instigation, Ohnesorg turned to investigating language acquisition (Ohnesorg, 1947, 1948a, 1948b), which became the topic of his life. The latter two works are unique in that Ohnesorg captured the longitudinal development of his own child's speech. In the tradition of Chlumský and Hála, Ohnesorg was also very prolific in writing detailed reviews of important phonetic works.

Another unpaid assistant was Věra Mazlová (1913–1950), who pursued phonetics since her studies between 1932 and 1938[60] and was appointed assistant shortly before the war in 1937.[61] Mazlová visited the Laboratory and the seminars not just for a single class, but during the whole period of her studies (namely, for six semesters). Moreover, Mazlová attended quite a few phonetic lectures:[62]

- Chlumský's "French Phonetics for Beginners" (two semesters), "Use of Phonetics in Teaching French" and "Melody of the French Language" (compare Table 1 above);
- Hála's "Comparative Phonetics of the Slavic Languages" (two semesters) and "Introduction to Czech Phonetics" (two semesters);
- Weingart's "Slavic Sound System" (two semesters), "Proto-Slavonic Consonantism", "Problems of Phonology and Sound Systems", "Issues in Czech Rhythmics and Metrics";
- Smetánka's "Sound System of the Czechoslovak Language" (two semesters).

Although Mazlová's study programme was the Czech and French languages, we can see that phonetics featured prominently in her curriculum: 21 classes in total! In her scientific work, Mazlová inclined especially towards dialectology (Mazlová, 1942, 1949). Unfortunately, she was forced to teach full time at various grammar schools until 1945, which seriously limited the time she could dedicate to her phonetic interests.[63]

6. Conclusion

The preceding sections presented a picture of how an institute can be established and brought to life, to the point when it becomes "a training centre for Central Europe" (Palková, 2000: 47), with numerous incoming international students. It was by no means an easy accomplishment, and support was necessary along the entire way. First and foremost, it was imperative to convince others of the usefulness of the new institution. In this respect, the support of other people, especially in the body of professors, was crucial for Chlumský. We should mention professors Josef Zubatý (1855–1931) and Josef Janko (1869–1947). These important figures were not phoneticians themselves, but they had a wide range of knowledge and could appraise the potential contribution of phonetics not

[59] Hála's expert opinion on Ohnesorg from 20.5.1951. Internal archive of the IPh.
[60] https://is.cuni.cz/webapps/archiv/public/book/bo/1924165860347712/327.
[61] Hála's request for Mazlová's leave from school from 15.5.1949. ACU, the FA CU fonds, box 41, inv. n. 480.
[62] Mazlová's study index. ACU, the FA CU fonds, box 41, inv. n. 480.
[63] Mazlová's CV. ACU, the FA CU fonds, box 41, inv. n. 480.

only to the field itself, but to linguistics in general. Although the ministry of education usually accepted proposals put forward by the body of professors, the decision was ultimately theirs. Therefore, it should not surprise us that Chlumský (1928: 4) gives thanks as well to a ministerial official who showed appreciation for his institute at its formation. Furthermore, the successful establishment of an institute also requires – besides great competence in the field – a certain kind of personality: persuasive in communication, persevering, with good organizational skills, and above all with unlimited enthusiasm and deep conviction. Chlumský definitely met these criteria.

Chlumský's assistants became important players in the phonetic world. While Chlumský was fundamentally connected to the Laboratory he had established and provided with equipment, Hála was similarly closely associated with the Institute of Phonetics that arose from the Laboratory after the Second World War. The subsequent development under Hála cannot be discussed here, as it belongs to a different chapter. Suffice it to say that Hála became a long-term director of the IPh with many scientific successes but also many disciples who further increased his influence. Ohnesorg was a member of the Prague team until 1955, when he was (willingly) transferred to Brno to take care of the local phonetic section at Masaryk University. He became professor in 1957. Straka decided to move even farther, settling down in France. He formed a Phonetic Institute in Strasbourg and was its director between 1945 and 1960 when he retired, a distinguished linguist. Mazlová was part of Hála's circle until her premature death in 1950.

Chlumský died on March 12, 1939, a few days before Czechoslovakia ceased to exist. The burial speech was delivered by Hála and the ceremony was held on the day Hitler's hordes marched through Prague. Shortly after, all Czech universities were closed down for the period of six years. Hála, Ohnesorg, Mazlová and many others were forced to pursue other occupations (for instance, working at the Academy).[64] In an attempt to sound less German, Ohnesorg began to use the name "Karel Orlík" when signing his publications. The exiled Straka was incarcerated in a concentration camp.

Yet, the story does not stop here. Ohnesorg remained Ohnesorg, and Straka returned home. The phonetic endeavour in Prague was getting stronger with good prospects for the future. We can say that the issue of Chlumský's succession was resolved successfully.

ACKNOWLEDGEMENTS

This research was supported by the Charles University project Progres Q10, *Language in the shiftings of time, space, and culture*. I would like to thank Alena Homolková from the Archive of Charles University for her great willingness to prepare all the materials I asked for. A thank-you note should also go to the students of the History of Phonetics course who helped me with the processing of archival documents over the past four years.

[64] Mazlová also taught phonetics at the Prague Conservatoire since 1942.

REFERENCES

Bartoš, L. (1966). K šedesátinám profesora K. Ohnesorga. In: *Sborník prací Filozofické fakulty brněnské univerzity – řada jazykovědná*, 15(A14), 7–13. Brno: Masarykova univerzita.

Ducháček, O. (1966). L'œuvre de Maxmilian Křepinský. *Études romanes de Brno*, 2(1), 9–21.

Frinta, A. (1909). *Novočeská výslovnost*. Praha: Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.

Frinta, A. (1916). *Fonetická povaha a historický vývoj souhlásky „v" ve slovanštině*. Praha: Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.

Gössel, G. (2006). *Fonogram 2. Výlety k počátkům historie záznamu zvuku.* Praha: Radioservis.

Hála, B. & Honty, L. (1931). La cinématographie des cordes vocales à l'aide du stroboscope et de la grande vitesse. *Otolaryngologia Slavica*, 3, 1–12.

Hála, B. & Sovák, M. (1941). *Hlas, řeč, sluch*. Praha: Česká grafická Unie.

Hála, B. (1923). *K popisu pražské výslovnosti*. Praha: Česká akademie věd a umění.

Hála, B. (1939). Za profesorem J. Chlumským. *Časopis pro moderní filologii*, 25, 249–255.

Hála, B. (1940). *Josef Chlumský*. Praha: Česká akademie věd a umění.

Hála, B. (1941). *Akustická podstata samohlásek*. Praha: Česká akademie věd a umění.

Hála, B. (1942). *Řeč v obrazech*. Praha: Václav Petr.

Hála, B. (1960). *Fonetické obrazy hlásek*. Praha: SPN.

Hála, B. (2018). *Dvě ofenzivy: Paměti bojů v jižním Tyrolsku a na Soči (1916–1917)*, edited by B. Nováková & P. Heřmánek. Praha: Epocha.

Chlumský, J. (1911a). Appareils nouveaux. *Revue de Phonétique*, 1, 68–78.

Chlumský, J. (1911b): *Pokus o měření českých zvuků a slabik v řeči souvislé.* Praha: Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.

Chlumský, J. (1912). Comparaison des tracés du phonographe et du petit tambour. *Revue de Phonétique*, 2, 213–250.

Chlumský, J. (1913). Méthodes pour obtenir le profil de la langue pendant l'articulation. *Revue de Phonétique*, 3, 167–173.

Chlumský, J. (1914). La photographie des articulations dessinées au palais artificiel. *Revue de Phonétique*, 4, 46–58.

Chlumský, J. (1920). Fonetické laboratoře v cizině a nově založená laboratoř pro experimentální fonetiku na české universitě v Praze. *Živé slovo*, 1, 19–22.

Chlumský, J. (1926). Ke sporu o českou kvantitu a přízvuk. *Listy filologické*, 53, 304–308.

Chlumský, J. (1927). Ke sporu o českou kvantitu a přízvuk II. *Listy filologické*, 54, 22–28.

Chlumský, J. (1928). *Česká kvantita, melodie a přízvuk*. Praha: Česká akademie věd a umění.

Chlumský, J. (1930). Fonografický a gramofonický archiv České akademie věd a umění v Praze. *Časopis pro moderní filologii*, 16, 189–192.

Chlumský, J., Pauphilet, A. & Polland, B. (1938). *Radiografie francouzských samohlásek a polosamohlásek.* Praha: Česká akademie věd a umění.

Janko, J. (1931). Několik slov o životě a působení Josefa Chlumského. *Časopis pro moderní filologii*, 17, 1–5.

Kratochvíl, M. (2009). *Lidová hudba v Československu 1929–1937.* Praha: Etnologický ústav AV ČR.

Kratochvíl, M. (2010). *Lidová hudba v nahrávkách Fonografické komise České akademie věd a umění* [PhD thesis]. Praha: FF UK.

Kurz, J. (1959). Slavistické dílo profesora Dr Antonína Frinty. In: K. Horálek, J. Kurz & M. Romportl (Eds.), *Acta Universitatis Carolinae – Philologica Supplementum, Slavica Pragensia I*, 3–13. Praha: Univerzita Karlova.

Mazlová, V. (1942). Systém hanáckých samohlásek. *Časopis pro moderní filologii*, 28, 137–144, 270–276.

Mazlová, V. (1949). *Výslovnost na Zábřežsku.* Praha: FF UK.

Mehnert, D., Pétursson, M. & Hoffmann, R. (2016). *Experimentalphonetik in Europa*. Dresden: TUDpress.

Miletić, B. (1930). Jozef Chlumský. *Južnoslovenski filolog*, 9, 319–327.

Neppert, J. & Pétursson, M. (2006). Death of a phonetics institute. The phonetics institute of the University of Hamburg. *The Phonetician*, 93/94, 43–46.
Ohnesorg, K. (1947). O vývoji dětské řeči a její fonetice. *Pedologické rozhledy*, 3, 65–91.
Ohnesorg, K. (1948a). *O mluvním vývoji dítěte*. Praha: Jednota českých filologů.
Ohnesorg, K. (1948b). *Fonetická studie o dětské řeči*. Praha: FF UK.
Ohnesorg, K. (1954). Profesor Hála šedesátníkem. *Časopis pro moderní filologii*, 36, 46–48.
Ohnesorg, K. (1959). Fonetika v díle Antonína Frinty. In: K. Horálek, J. Kurz & M. Romportl (Eds.), *Acta Universitatis Carolinae – Philologica Supplementum, Slavica Pragensia I*, 15–20. Praha: Univerzita Karlova.
Ohnesorg, K. (1973). Český fonetik Josef Chlumský. In: *Zprávy Kruhu přátel českého jazyka*, 1–5. Praha: Kruh přátel českého jazyka.
Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.
Palková, Z. (2000). 80 years of phonetics at Charles University Prague. *The Phonetician*, 81, 47–50.
Panconcelli-Calzia, G. (1914). Annotationes phoneticae. *Vox*, 24(3), 147–168.
Petráň, J. (1997a). Filozofická fakulta 1848–1882. In: J. Havránek (Ed.), *Dějiny Univerzity Karlovy III (1802–1918)*, 155–180. Praha: Karolinum.
Petráň, J. (1997b). Filozofická fakulta 1882–1918. In: J. Havránek (Ed.), *Dějiny Univerzity Karlovy III (1802–1918)*, 257–304. Praha: Karolinum.
Polland, B. & Hála, B. (1926). *Artikulace českých zvuků v rentgenových obrazech (skiagramech)*. Praha: Česká akademie věd a umění.
Romportl, M. (1985). *Základy fonetiky*. Praha: SPN.
Roques, G. (1994). Georges Straka (1910–1993). *Revue de linguistique romane*, 58, 281–288.
Skarnitzl, R. & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18, 7–11.
Strouhal, Č. (1902). *Akustika*. Praha: Jednota českých matematiků.
Suchý, K. (1934). Jak vzniká náš národní fonetický archiv. *Světozor*, 34(27), 6–7.
Swiggers, P. (1993, Ed.). *Georges Straka, Notice biographique et bibliographique*. Louvain: Centre international de dialectologie générale.
Swiggers, P. (1994). Georges Straka (1910–1993). *Orbis*, 37, 593–602.
Štemberková, M. (2011). *Univerzita Karlova*. Praha: Karolinum.
Šturm, P. (2015). The Prague historical collection of tuning forks: A surviving replica of the Koenig tonometre. In: R. Hoffmann & J. Trouvain (Eds.), *Proceedings of the First International Workshop on the History of Speech Communication Research*, 95–105. Dresden: TUDpress.
Šturm, P. (2019). The contribution of Czech phonetics to laryngeal investigation. In: *Proceedings of the 19th International Congress of Phonetic Sciences*. 1903–1907. Canberra: ASSTA.
Volín, J. (2014). Speech sound structure studies in Prague: Differences in approaches and conflicts between methods. *La Linguistique*, 2014/2, 83–100.

List of abbreviations

ACU = Archive of the Charles University
FA CU = Faculty of Arts of Charles University
IPh = Institute of Phonetics
MIA CAS = the Masaryk Institute and Archive of the Academy of Sciences of the Czech Republic
CASA = the Czech Academy of Sciences and Arts

RESUMÉ

Aktuální číslo časopisu Phonetica Pragensia vychází při příležitosti stoletého výročí Fonetického ústavu FF UK. Cílem článku je podat historický kontext k příspěvkům, které následují, a představit tradici dlouhodobě pěstovaného fonetického experimentálního výzkumu, jemuž v Praze razili cestu právě zakladatelé ústavu. Díky důrazu na archivní zdroje článek přináší v porovnání s doposud publikovanými přehledy důkladnější osvětlení ustavujících let pražského fonetického pracoviště. Ukazuje mimo jiné spletitost procesu, jakým je zakládání a následné budování a upevňování nového ústavu, což může být přínosné i pro širší publikum.

Pavel Šturm
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: pavel.sturm@ff.cuni.cz

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 27–56
PHILOLOGICA 2/ PHONETICA PRAGENSIA

MYTH BUSTED? CHALLENGING WHAT WE THINK WE KNOW ABOUT CHARISMATIC SPEECH

JAN MICHALSKY and OLIVER NIEBUHR

ABSTRACT

Charisma is a complex phenomenon. This fact manifests itself not least in an abundance of myths, half-truths, and unanswered research questions. Most charisma myths have not been uncontroversial, and since empirical investigations have advanced quickly over the past years, we take the opportunity in this paper to revisit ten of the most important myths that relate primarily, but not exclusively, to the linguistic and phonetic aspects of charisma, such as the interactions between verbal and nonverbal and between segmental and prosodic cues, as well as the roles of breathing and fundamental frequency in charisma perception. The result is a very diverse picture. Some myths, including very old ones, can be accepted. Others must be rejected in the light of contradicting empirical results. The status of some myths remains unsettled. Furthermore, in discussing that diverse picture, our paper points towards knowledge gaps in research and practice and gives concrete directions as to where to go from here.

Key words: Charisma, speech prosody, rhetoric, public speaking, posture, breathing, personality traits

Introduction

Research conducted in the field of charismatic leadership and speech is exemplary for the discrepancy between what we believe we know and what is actually empirically grounded. Furthermore, what we think we know has become so prevalent over time that it resulted in well-known, often undisputed myths. In this paper we address ten of the most frequent charisma myths along with the questions: What makes charisma so susceptible to myth-building? Is there empirical evidence to support prevailing statements about charisma? And if not, what may have caused these misconceptions? First of all, we find that charisma itself has been a myth from the start and to a certain degree remains a myth even today. The idea of charisma as an instrument of persuasion dates back to classical Aristotelian rhetoric (Antonakis et al., 2016). Early research viewed charisma as a mystical and even magical or alchemistical gift endowed at birth to a selected few (Weber 1947; Gemmill & Oakley 1991: 119; see Antonakis et al., 2016 for an overview). Thus, charisma, by definition, exceeded the grasp of scientific understanding in that it was not a concrete skill but a variety of different subjective traits that, together, create

https://doi.org/10.14712/24646830.2019.17

a charismatic impression. Charisma was described as a social illusion (Gemmill & Oakley 1991: 119) that magically empowered – often divinely chosen – leaders to pave the way out of a crisis (Weber, 1947). Accordingly, research on the mechanisms of charisma is relatively scarce. How can one study something that is magic and impossible to get hold of?

The scarcity of earlier empirical investigations on charisma contrasts with the need for a consistent concept of charisma and the understanding of its mechanisms when it comes to its use in economics and leadership training. It is commonly believed that charismatic leaders possess extraordinary abilities to motivate followers and to assert influence (Weber, 1947; Etzioni, 1964; House, 1977; Bass, 1985; Antonakis et al., 2016), which sparked a great deal of interest in teaching charisma and in using it as a tool for political-career and business development. This can be deduced from the vast body of popular advice literature on the topic (e.g., Mortensen, 2011; Soorjoo, 2012; Fox Cabane, 2012; Volkmann, 2013; Peters, 2015; Amon, 2016, inter alia). However, the lack of empirical research limits the applicability of the concept of charisma to everyday situations. In particular, coaches and consultants, as well as politicians and managers attempted to understand the sources of perceived speaker charisma and developed techniques to teach these sources (e.g., Fox Cabane, 2012). Over the course of time, this desire to understand that "ineffable quality that attracts, fascinates, and influences people around you" (Peters, 2015:1) without an established research paradigm or empirical background has led to a large number of assumptions based on impressionistic, anecdotal, and subjective observations, or on research that does not meet modern scientific standards. Furthermore, since those assumptions remained unchallenged by the scientific community and were continuously shared in an expanding scene of business coaching and consulting without ever having been testable hypotheses, they were declared common knowledge and became myths.

The research on charisma has expanded considerably over the past 10 to 15 years. While the earlier studies in political science, social science, and psychology laid the ground work to unravel the nature of charisma (Weber, 1947; Davies, 1954; Etzioni, 1964; Tucker, 1968; House, 1977; Bass, 1985), recent advances in leadership studies arrived at an operationalizable definition of charisma (Antonakis et al., 2016). Furthermore, charisma has been conceptualized in a set of modern as well as classical rhetorical devices that could be and have partly already been empirically tested (cf. Shamir et al., 1994; Emrich et al., 2001; Antonakis et al., 2011, 2015, 2016). In spite of this vast progress in delivering empirically grounded results and insights into charismatic leadership and speech, many researchers as well as practitioners (i.e. leaders and speakers) still rely on the established myths; the transfer from empirical research to everyday application and education is slow.

Interdisciplinary research on various facets of charisma has reached a point where it is deemed fruitful to revisit and reevaluate some of the most common myths in order to assess the actual status quo of what is known about charisma. This is precisely the objective of this paper. Moreover, and more importantly in fact, we want to challenge what is commonly declared to be known, namely 10 of the most frequent myths of charismatic leadership, particularly with respect to charismatic speech and delivery. The latter specification already implies that the 10 myths addressed here not do represent *the* top 10. It would probably be difficult to define on objective grounds and/or with reference to some

external criteria what the top-10 myths about charisma actually are. Should we look at their persistence, i.e. for how long they have already been around, or at their frequency of occurrence in literature or the internet, or should we perhaps even rely on an estimate of how harmful or beneficial they are from a socio-economic perspective? As all of this seems inappropriate, the 10 myths we address here have been selected on a subjective basis insofar as they reflect the authors' own research activities and the field of research of the special issue in which this paper is published, i.e. phonetics or, more generally, communication signals. Therefore, we by no means claim that the selected 10 myths are the most important ones, nor does our selection imply that other myths are not worth being addressed and revisited.

In the following, we investigate which statements about charisma have passed the test of time and still hold to scientific standards, which statements have to be adjusted to fit empirical data, and which statements have to be rejected entirely, either due to a shift in the concept of charisma and how it is conveyed, or for being a general misconception. Furthermore, we do not just want to confirm or reject common myths based on empirical evidence. We also seek to explain why a myth may be considered true, based on the mechanisms of charisma, as well as why some myths have emerged at all.

Myth 1: Charisma makes a difference

The interest in making charisma a learnable skill comes from the assumption that charismatic leaders possess extraordinary powers to influence people in ways uncharismatic leaders cannot (cf. Weber, 1947; House, 1977; Bass, 1985; Antonakis et al., 2016). Furthermore, these charismatic ways of influence are said to not only differ from, but also have at least equal effects as authoritative leaders can assert through the power given by their position in the social hierarchy, as well as the effects that transactional leaders can achieve through the use of incentives (Howell & Frost, 1989; Judge & Piccolo, 2004). There is common ground in the advice literature that charisma makes a difference and that leaders should strive to become more charismatic. As is stated provocatively by Peters (2015: 29): "*Charisma works* like magic, it can put you in front of other people even though you know less than others (most of the country leaders will agree on this)".

That charisma makes a difference is indeed supported by empirical evidence. One of the first empirical studies by Howell and Frost (1989) shows that a charismatic leadership style increases the quality of output and efficiency of participants. Furthermore, Howell and Frost (1989) compared leadership styles and found that the effects of charismatic leaders outperformed those of compassionate and even structuring and hence more authoritative leaders. Further evidence is provided by Towler (2003). She found that HR personnel instructed by leaders who received charisma training performed with higher precision and produced greater task quality. A recent study by Antonakis et al. (2015) also supports the extraordinary effectiveness of charismatic leaders by showing that followers instructed by charismatic leaders work much more efficiently while retaining a high level of quality. Furthermore, Antonakis et al. (2015) compared the charisma effect to the influence of financial incentives and found that they increased productivity to the same degree, making a charismatic leadership style as effective as a transactional style without the additional costs; or in the words of Antonakis published in an online transcript

by Pangambam (2016): "This charisma result is crazy because it's not well explained by current economic theory. We got increased performance, basically for free. And charisma significantly decreased production costs. We got increased performance without paying economic incentives." Additionally, studies show that vocal features of charisma alone exhibit a significant charisma effect by encouraging listeners to do more voluntary work, influencing their choice when booking a sightseeing trip or choosing healthy fruits over unhealthy sweets (Fischer, 2018), as well as affecting how willing they are to follow directions to a destination given by a car navigation system (Niebuhr & Michalsky, 2019).

Not only do we have increasing empirical evidence that charisma *does* works but also *why* and *how*. In contrast to earlier descriptions, we can assume that the power of charisma is neither divine nor magical nor entirely subjective and indescribable. As Antonakis et al. (2016) define it, charisma is a device for emotion-laden, values-based, symbolic leadership signaling. Accordingly, charisma conveys that leaders are competent and confident in their abilities, convinced of the vision they entail, emotionally invested in, and passionate about their goals and agenda and, finally, also able to signal these properties through ways of communicating. Followers are inspired by passion, convinced of their common goal and vision through confidence in their leader, and consequently develop an intrinsic motivation to work for their common goal rather than merely because they have to obey orders. This is also supported by empirical evidence. In the study by Howell and Frost (1989), the participants in the charisma group reported higher satisfaction with both the task and the experimental environment in general. They were also much less or not at all affected by the expressed motivation of their peers in the groups, which is an indicator of robust intrinsic motivation. The charisma group in Towler's (2003) study also reported higher overall satisfaction. Furthermore, they reported perceiving the charismatic instructor as more efficient, competent, and convincing. Ning (2019) shows that participants in a brainstorming workshop rate themselves as more intelligent, unconventional, and capable, when the workshop is given by a more charismatic moderator. Lastly, the study by Niebuhr and Michalsky (2019) shows that listeners even project the charismatic features of trustworthiness and competence to a computer system when receiving instructions with a charismatic tone of voice.

Bottom line: The myth that charisma makes a difference is valid. The assumption that charismatic leaders inspire followers to achieve more and higher quality work is supported by a number of empirical studies. Their results even suggest that charismatic leaders not only surpass uncharismatic ones, but also that the effect of charismatic leadership outperforms the classical structuring leadership style that relies on authority from the social hierarchy, and that the effect of charismatic leadership achieves the same results as financial incentives given in a transactional leadership style. Furthermore, studies suggest that leaders who communicate in a charismatic way are perceived differently by their followers in terms of motivation, passion, confidence, and competence, which significantly affects self-reported satisfaction, motivation, confidence, intelligence, and capability.

Myth 2: Charisma is a divine talent of a few gifted people that only surfaces during a crisis

Max Weber describes charisma based on the collective prior observations and concepts as "an extraordinary power, giving leaders salvationist qualities to deliver followers from great upheaval" (Weber, 1947, 1968). Accordingly, in this and earlier accounts, charisma was not only assumed to be a magical and indescribable feat, it was also considered an innate talent that only a few gifted people were able to develop during times of great need. Moreover, even those chosen few could not deliberately improve or change their charisma through training. It emerges and is shaped as a reaction to difficult times, specifically a crisis, and is spawned and enhanced by the people's need for a charismatic leader. Accordingly, charismatic speech should not be learnable or improvable by anyone, not even by those who possess the innate gift.

This perspective has been challenged from its inception by several researchers from psychology through the social sciences to business and management research; see, for example, Etzioni (1961), House (1977), or Shamir and Howell (1999). In addition, the modern advice literature arrives at the claim that charisma can in fact be learned by anyone:

"There's an often repeated myth that you're either born a great pitcher or you're not.[1] This myth simply provides a justification for not preparing properly and an excuse for why pitches fail. The truth is that these so-called naturals put in days, and sometimes weeks, of preparation and use an array of proven strategies and techniques to consistently win over their audiences" (Soorjoo, 2012:xv).

Furthermore, it is suggested that charisma has to be trained even by the most proficient natural talents:

"We understand that proficiency at chess, singing, or hitting a fastball requires conscious practice. Charisma is a skill that can also be developed through conscious practice [...]. I know that a person's charisma level can be changed because I've helped countless clients increase theirs in this way" (Fox Cabane, 2012: 7).

What was assumed by earlier studies and additionally derived from anecdotal evidence in the advice literature has since been supported by empirical studies. The study by Howell and Frost (1989) investigated the results of charismatic, compassionate, and structuring leadership by training actors such that they were able to consistently apply the respective leadership styles. The results suggest that a charismatic leadership style can be convincingly learned and displayed by trained actors. Frese et al. (2003) extended the study of learnability of charisma to top managers and business leaders in a controlled design. They found that the charisma group improved in all relevant parameters of charisma such as communicating a vision, developing a collective identity, and having a stronger, more confident and more dynamic and expressive appearance as well as a so-called "captivating" tone of voice. Towler (2003) pushed this line of research further by teaching business students and achieved the same effects, i.e. an increase in symbolic

[1] A "pitch" is a specific form of public speech given in business contexts. It "is usually less than two minutes in length, provides an initial glimpse of [a] venture idea with the goal of engaging the investor in further conversation and, ultimately, obtaining financing" (Clingingsmith & Shane, 2017: 5164). In a more general sense, a pitch is any kind of public oral presentation that aims at persuading listeners.

communication and a captivating tone of voice through charisma training. The study by Antonakis et al. (2011) was again aimed at managers and business leaders. However, they tried to teach charisma through a very specific set of Charismatic Leadership Tactics (CLTs) to make charisma training more efficient and tangible. Their study shows that all CLTs could be successfully learned by naïve speakers, and that they significantly increased the participants' perceived leader prototypicality.

The second author of this paper runs a 12-week course on *Persuasive Communication and Negotiation* that is mandatory for all master's students in electrical and business engineering at the University of Southern Denmark. It starts with the verbal CLTs and, based on this foundation, puts the focus on nonverbal aspects of body language and, in particular, speech melody. At the end of the lecture term, the students' presentations are rated with respect to perceived speaker charisma by an expert panel of lecturers and company leaders as well as by a sample of naïve listeners in an online experiment (between 50–100 people each year). Both experts and naïve listeners receive paired stimuli representing each student's baseline performance at the beginning of the course and his/her trained performances at the end of the course. The order of the stimuli within a pair as well as of the pairs themselves is randomized, and pairs are presented several times. The listeners' task is to indicate in which of the two compared presentations the speaker is more charismatic and to rate the higher level of charisma on a scale from 1 to 10. Results show for all classes taught so far that the students have an about 40–90 % higher perceived charisma level at the end than at the beginning of the course, in the ears of experts even more so than in the ears of naïve listeners. Research based on automated acoustic charisma quantification shows additionally that a 4-hour intensive training of a speaker's voice can significantly increase speaker charisma by 10–50% and that female speakers benefit from such training more than male speakers do (Niebuhr et al., 2019).

Bottom line: The myth is busted that charisma is an innate talent that only manifests in times of crisis. Several studies have shown that signaling charisma can be trained by naïve speakers ranging from professional actors, through managers and business leaders to business and engineering students. However, there seem to be restrictions as to the degree to which charisma can be learned. As Antonakis et al. (2016) point out, proficiency in charismatic communication may be related to intelligence and/or creativity, as some CLTs such as metaphors and storytelling require a higher degree of creativity, innovation, and planning. Furthermore, as we discuss in the next chapter, the ability to convey charisma may be affected by personality or, as we see in Myth 10, general social skills or professional mindsets.

Myth 3: Charismatic communication is the expression of a charismatic personality

There is one common thread that connects all previous investigations of charisma. Weber (1947) described charisma as an extraordinary property of charismatic leaders. Charismatic leaders are said to be able to connect with their followers (Davies, 1954) and to possess a vision as well as confidence in their ideals and competence (Tucker, 1968). Furthermore, charismatic leaders are idols and exemplary personalities who care about their image (House, 1977). What all these assumptions have in common is that charisma

is something that charismatic leaders *have* and which constitutes a part of their personality. That is, charisma is considered a personality trait. This basic notion was already largely rejected in the discussion of the last myth. As we have seen, charisma can be learned and improved. As described above, Antonakis et al. (2011, 2016) reject charisma as a property of charismatic leaders. Rather, charismatic leaders are not charismatic per se but possess the ability to communicate in a charismatic way. Consequently, in the studies by Howell and Frost (1989), Frese et al. (2003), and Towler (2003), the participants did not learn to become more charismatic people. They acquired communicative skills to convey charisma.

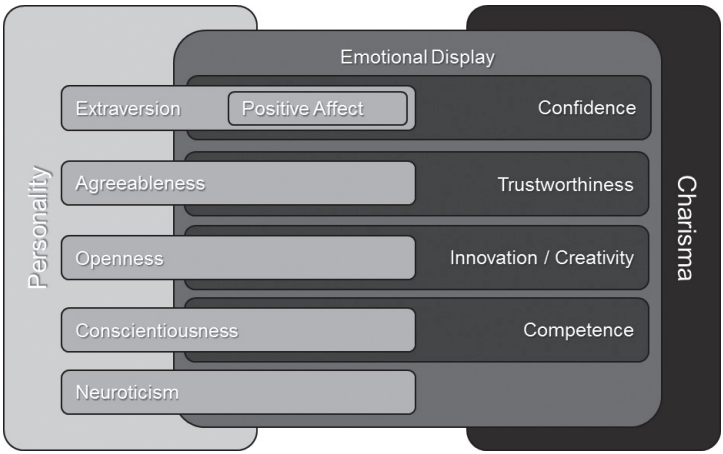

Figure 1. Intersections between personality types according to the Big Five Personalities paradigm (Costa & McCrae 1992) and features of perceived charismatic personality (Antonakis et al. 2016).

However, even without perceiving charisma as a personality trait, there may still be traits that significantly influence a speaker's ability to acquire a charismatic way of communicating. That is, certain personality types foster ways of communicating that coincide with charismatic speech; see Figure 1 and Peters (2015). There are in fact two integral parts to charismatic speech according to Antonakis et al. (2016): confidence and self-assuredness, as well as passion and display of emotions. When it comes to the big five personality traits that are commonly used for assessing speaker personalities (Costa & McCrae, 1992; Sharma et al., 2013), both confidence and displaying emotions relate to the personality trait of *extraversion*. When it comes to the phonetic properties of charismatic speech, we find similarities between the phonetic manifestations of extraversion and charismatic speech. Acoustic characteristics associated by listeners with extraversion, such as an expanded fundamental frequency (f_0[2]) range, an elevated f_0 mean, more frequent and deeper final falls, as well as a higher speaking rate, resemble the acoustic characteristics of charisma (Michalsky et al., 2019). Furthermore, extraversion is linked

[2] Fundamental frequency, or f_0, is the acoustic correlate of the vocal-fold vibration frequency in speech production; f_0 is used by listeners as the main source of pitch perception. Thus, in a nutshell, f_0 movements in speech represent a speaker's speech melody.

to *positive affect*, a trait which is also related to higher confidence, higher personal goals, as well as to experiencing and expressing positive emotions more frequently (Costa & McCrae, 1992; Curhan & Brown, 2012). The trait of *agreeableness* also relates to concepts relevant for expressing charisma. Agreeable speakers are assumed to be kinder and warmer, which relates to the ability of charismatic leaders to connect with people and, furthermore, to show a greater ability to express and develop trust (Costa & McCraw, 1992; John & Srivastava, 1999). *Conscientiousness* is related to self-discipline and overall job performance (Costa & McCraw, 1992; John & Srivastava, 1999; Barrick & Mount, 1991). Lastly, *openness* can be regarded as a measure of imaginativeness and divergent thinking (Costa & McCrae, 1992; John & Srivastava, 1999). Since a significant portion of the CLTs that result in the necessary emotional symbolic communication is contributed by strategies such as metaphors and storytelling, creativity and imaginativeness constitute integral parts of charismatic communication (Antonakis et al., 2011).

Bottom line: The myth that charismatic speech and communication are mere expressions of a charismatic personality has been largely busted in its strict form. Charisma itself is a way of communicating that can be learned, improved, and implemented independently of personality traits. There are good reasons to assume that certain personality traits naturally lend themselves to support charismatic speech. However, we saw that there is no single charismatic personality type. Rather, it is a mixture of different personality types that supports speaker charisma. People may communicate in charismatic ways through different strategies by approaching charisma either through the self-assured and confident facet of being extraverted, through the passionate and emotional facet of being positively affective, through the expressive/symbolic facet of being high in openness, through the trustworthy facet of being agreeable, through the competent facet of being conscientious, or through a combination of any of these. It is unlikely that the majority of charismatic leaders possess a charismatic personality that includes all or only just the majority of these traits; and it is even more unlikely that this is required to learn a charismatic performance. That is, charisma itself is not necessarily a matter of personality.

Myth 4: How we say something is more important than what we say

Another frequently reappearing assumption about charisma is that the delivery of a message matters more to the listener than the message's content. However, we first have to establish where to draw the line between *how* and *what*, since this issue in itself is a controversial topic (see Figure 2). From a linguistic perspective, it is reasonable to distinguish between linguistic content and paralinguistic delivery. However, in classical rhetoric and, in fact, in the majority of psychology, social-science, and management/leadership studies, content and delivery are already separated at a linguistic level. Following classical rhetorical research as well as Antonakis et al. (2016), only the speaker's propositions are regarded as content, whereas linguistic rhetorical devices like rhetorical questions, lists, contrasts as well as metaphors, analogies, and devices of storytelling are all part of the delivery. Although this distinction is most common and also adopted in this paper, there are other approaches arguing that rhetorical strategies belong to the

content rather than the delivery (Shamir et al., 1994). Following Antonakis et al. (2016), the myth would claim that how we say something both in term of voice as well as rhetorical strategies and visual cues outweighs the propositions of a speech. This is supported by the aforementioned study by Antonakis et al. (2015) who tested the effect of the same propositions delivered using different linguistic, visual, and prosodic devices. Although the speakers' propositions were identical, the delivery strategy significantly affected perceived charisma.

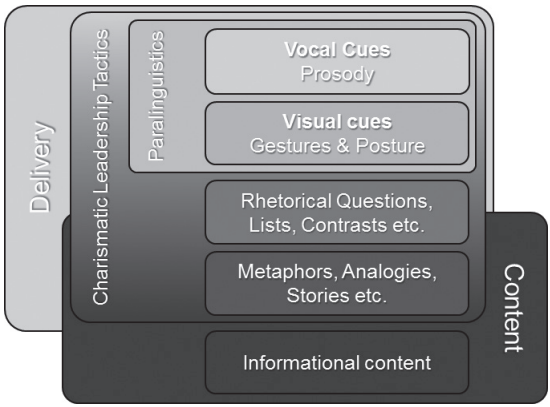

Figure 2. What we say and how we say it. Different classifications of content and delivery.

The other common distinction is the separation of linguistic content from paralinguistic delivery, with paralinguistics encompassing both vocal and visual features, but not linguistic rhetorical strategies. Accordingly, the myth would claim that how we say something in terms of acoustic-prosodic and visual (i.e. overall nonverbal) cues outweighs the propositional as well as the linguistic content. This distinction is frequently found in the advice literature. As Soorjoo (2012: 20) points out: "Yet, when it comes to preparing a pitch, most people tend to focus on the content of their speech and their PowerPoint... This is one of the principal reasons why most people deliver bad pitches". Fox Cabane (2012) also claims that "nonverbal modes of communication are hardwired in our brains, much deeper than the more recent language-processing [i.e. word-related] abilities, and they affect us more strongly" (p. 89). This common assumption manifests itself also in proverbs like "hitting the right note".

The dominant role of vocal features in the field of charismatic persuasion becomes apparent in several empirical studies. Holladay and Coombs (1993) found that if there is an apparent contradiction between a speaker's verbal and non-verbal message, the latter is more likely to influence a listener's perception (see also Holladay & Coombs, 1994). Towler (2003) conducted a principal component analysis to separate different contributors to charisma and found that vocal features provided an independent and crucial effect. Pentland (2008) investigated the impact of paralinguistic cues including both gestures and acoustic-prosodic features in several social settings and found paralinguistic cues to successfully predict the success of investor pitches, the exchange of business cards after a meeting, the success in acquiring new customers, and even the exchange of phone

numbers after speed dating. Several comparable findings on the dominance of delivery over content have been made (Awamleh & Gardner, 1999; Gregory and Gallagher, 2002; Park et al., 2014). Just recently, Caspi et al. (2019) conducted two experiments on the topic and found that delivery significantly outweighs the content when it comes to a first impression of a speaker, which is supported by the findings of McAleer et al. (2014), who found that the delivery of a simple "hello" already critically affects the impression we make about a speaker personality. Sometimes even the visual cues such as gestures and posture are excluded from the *how* and the major role is solely attributed to the acoustic-prosodic features of charisma. Amon (2016) claims that "there is a superiority of the audible impression over the visible. The moment you open your mouth, all the visible elements become mere decoration" (Amon 2016: 12, the authors' translation). In their multimodal analysis of speaker charisma, Scherer et al. (2012) found that auditory cues alone affect perceived charisma and enhance and even shift the interpretation of visual cues. Chen et al. (2014), who conducted a similar analysis, arrived at the conclusion that auditory cues outweigh visual cues as a predictor for charisma. Lastly, the studies done by Fischer (2018) as well as Niebuhr and Michalsky (2019) show that computer voices possess charismatic influence even if the lexical material is identical, visual cues are absent, and only acoustic cues serve to signal charisma.

Bottom line: Regardless of whether we include rhetorical strategies in the delivery side of speech, restrict ourselves to paralinguistic cues, or to acoustic-prosodic features alone, empirical research supports that *how* we deliver a speech significantly contributes to its charismatic and persuasive impact. Furthermore, we can assume that the delivery is more important in signaling charisma. However, although the *how* is essential for a charismatic performance, to date there is not a single study that actually compares the charismatic effect of content against the charismatic effect of the delivery for neither definition of delivery. Following Emrich et al. (2001) it is possible that delivery is crucial for the immediate impact of a charismatic performance but the effects diminish in the long run, if not supported by the content (see also Caspi et al., 2019). Accordingly, although this chapter strongly suggests that the *how* outweighs the *what,* we can neither completely reject nor accept the myth in this simple form.

Myth 5: Lower voices are more charismatic

"How to Train Your Voice to Be More Charismatic?" In answering this question, Nancy Daniels (2013) points her readers to the study of Mayew et al. (2013). Based on voice-pitch analyses of 792 leaders (CEOs of major companies) around the globe, and controlling for other confounding factors, Mayew et al. conclude that low voices make better leaders. More specifically, speakers showing an interquartile decrease in f_0 level of 22.1 Hz enjoy longer tenures (about 151 days longer), lead larger and higher-valued companies (by about $440 million) and, thus, earn more money (about $187,000 more per year). Carnegie and Esenwein (2011: 32) also criticize in their rhetoric manual the fact that "most speakers pitch their voices too high" while presenting; and Barker (2011) paraphrases the same criticism in the form of an imperative: "Create vocal music that is lower in tone, slower and softer, and you will create rapport more easily" (p. 14), which

is later in the book narrowed down to f_0 alone: "Lower your tone. A thin, high-pitched voice will suggest a lack of authority or confidence" (p. 175).

Together, these interdisciplinary findings, statements and instructions seem to form a coherent whole: The lower-pitched you speak the more charismatic you sound. In fact, exactly the opposite is true. When looking beyond management and psychology studies and rhetoric manuals and anecdotes, readers will quickly find consistent evidence from the experts in that matter, i.e. speech scientists, that the correlation between a speaker's f_0 level and his/her perceived speaker charisma is positive, not negative. This finding was made, for example, by Touati (1993), Strangert and Gustafson (2008), Biadsy et al. (2008), Rosenberg and Hirschberg (2009), D'Errico et al. (2013), Berger et al. (2017), Jokisch et al. (2018), and Niebuhr and Skarnitzl (2019) whose studies cover languages that range from English to German and Swedish to Italian, French, and Arabic.

In the light of such obvious and abundant counterevidence, why does the myth that lower voices are more charismatic still persist? There are several reasons. First, charisma is a fuzzy semantic concept, and studies advocating lower pitched voices often do not investigate charisma, at least not in its current prototypical sense. Recall that charisma is defined as the ability to gather and win over people and determine their opinions, attitudes, and actions, without exercising authority and control and without using formal mechanisms (cf. Antonakis et al., 2016). As Smith (2010) puts it, charisma "equals persuasion with force", with persuasion being based on emotional contagion. In contrast, studies advocating lower pitched voices often refer to terms like dominance, authority, and power. In short, it is the lack of a clear distinction between dominance and authority on the one hand and charisma on the other that creates the inconsistency in voice-pitch related recommendations to speakers. A low-pitched speaker conveys power and authority and, on this basis, tells people what to think and do. A high-pitched speaker conveys charisma and, on this basis, makes people adopt his or her point of view so that the intended thoughts and actions are elicited on a voluntary basis.[3]

Figure 3 shows the pitch levels of two undoubtedly charismatic speakers, Barack Obama and Steve Jobs, in relation to the frequency of occurrence of voice-pitch levels among the populations of male and female American English speakers. As can be seen, both Obama and Jobs speak at such high pitch levels (217 Hz and 232 Hz, respectively) that they already fall in a voice-pitch range that is characteristic of female speakers in American English (see Niebuhr et al., 2016; D'Errico et al., 2019 for the sources of the mean values of the two speakers). Although these female speakers were reading calibration sentences whereas Obama and Jobs were giving public speeches (with a louder and hence inherently higher-pitched voice), this is still a remarkable observation against the background of a myth claiming that it is a low-pitched voice that makes a charismatic speaker.

The second reason for the persistence of the myth that lower voices are more charismatic lies in the confusion of local and global pitch levels. While a speaker's global pitch

[3] However, note that, although dominant and authoritative speakers have a lower voice than charismatic speakers, their average pitch level is not extremely low but still within the lower mid of their pitch range, probably because of a high loudness level and a way of speaking that is meant to convey urgency and righteous indignation. Humble speakers can still have a lower pitch level than dominant/authoritative speakers, especially in dialogue situations (D'Errico et al., 2019).

level should be overall higher to be more charismatic, charismatic speakers must also be able to get down to the bottom of their individual pitch ranges at certain local points in their sentences. This primarily applies to the pitch valleys in between to expressively stressed, high-pitched words and, in particular, to the ends of sentences. This is very well described by Fox Cabane (2012: 89): "imagine an assertion: a judge saying 'This case is closed'. Feel how the intonation of the word 'closed' drops. Lowering the intonation of your voice at the end of a sentence broadcasts power. When you want to sound superconfident, you can even lower your intonation midsentence." In accord with this statement, Mixdorff et al. (2018) showed that charismatic speech means a raising of pitch peaks and the speaker's overall pitch level and, at the same time, a lowering of the "baseline f_0", i.e. those local levels at which a speaker begins his/her pitch rises towards stressed words and ends his/her sentence-final pitch falls.

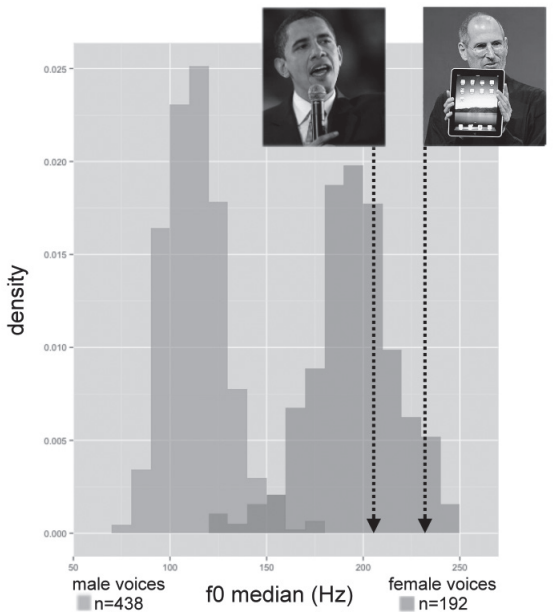

Figure 3. f_0 medians of Steve Jobs and Barack Obama during public speeches, compared to the f_0 medians of 630 male and female speakers of American English who read the two calibration sentences of the TIMIT database. The histogram was edited after Liberman (2013). The photographs were edited and added to the figure based on a Wikipedia creative commons license.

The third reason is related to the relevance of the pitch-level factor for a speaker's charismatic impact. Although the positive correlation between pitch level and speaker charisma is strong and significant across languages, the perceptual relevance of the factor pitch level is actually rather low. Several perception experiments arrived at the conclusion that other factors like speaking rate, pitch range, pause frequency and duration, as well as filled pause characteristics have a much stronger effect on listeners' speaker charisma ratings than a speaker's global voice-pitch level (Berger et al., 2017; Niebuhr et al., 2017). Thus, even if speakers follow the recommendation to lower their pitch level instead of

raising it in order to sound more charismatic, this wrong decision has no strongly negative effect on their overall charismatic impact, as long as they perform well along other more important voice factors.

Bottom line: The myth that lower voices are more charismatic is busted by cross-linguistic empirical evidence from the speech sciences. Speakers should raise rather than lower their global pitch level. However, local pitch levels that are reached between expressively stressed, high-pitched words and at the ends of utterances should indeed be lowered to the bottom of a speaker's individual pitch range, i.e. where the voice starts getting creaky (i.e. irregular and crackling) in the case of male and breathy in the case of female speakers.

Myth 6: A clear pronunciation supports perceived speaker charisma

"Distinct and precise utterance is one of the most important considerations of public speech. How preposterous it is to hear a speaker making sounds of 'inarticulate earnestness' under the contended delusion that he is telling something to his audience! [...] Telling means communicating, and how can he actually communicate without making every word distinct?" (Carnegie and Esenwein, 2011: 146). Not every statement is as strong as the one above from *The Art of Public Speaking*, but virtually every rhetoric manual urges its readers to "clearly articulate every phrase and word" and to internalize that "good articulation conveys competence and credibility" (Mortensen, 2011: 158) and, thus, "is imperative to develop charisma" (Camper Bull, 2010: 138); see also Frese et al. (2003).

The contribution of articulatory precision to a speaker's charismatic impact is, unfortunately, not as well studied as the contribution of prosodic features like loudness, speaking rate, and pitch level or range, but the few studies that exist basically back up the rhetorical statements on articulation. For example, Niebuhr (2017) conducted a perception experiment whose stimuli included, amongst other things, a naturally produced systematic variation in the degree of speech reduction. Three controlled reduction steps were created: (i) sentences in which each word is pronounced in its full, dictionary-like fashion; (ii) sentences in which each word is pronounced like in an informal everyday conversation, i.e. slightly reduced in the case of content words and moderately reduced in the case of function words; (iii) sentences in which both content and function words are all equally pronounced as strongly reduced as possible. Results show that a constantly strong reduction makes speakers sound significantly more absent-minded, stressed, and clumsy and less trained/skilled, sociable, educated, optimistic, and sincere, i.e. overall less charismatic. This perception evidence is in accord with production evidence from a comparison of Steve Jobs and Mark Zuckerberg. Steve Jobs, who is perceived to be more charismatic than Mark Zuckerberg both by representatives of the media and listeners in a controlled perception experiment (Niebuhr et al., 2018a), performs significantly better than Zuckerberg in acoustically distinguishing his voiced and voiceless stop consonants (/p t k/ vs /b d g/) as well as the different vowel qualities of American English. The acoustic vowel space that Jobs uses in his speech is at least 32.7% larger than that of Zuckerberg (Niebuhr and Gonzalez, 2019). Furthermore, Jobs' speech includes 28.3% fewer instances of post-lexical assimilation of alveolar consonants (/t d n/) to either bilabial or velar plac-

es of articulation than Zuckerberg's speech (Niebuhr et al., 2018a). This applies to content words; the relative difference between the two speakers is even larger for function words.

Finally, that articulatory precision contributes to a speaker's charismatic impact also makes sense in terms of the fundamental ethological principle of the Effort Code (Chen et al., 2002). In a nutshell, the Effort Code conceptualizes that the importance of a certain matter is positively correlated with the energy that is invested in addressing it (Gussenhoven, 2016). In speech, this means that whatever a speaker considers (more) important is realized with great(er) articulatory effort; and greater articulatory effort, in turn, "tends to create more elaborate and more explicit phonetic realisations" (Chen et al., 2002: 211), i.e. a clearer pronunciation. Thus, in terms of the Effort Code, the charisma effect of articulation is explained by a clearer pronunciation being an implicit signal of "I have something important and meaningful to say" and/or "you, my listeners, are important to me". Barker (2011: 176) explains speaking clearly to his readers as follows: "Make sure all the consonants are clear when you are speaking (all the letters that are not A, E, I, O or U)". Such a restriction of articulatory effort to consonants alone does not make sense in terms of the Effort Code; nor is there, to the best of our knowledge, any empirical evidence that persuasion relies more on consonant than on vowel articulation. Therefore, such recommendations should be treated with caution. The actual reason for Barker's focus on consonants is that they, unlike vowels, most often include a sensible contact between active and passive articulators, which makes self-monitoring and articulatory control easier for speakers (Abercrombie, 2000).

Bottom line: The myth that a clear pronunciation supports perceived speaker charisma is valid. It is consistent with empirical evidence as well as with theoretical concepts like the Effort Code. If one wants to qualify the myth, then by noting that realizing each and every word with a dictionary-like pronunciation can reverse the positive effect of a clear pronunciation and attenuate speaker charisma again. For example, Niebuhr (2017) found that such an "overarticulated" way of speaking makes a speaker sound more vain and less composed and sincere. However, without special training or instruction, native speakers are unlikely to attain this overarticulated level of pronunciation, as speech reduction probably belongs to the universal characteristics of spoken language (Clopper and Turnbull, 2018) and comes naturally to speakers through semantic or frequency effects or biomechanical and physiological limitations of the speech production apparatus (Cangemi et al., 2018). Therefore, it is reasonable to assume that the practical risk of strongly and constantly overarticulating one's presentation is low, except, maybe, for non-native speakers, but this is a question that needs to be addressed in future studies. Until then, all speakers should aim at a "crisp clear pronunciation" (Seet, 2013) when performing a speech. Whether this applies to the same degree for vowels and consonants is also a matter of future research.

Myth 7: Filled pauses are bad for perceived speaker charisma

It is a common statement in rhetoric manuals that speakers should, as much as they can, avoid all the *errs*, *uhs*, *urns*, *ums*, and *mhs* in their speech that are referred to as filled pauses (or hesitation markers/disfluencies). For example, Sprague et al. (2013: 336) make

the following recommendation in their *Speaker's Handbook*: "Do not be afraid to pause between sentences or thoughts when you speak. But avoid filling those pauses with distracting and meaningless sounds and phrases [...]". Similarly, Soorjoo (2012: 26) states that silent pauses are an effective way to "eliminate distracting nonwords such as *ums* and *uhs*" from a speaker's speech. Learning to self-monitor one's speech and, on this basis, to anticipate and replace filled pauses by silent pauses is also a key point in the "3 tips to eliminate filled pauses from your professional presentation" by Bell (2011).

At first glance, these strong statements and specific recommendations are backed up by empirical evidence from the speech sciences. For example, Biadsy et al. (2008) as well as Rosenberg and Hirschberg (2009) found, across languages, a negative correlation between filled pauses (and self-repairs) on the one hand, and charisma-related ratings of a speaker on the other. The works of Niebuhr et al. (2016, 2019) are consistent with these findings. Comparing the more charismatic Steve Jobs with the less charismatic Mark Zuckerberg revealed that the frequency of filled pauses (and other disfluencies) represents one of the biggest differences between the two speakers, with Jobs using 46.2% fewer filled pauses than Zuckerberg. Moreover, indirect perception evidence of Niebuhr and Fischer (2019) suggests that filled pauses are one of the major factors for the perceived charisma differences between the two speakers.

At second glance, however, the strong statements in rhetoric manuals need to be qualified in at least two respects. First, filled pauses per se are not bad – either for the comprehension and memory of a speaker's messages, or for his/her perceived charisma and related traits. Rather, the opposite is true. Filled pauses fulfill important communicative functions. They facilitate the listeners' cognitive processing of the upcoming information (in that they typically occur before less frequent words and/or new information); Corley and Hartsuiker (2003) call this the "um advantage". In addition, they indicate to listeners through their specific phonetic form how long they will have to wait until the speaker continues talking (Fox Tree, 2001) and whether the speaker continues with the same or a different message (Fischer, 2000). Furthermore, Fischer (2000) and Fruehwald (2016) stress that filled pauses serve important social functions in speech, such as mitigating potentially impolite utterances (Levinson, 1983; Schegloff, 2010) and showcasing a speaker's affiliation to a specific cultural or social group.

The second, more important reason why filled pauses should not simply be eliminated from a speaker's speech is that they convey spontaneity and listener-orientation. That is, they are critical "contact signals" (cf. Fischer, 2006). In accordance with that, Novák-Tót (2016) found that the former CEO of Hewlett-Packard, Meg Whitman, sounded less charismatic in the ears of listeners than the CEO of IBM, Virginia Rometty. This difference was, amongst others, traced back to the frequency of filled pauses, but not in the sense that Whitman used significantly more filled pauses than Rometty. Rather, Whitman used almost no filled pauses at all within almost 20 minutes of analyzed speech, as compared to 13 filled pauses in the case of Virginia Rometty (and 35 in the case of Steve Jobs; see Novák-Tót et al., 2017). More in-depth analyses in separate perception experiments show that the complete absence of filled pauses makes a speaker appear self-referred, arrogant, distant, and sounding as if s/he were reading a text rather than presenting a message.

Thus, recommendations of manuals that speakers must try to eliminate filled pauses altogether from one's speech are wrong and should not be followed. Obviously, it is the

dose that makes the poison; and indeed some rhetoric manuals do point readers to that fact, but only in unspecific ways that are of limited help for speakers. For instance, Bell (2011) briefly notes that "using a filled pause one in 100 words is not problematic, using filled pauses one in five words is a big problem." (p. 12). Bell is one of few who provide the reader with specific numbers. Yet, the numerical range is huge and hence of limited practical help, and the statement is still oversimplified. Niebuhr and Fischer (2019) show in a charisma-related study on filled pauses that it is not the mere total number or relative frequency of filled pauses that matters for a speaker's impact on listeners, but the duration of filled pauses and the degree to which they are realized as a nasal element (e.g., *mmm*). The shorter and more nasal filled pauses are, the more do listeners underestimate their actual physical number and frequency in a speaker's speech and the higher they rate the speaker's presentation performance. In other words, rather than trying to get rid of filled pauses (at the additional risk of losing listener-orientation and spontaneity), speakers should rather learn to produce short and nasalized filled pauses. That is, long *errs* and *uhs* should be replaced by shorter *ums* and *mhs*. More specifically, filled pauses with duration up to one syllable (300–400 ms) have only a marginally negative impact on a speaker's perceived performance, and filled pauses that consist more of a nasal than of a vowel sound can even add to a speaker's perceived performance (Niebuhr & Fischer, 2019). To what extent this is language-dependent still needs to be determined. For now, it seems to hold at least for Western Germanic languages.

Bottom line: The myth that filled pauses are bad for perceived speaker charisma is busted, at least in this general form. Filled pauses perform important communicative functions, and trying to reduce one's filled pauses is only useful if their number is exceptionally high (> 8 items per minute). Working on the quality of filled pauses is more effective in terms of improving speaker charisma.

Myth 8: Belly breathing and an upright posture support speaker charisma

There is hardly any rhetoric manual without a chapter of 10 pages or more that is specifically dedicated to breathing. In books like Kraus (2015) and Volkmann (2013), the breathing chapter represents 8–17% of the entire text. In these chapters, authors often stress the relevance of the so-called "belly breathing" that relies on the speaker's diaphragm rather than on his/her inter-costal muscles whose activity is associated with "chest breathing". For example, Fox Cabane (2012: 192) reminds her readers: "make sure you're breathing deeply into your belly". Similarly, Carnegie and Esenwein (2011: 223) claim that "deep breathing – breathing from the diaphragm – give[s] the voice a better support [and] a stronger resonance" both of which are assumed key features of the art of (persuasive) public speaking. Likewise, it is concluded in Speech-and-Voice (2019) that "For optimal voice usage and projection, proper breathing must come from the midsection or diaphragm" (see also Goman, 2008 and Volkman, 2013). Barker (2011) draws a direct connection between belly breathing and persuasive (charismatic) speech by stating that "the deepest kind of breathing, which works from the stomach rather than the upper part of the lungs [...] works wonders for the voice: it gives it depth and power, and makes for a more convincing delivery" (pp. 132–133).

Furthermore, the beneficial effect of belly breathing on public-speaking performance is often linked to an upright posture. Figure 4 shows three examples of breathing exercises from online and offline public-speaking manuals. They all recommend belly breathing, and, while instructions for experiencing and training belly breathing occasionally also include sitting and lying postures, the ultimate application of belly breathing during public speaking, as well as the related warm-up, is always closely tied to a standing posture. Standing upright, so rhetoric manuals claim, supports belly breathing in a charismatic speaking scenario and, moreover, "communicates a message of confidence" (Hargrave, 1995: 52) and similar desirable traits of a charismatic speaker (which is why an upright posture is also addressed additionally in the chapter(s) on body language in rhetoric manuals). As Fox Cabane says: "Be the big gorilla" (p. 251).

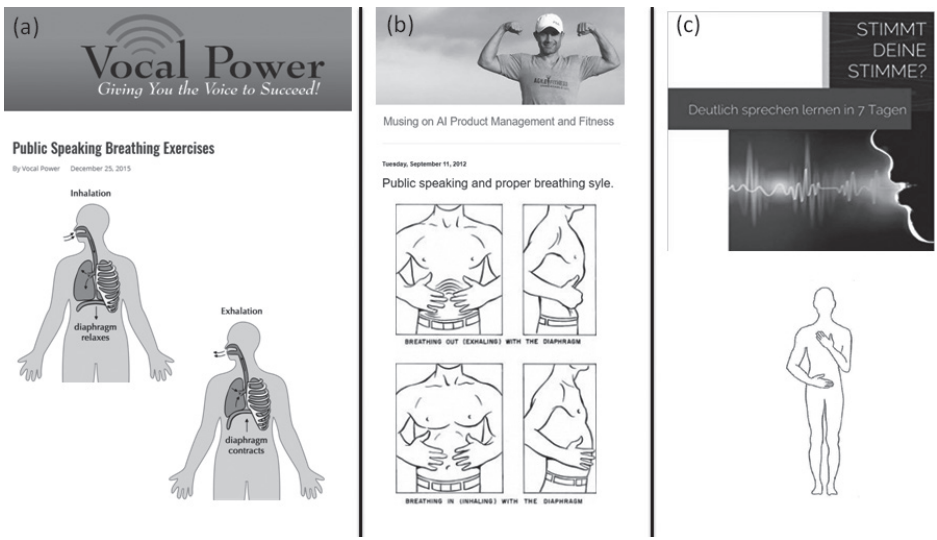

Figure 4. Three examples of belly-breathing instructions in the advice literature illustrated by means of a standing posture; (a) http://engage.vocalpower.ca/engage/public-speaking-breathing-exercises-2340, (b) http://www.manjeetjakhar.com/ 2012/09/public-speaking-and-proper-breathing.html, (c) Nico Kraus (2015); links were last accessed on June 3rd, 2019.

Barbosa and Niebuhr (forthcoming) recorded time-aligned speech and breathing signals (both belly and chest) of 18 native speakers of German, 9 men and 9 women, all of them early-stage entrepreneurs with experience in public speaking. The speakers were recorded while giving an investor pitch in front of an audience of peers, once while sitting and once while standing (the order was balanced across speakers). A sample of 21 German listeners rated the speakers' performances in terms of two criteria: (1) how charismatic (persuasive/confident/inspiring/passionate) does the speaker sound? (2) how resonant (relaxed/rich/sonorant/full) is the speaker's voice? Ratings were made on a scale from 1–6 (German school grading system). Additionally, acoustic-prosodic measurements positively correlated with speaker charisma were taken, including the levels, variation, or ranges of f_0 and intensity.

Results show that the acoustic-prosodic measures do not benefit from belly breathing and an upright posture and that speakers do not sound more charismatic when they mainly rely on belly breathing and present in a standing rather than sitting posture. Rather, the opposite was found. That is, speakers who mainly used chest breathing while presenting were those whose acoustic-prosodic measurements and perceived-charisma ratings went up. Belly breathing had an effect as well, but only in terms of the perceived resonance of a speaker's voice. The more a speaker relied on belly breathing while presenting, the more resonant was his/her voice perceived by listeners. Barbosa and Niebuhr have, in the meantime, more than doubled the speaker sample and extended it to Danish and Russian speakers. The overall results pattern remains the same according to pilot tests.

Bottom line: The myth that belly breathing and an upright posture support speaker charisma is busted. It is true that belly breathing has a favorable effect on voice quality, which is also consistent with studies on singing and speech pathology (Salomoni & van den Hoorn, 2016; Thorpe et al., 2001; Xu et al., 1991). However, this favorable effect does not include those acoustic-prosodic parameters that listeners use when rating speaker charisma. Given that, what the experiments of Barbosa and Niebuhr have falsified (based on the currently analyzed data) is the following implicit conclusion of rhetoric: Belly breathing is good for the voice and, therefore, it must also be good for charismatic speech. In fact, it is chest breathing that is good for charismatic speech; and this finding makes sense if it is looked at from the following angle: The positive effect of belly breathing in singing and speech pathology is associated with maintaining a powerful (i.e. loud), long exhalation phase. However, when it comes to speaking skills, it is the shorter rather than the longer prosodic phrase that makes speakers sound more charismatic (Biasdy et al., 2008; Rosenberg and Hirschberg, 2009). Thus, if charismatic speakers have to split up their messages into small acoustic sound bites of 2–3 s, why should they then benefit from belly breathing? Such short, impulse-like speech bites, often combined with very short, intensive inhalation phases, are better supported by chest breathing, for example, due to the intercostal muscles having a larger number of fast muscle fibers (Polla et al., 2004). Thus, the presented empirical evidence suggests to not invest too much time in learning to use and control belly breathing for public speaking. It gives speakers no measurable or perceivable advantage. The same applies to an upright posture. It is safe to give a charismatic presentation while sitting, at least in terms of speech acoustics and perception.

Myth 9: A charismatic performance requires intensive training on part of the speaker

The *Speaker's Handbook* (Sprague et al., 2013: 327) uses a salient red text box to warn its readers that "adequate practice is paramount to successful speaking". "Sitting and thinking about your speech, or reading over your outline or notes, is no substitute for rehearsing the speech aloud." (Sprague et al., 2013: 326). Barker (2011: 128), in his instructions to *Improve Your Communication Skills*, makes a similar point by raising readers' awareness for the fact that "there is a world of difference between thinking your presentation through and doing it. You may think you know what you want to say, but until you say it you don't really know. Only by uttering it aloud can you test whether you understand what you are saying. Rehearsal is the reality check" (Barker, 2011: 128).

That a charismatic presentation performance requires intensive oral training is not a simple myth. It seems to be an axiom. In addition, most empirical research asks not *if* but rather focuses on *how*, *which*, and *when* speaker feedback should be given in the context of presentation rehearsal (e.g., Batrinca et al., 2013). What is neglected are the questions of how much oral rehearsal is actually needed and if it is needed at all under certain conditions, depending on the occasion, the speaker's educational background, his or her personality traits, and previous experience with public speaking. Soorjoo (2012: 76) states: "The more you rehearse, the better you will perform". It is not that straightforward, unfortunately. For example, some experienced charismatic speakers insist that intensively practicing an oral presentation is harmful for them as it reduces their spontaneity, flexibility, naturalness and, ultimately, also their fluency, because their minds are constantly distracted by trying to remember how the current argument was paraphrased most successfully during previous rounds of rehearsal. So, either such speakers do not practice enough to overcome these problems, or their experience of public speaking and the delivery patterns that they have internalized and automated while learning to become a charismatic speaker allow them now to largely skip an intensive oral rehearsal of individual presentations.

Such reports and reflections at least cast some doubts about the two general assumptions that underlie handbook statements like those cited above: (1) everyone benefits from intensive oral preparation; (2) the more often you rehearse your presentation, the better (i.e. more charismatic) you will be in the end. The present paper also has no empirical evidence to confirm or qualify the myth of intensive preparation. However, what we can do here is to pick up on two important restrictions that are left out in connection with this myth in many rhetoric manuals.

First, intensive oral rehearsal is only effective if the speaker rehearses in front of an audience. Niebuhr and Tegtmeier (2019) conducted a series of experiments in which they let their entrepreneurship students rehearse investor pitches in different conditions, i.e. alone in a quiet room (which is in fact the prototypical rehearsal condition), in front of a real audience (of peers and friends), and in front of an audience of virtual speakers in a virtual-reality presentation training environment. They found that the prototypical rehearsal condition, i.e. presenting alone in a quiet room to no audience or only an imaginary one, fails to make speakers significantly better (according to listener ratings in a perception experiment). It requires an audience to make one's presentation performance more charismatic after repeated rehearsal and, noteworthy from a practical perspective, it makes no significant difference whether this audience is real or virtual.

Second, Niebuhr and Tegtmeier showed that too intensive rehearsal causes what they call a "speech erosion effect", i.e. a significant reduction in the charismatic performance of the presentation, which is then also carried over by speakers into the actual presentation event. The speech-erosion effect already sets in if the pitch is practiced aloud more than three times in a row. However, rehearsing in front of a real or virtual audience can attenuate the speech-erosion effect. Only a few rhetoric manuals like the *Speaker's Handbook* of Sprague et al. (2013) caution their readers against this speech-erosion effect. Sprague et al. (2013: 326) state that "some speakers rehearse their pitch so much that is becomes mechanical", and they consider this "over-preparation" a "common pitfall".

Bottom line: Is the myth that a charismatic performance requires intensive training correct? The answer to this question is yes and no. Until proven otherwise, it is reason-

able to assume that intensive rehearsal of a presentation is beneficial for the charismatic impact of a speaker in the actual speech. The data from Niebuhr and Tegtmeier (2019) provide first empirical evidence for the positive effect of rehearsal. However, rehearsing per se is not always positive. It can even make a speaker significantly less charismatic. Presentation rehearsal is beneficial only when it takes place in front of an audience and when the presentation is practiced no more than three times in direct succession. We recommend that speakers take a half day break between their rehearsal dyads or triplets.

Myth 10: Engineers are less charismatic

Anthony Fasano (2013) starts his rhetorical work *Wow the Crowd: Anthony Fasano's Guide to Public Speaking for Engineers* with the following anecdote: "Even though I have been a professional speaker for three years as of the publication of this guide, I still introduce myself as an engineer. People often joke and say, 'You can't be an engineer, engineers don't speak well in front of an audience.' This is one of the reasons that I wanted to prepare this comprehensive guide on public speaking for engineers".

That engineers are less charismatic than other speakers of the same sex and age but with a different profession and academic background is probably more a cliché than a real myth (in the sense of the introductory definition) and, in particular, not a topic that is addressed in many rhetoric manuals. Nevertheless, we decided to include this point here in our 10 myths about speaker charisma because of its societal and economic relevance. Both authors have university affiliations to technology and engineering departments. Moreover, the second author gives mandatory university courses in *Persuasive Communication and Negotiation* to business and electrical engineering students and regularly works with engineers in start-up incubators across different countries. Against this background, it is the authors' joint experience that engineers typically base their career on the mindset that good ideas, constructions, and technologies sell themselves. They would not require a persuasive person who sells them, just someone who is able to transform all facts and figures into intelligible spoken language and/or a text-loaded Power Point presentation.

We have included the present section in this paper for two reasons; first, to emphasize that good ideas, constructions, and technologies do *not* simply sell themselves. They *do* require someone who is able to push these good ideas/constructions/technologies through to investors, supervisors, and even the team who is eventually in charge of implementing them. Soorjoo (2012) explains this fact very clearly in chapter 1 of his book on how to pitch, get funded, and win clients. The second reason is that we have initial empirical evidence that engineers are indeed less charismatic than otherwise similar speakers with a different profession and academic background. As part of the second author's charisma training, a performance score is calculated for each speaker, based on a recorded (unscripted) presentation in front of a real audience (of typically 10–30 listeners); see Niebuhr et al. (2019). This performance score decomposes the speaker's speech signal into 16 acoustic parameters, assesses, for each parameter separately, how well the speaker performs and then provides a single total performance value to which each parameter contributes according to its power in triggering perceived speaker charisma. Currently, there are 466 such performance scores in the speaker database. Figure 5 shows the pat-

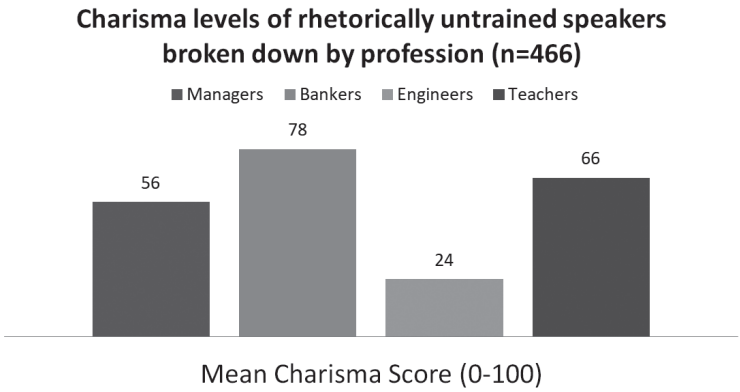

Figure 5. Rounded mean charisma scores/levels of rhetorically untrained speakers broken down by the 466 speakers' professional backgrounds, i.e. managers, bankers, engineers, and teachers.

tern that emerges, if this database is broken down by professions. The engineers' mean charisma score is significantly lower (on average 64% lower) than that of speakers with other professions (23.6). Speakers with a banking/economic background score highest (78.4), followed by teachers (66.1) and speakers with a management background (55.8).

Bottom line: Yes, engineers are less charismatic, at least in terms of initial empirical evidence from a sample of 466 advanced Western Germanic student or early-career speakers whose presentation delivery was analyzed with respect to their tone-of-voice performance. Obviously, this excludes body language, the design of presentation slides, and how the message itself is put into words; cf. the Charismatic Leadership Tactics of Antonakis et al. (2011, 2016). However, firstly, the speaker's tone-of-voice is a major factor for perceived speaker charisma and, secondly, there is no counterevidence (either anecdotal or empirical) that engineers perform better along these excluded factors than in their tone of voice. Further research is needed at these points. The poor performance of engineers in charismatic speech is potentially a loss in terms of a society's innovation and leadership (and hence economic and wealth) potential. It seems worth tackling that problem, for example, by including mandatory charisma courses into engineering education programs or by increasing the inherent motivation to take part in such courses. Such a motivation booster could be the recent finding of a significant correlation between charisma scores and course grades of engineering students (Niebuhr and Michalsky 2019b).

Discussion

In this paper we have addressed ten of the most common and often repeated myths about charisma and the way charisma manifests itself in speech. For each of the 10 myths, we have investigated its potential origin, reviewed the corresponding recommendations from the advice literature, i.e. primarily rhetoric manuals, and explained whether or not it is supported by empirical evidence. The 10 myths we investigated contribute to the demystification of charisma and the establishment of a measurable research and training

object. We have shown that the existing charisma myths are not categorically false. Several fundamental assumptions in the advice literature such as "everyone can improve his/her charismatic performance" or "charismatic performances require intensive training" are supported by modern cross-disciplinary research. However, there are many misconceptions when it comes to the phenomenological details. The advice to lower the voice supports charisma-related concepts like dominance and authority, but not charisma itself. The assumptions to increase a charismatic tone of voice through an upright posture, belly breathing and fewer/no filled pauses are all directly contradicted by empirical research. An upright posture could still be useful when it comes to a speaker's visual charisma, but it does not enhance the acoustic charisma triggers.

The upright posture is a good example of the state-of-the-art in understanding perceived speaker charisma. Two decades after the first empirical studies, we only have fragmentary knowledge about what speaker charisma actually is. There are three main reasons for this. First, the definition of charisma is still too vague. It is not just necessary to define the constituting features of charisma, as Antonakis et al. (2016) did, but also to separate charisma from related concepts. That is, we also need to state clearly at some point what charisma is not and why. Even empirical research still confuses charisma with concepts like attractiveness, dominance, assertiveness, power, likability, charm, leadership, and eloquence, all the more so in interdisciplinary studies. Semantic correlation analyses like those of Rosenberg and Hirschberg (2009) and Weninger et al. (2012) will help delimit the research object 'charisma' more thoroughly and they must be continued in the future.

Secondly, perceived speaker charisma is a complex, multifaceted phenomenon that requires an interdisciplinary approach. So far, scientific studies have barely acted on this cross-disciplinary potential. For example, concluding from the presented phonetic findings that speakers can sit while giving a speech and still be as charismatic as with a standing posture would be premature without taking into account the visual cues to charisma that, however, belong to a different research discipline. Additionally, factors like attire (Brem & Niebuhr, 2019), size and age (Niebuhr et al., 2018b), culture (Ning, 2019), and the technical properties of signal transmission (cf. Gallardo & Weiß, 2017) also play a role in charisma perception. Thus, besides linguistics and phonetics, charisma involves social sciences, political sciences, business sciences, psychology (social, personality and organizational), as well as ethnology, biology, aesthetics, media science, physics, and, last but not least, pedagogy. A broad and in-depth understanding of speaker charisma can only emerge from close collaborations between these research disciplines, preferably already at the stage of study design.

Thirdly, in order to analyze a complex multi-modal phenomenon such as speaker charisma, special measurement methods are needed; in particular those methods that allow for precise monitoring of acoustic, articulatory or cognitive processes and that are at the same time non-intrusive, adaptive and, preferably, mobile. Digital technologies ranging from virtual reality through smart phone apps and mobile EEGs to posture and gesture analyses with MS Kinect (Chen et al., 2014) have only emerged in recent years and will contribute greatly to advance charisma research in the future. The use of Respiratory Inductance Plethysmography (RIP, Włodarczak & Heldner, 2016) to investigate and debunk Myth 8 is exemplary for how technological innovations can advance charisma

understanding. The authors of this paper developed two instrumental-phonetic procedures relevant to charisma research (see Figure 6). PASCAL (Prosodic Analysis of Charismatic Speech: Assessment and Learning) breaks down a speaker's acoustic voice and melody profile into 16 charisma relevant parameters, allows to track and visualize these parameters in real time, and gives (sex-specific) feedback based on an algorithm that has been trained on perception data from hundreds of listeners from Western Germanic languages (cf. Niebuhr et al., 2019). MARRYS (Mandible Action-Related RhYthm Signals) is a special headgear for the measurement of speech rhythm. It builds on findings of Erickson and Kawahara (2016) that mandible movements (i.e. the dynamics and degrees of mouth opening) in speech are a robust correlate of perceived syllable prominence. On this basis, MARRYS will be used for future research on Myth 6.

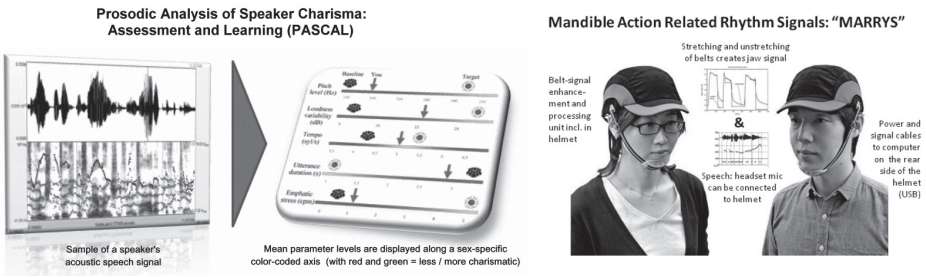

Figure 6. Illustration of two instrumental-phonetic methods PASCAL (left) and MARRYS (right) developed by the authors and their co-workers for the measurement, analysis, and assessment of charisma features.

Another technology-driven way of investigating charisma and its impact on listeners is using robots and talking machines. Works of Fischer (2018) or Niebuhr and Michalsky (2019) show that listeners interpret the acoustic charisma cues in synthetic voices in the same way as for human speakers. But, unlike for human speakers, the acoustic output of talking machines can be precisely controlled and exactly replicated. Furthermore, external factors influencing charisma such as sex, size, skin color, attire, or attractiveness are less likely to bias listener ratings of perceived charisma (if at all) if the "speaker" is a robot or a talking machine. Current experiments investigate the cross-cultural differences in the perception of charismatic voices by means of talking robots.

When it comes to practical challenges and questions for future research, revisiting the 10 myths yielded valuable insights. Although we know that a raised rather than lowered f_0 level is required in charismatic speech, Niebuhr and Skarnitzl (2019) showed that we still do not fully understand what the acoustic correlate of perceived f_0 level is. A study by Niebuhr et al. (2018a) showed additionally that also local f_0 events such as pitch accents and their f_0 shape contribute to the perceived f_0 level. The interaction between f_0, perceived pitch, and factors like speaking rate and vowel transitions is also not fully resolved (Michalsky, 2016; Barnes et al., 2012). Another unsettled issue is that of clear pronunciation. We showed that articulatory precision is important. However, when it comes to the links between articulatory precision, articulatory effort, and speaking rate, the parametric interplay in charisma perception is anything but well researched. We have also only

scratched the surface with respect to pauses, breathing, hesitations and filled pauses. For example, what about dental clicks (i.e., sucking one's teeth) as pause fillers and the prosody of filled pauses? Initial evidence suggests that dental clicks are a real "charisma killer" in that they are interpreted by listeners as a signal of self-punishment or dissatisfaction of the speaker with his/her own current performance. Another big question concerns Myth 4 and the actual relation between charismatic effects of delivery and content. Finding answers to this question also means looking in more detail at how charisma is neurologically or cognitively processed. These sciences have barely been involved in charisma research so far (see Schjoedt et al., 2011 for one of the few exceptions).

Over and above the provided research overview, the present paper touched upon several practical questions in charisma or, more generally, leadership training. We know that charisma can be learned and makes a difference, but we have little knowledge about how and why charisma works from a cognitive point of view. We do not know for how long a charismatic influence persists, whether charisma has only a short-term or a long-term effect, and even less is known about cross-cultural charisma effects. In addition, charisma probably also varies across individuals. As is described in Myth 3, little is known about how a speaker's personality affects the learnability and expression of charisma, and barely anything is known about how a listener's personality affects the charisma perception of a speaker. The same set of questions can be asked for speaker and listener sex, although a lot of research has already gone into them. In speaker training, the lack of research in all these areas bears the risk of overgeneralizing charisma instructions across individuals, languages and cultures.

We are also just beginning to grasp how charisma is best trained and learned. Antonakis et al. (2011) identified, in the form of their CLTs, key elements for a charismatic impression. Phonetic research elaborated on the CLT element named "animated voice" by identifying the specific acoustic features of a charismatic tone of voice. However, there is still much to be done. Linguistic research is inconclusive about the persuasive effect of many CLTs such as metaphors (Sopory & Dillard, 2002) and rhetorical questions (Ahluwalia & Burnkrankt, 2004), and phonetic research still has to address questions of speaker sex, culture, language, personality and many more.

Moreover, further research has to be done on the order in which charisma techniques should be trained. Learning lexical CLT strategies before addressing the speaker's "animated voice" seems to yield overall stronger improvements than in the opposite order, since some rhetorical CLTs can prime certain prosodic strategies. Additionally, male speakers should focus more on lexical and female speakers more on tone-of-voice improvements (Niebuhr & Wrzeszcz, 2019). Research on Myth 4 has also implications on how to weight the training of non-verbal and verbal strategies. How long do we have to train charisma in general to achieve the best effect in as short a time as possible, and how much training leads to sustainable improvements? Are there other training areas that indirectly foster charisma, such as the training of creativity, imagination or expressivity? Lastly, charisma is still regarded as a skill restricted to leadership and useful only to top managers, business leaders and politicians. However, what about teachers, advisors, consultants, actors or physicians? Beyond Myth 10, which professions are actually at a disadvantage when it comes to a lack of charisma?

Finally, technological advances not only provide advantages for research on speaker charisma. They can also take the assessment and learning of charisma to a new level. Measurement techniques like RIP, PASCAL, and MARRYS, and the use of speech synthesis as a learning tool all hold a big potential for training charisma, as they allow to assess charisma and identify specific weaknesses and strengths on detailed objective grounds on which personalized, effective trainings can be built. That is, measurement and visualization techniques can give feedback on a phenomenon that is otherwise largely subjective and difficult to grasp and train. They make the soft skill charisma far less soft and, thus, pave the way for a new era of leadership and public-speaking training that is shaped by science and digitization.

ACKNOWLEDGMENTS

We are greatly indebted to Radek Skarnitzl and our two anonymous reviewers for their insightful and constructive comments on an earlier draft of this manuscript. They helped us a lot to make the paper understandable and relevant across disciplines. Furthermore, thanks are due to Heike Schoormann and Pauline Welby for their careful proofreading of the revised manuscript. We would also like to thank Jana Neitsch, Stephanie Berger, Jørgen Jakob Friis, Cordula Vesper, Jana Voße, and, in fact, all participants in our charisma-training seminars for many inspiring, exciting, and sometimes challenging discussions. Additional thanks go to Lars Penke, Thomas Schultze-Gerlach, and Julia Stern as well as to the whole department of personality psychology and organizational psychology at the University of Göttingen for insightful discussions about Myth 4 and to Alexander Brem, the Chair of Technology Management of the FAU Nuremberg-Erlangen for his committed aid in investigating phonetic charisma. Finally, we would like to express our special gratitude to Dante and Ernst for their patience and continuous support and encouragement. This work was partly funded by the Danish Council for Independent Research under Grant No. 7059–00101A.

REFERENCES

Ahluwalia, R. & Burnkrankt, R. E. (2004). Answering Questions about Questions: A Persuasion Knowledge Perspective for Understanding the Effects of Rhetorical Questions. *Journal of Consumer Research*, 31, 26–42.

Amon, I. (2016). *Die Macht der Stimme*. Munich: Redline.

Antonakis, J., Bastardoz, N. & Jacquart, P. (2016). Charisma: an ill-defined and ill-measured gift. *Ann. Rev. Organ. Psychol. Organ. Behav.*, 3, 293–319.

Antonakis, J., Fenley, M. & Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Acad. Manag. Learn. Educ.*, 10, 374–396.

Antonakis, J., d'Adda, G., Weber, R. & Zehnder, C. (2015). "Just words? Just speeches?" On the economic value of charismatic leadership. In: *Working Paper. Department of Organizational Behavior*, University of Lausanne.

Awamleh, R. & Gardner, W. L. (1999). Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance. *The Leadership Quarterly*, 10, 345–373.

Barbosa, P. A. & Niebuhr, O. (submitted). Persuasive speech is a matter of acoustics and chest breathing only. *Journal of Speech Sciences*.
Barker, A. (2011). *Improve Your Communication Skills*. London: Replika Press.
Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, 3, 337–383.
Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
Bass, B. M. (1985). *Leadership and Performance Beyond Expectations*. New York: Free Press.
Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P. & Scherer, S. (2013). Cicero – towards a multimodal virtual audience platform for public speaking training. In: *Proceedings Intelligent Virtual Agents 2013*, 116–128, Edinburgh, UK.
Bell, R. L. (2011). Is your speech filled with um? 3 tips to eliminate filled pauses from your professional presentation. Retrieved from https://www.researchgate.net/publication/261551832_Is_your_speech_filled_with_um_3_tips_to_eliminate_filled_pauses_from_your_professional_presentation (last access: 13 June, 2019).
Berger, S., Niebuhr, O. & Peters, B. (2017). Winning over an audience – a perception-based analysis of prosodic features of charismatic speech. in: *Proceedings 43rd Annual Meeting of the German Acoustical Society*, 793–796.
Biadsy, F., Rosenberg, A., Carlson, R., Hirschberg, J. & Strangert, E. (2008). A cross-cultural comparison of American, Palestinian, and Swedish perception of charismatic speech. In: *Proc. Speech Prosody 2008*, 579–582. Campinas, Brazil.
Brem, A. & Niebuhr, O. (2019). Dress to Impress? On the Interaction of Attire with Prosody and Gender in the Perception of Speaker Charisma. In: M. Barkat-Defradas, B. Weiss, J. Trouvain & J. J. Ohala (Eds.), *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers*. New York: Springer Nature.
Camper Bull, R. (2010). *Moving from Project Management to Project Leadership: A practical guide to leading groups*. Boca Raton: CRC.
Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. & Zellers, M. (Eds.) (2018). *Rethinking Reduction: Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation*. Berlin: Walter de Gruyter.
Carnegie, D. & Esenwein, J. B. (2011). *The Art of Public Speaking*. London: Walking Lion.
Caspi, A., Bogler, R. & Tzuman, O. (2019). "Judging a Book by Its Cover": The Dominance of Delivery Over Content When Perceiving Charisma. *Group & Organization Management*. DOI: https://doi.org/10.1177/1059601119835982.
Chen, A., Gussenhoven, C. & Rietveld, T. (2002). Language-specific uses of the Effort Code. In: *Proc. Speech Prosody 2002*, 211–214.
Chen, L., Feng, G., Joe, J., Leong, C. W., Kitchen, C. & Lee, C. M. (2014). Towards automated assessment of public speaking skills using multimodal cues. In: *Proceedings 16th International Conference on Multimodal Interaction* (Istanbul).
Clingingsmith, D. & Shane, S. (2017). Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States. *Management Science*, 64(11), 5164–5179.
Clopper, C. G., Turnbull, R., Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., & Zellers, M. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In: Cangemi, F. et al. (Eds.), *Rethinking Reduction: Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation* (pp. 25–72). Berlin: Walter de Gruyter.
Corley, M. & Hartsuiker, R. J. (2003). Hesitation in speech can …um… help a listener understand. In: *Proc. 25th Annual Meeting of the Cognitive Science Society*.
Costa, P. T. & McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.
Curhan, J. R. & Brown, A. D. (2012). Parallel and divergent predictors of objective and subjective value in negotiation. In: K. S. Cameron & G. M. Spreitzer (Eds.), *Oxford Handbook of Positive Organizational Scholarship* (pp. 579–590). New York, NY: Oxford University Press.

D'Errico, F., Niebuhr, O. & Poggi, I. (2019). Humble voices in political communication: A speech analysis across two cultures. *Lecture Notes in Computer Science 11620*, 361–374.
Daniels, N. (2013). How to train your voice to become more charismatic. Retrieved from https://like3n-et.blogspot.com/2015/09/how-to-train-your-voice-to-be-more_14.html (last access 13 June, 2019).
Davies, J. C. (1954). Charisma in the 1952 campaign. *Am. Polit. Sci. Rev.*, 48, 1083–1102.
D'Errico, F., Signorello, R., Demolin, D. & Poggi, I. (2013). The perception of charisma from voice. A crosscultural study. *Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 552–557. Geneva, Switzerland.
Emrich, C. G., Brower, H. H., Feldman, J. M. & Garland, H. (2001). Images in words: Presidential rhetoric, charisma, and greatness. *Administrative Science Quarterly*, 46, 527–557.
Erickson, D. & Kawahara, S. (2016). Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistics Vanguard*, 2(1): https://doi.org/10.1515/lingvan-2015-0025.
Etzioni, A. (1964). *Modern Organizations*. Englewood Cliffs, NJ: Prentice Hall.
Fasano, A. (2013). Wow the Crowd: Anthony Fasano's Guide to Public Speaking for Engineers. Retrieved from https://engineeringmanagementinstitute.org/wow-crowd-engineers-guide-public-speaking/ (last access: 13 June, 2019).
Fischer, K. (2000). *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Markers*. Mouton de Gruyter.
Fischer, K. (2018). Talking to robots. In: Elmentaler, M. & Niebuhr, O. (Eds.), *An den Rändern der Sprache. Notes of a lecture series at Kiel University*. Retrieved from https://www.uni-kiel.de/ pressemeldungen/index.php?pmid=2018–084-rv-sprache&pr=1. (last access 13 June, 2019).
Fox Cabane, O. (2012). *The Charisma Myth: How Anyone Can Master the Art and Science of Personal Magnetism*. New York: Penguin.
Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, 29, 320–326.
Frese, M., Beimel, S. & Schoenborn, S. (2003). Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision. *Personnel Psychology*, 56, 671–697.
Fruehwald, J. (2016). Filled Pauses as a Sociolinguistic Variable. *U. Penn Working Papers in Linguistics*, 22, 41–49.
Gallardo, L. F. & Weiß, B. (2017). Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution. In: *Proc. Interspeech 2017*, 904–908.
Gemmill, G. & Oakley, J. (1992). Leadership: an alienating social myth? *Hum. Relations*, 45, 113–129.
Goman, K.G. (2008). *The nonverbal advantage – Secrets and science of body language at work*. San Francisco: Berrett-Koehler.
Gregory, S. W. Jr. & Gallagher, T. J. (2002). Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. *Soc. Psychol. Q.*, 65, 298–308.
Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8(2), 425–434.
Hargrave, J. (1995). *Let me see your body talk*. Dubuque: Kendall/Hunt.
Holladay, S. J. & Coombs, W. T. (1993). Communicating visions: An exploration of the role of delivery in the creation of leader charisma. *Management Communication Quarterly*, 6, 405–427.
Holladay, S. J. & Coombs, W. T. (1994). Speaking of visions and visions being spoken an exploration of the effects of content and delivery on perceptions of leader charisma. *Management Communication Quarterly*, 8, 165–189.
House, R. J. (1977). A 1976 theory of charismatic leadership. In: J. G. Hunt & L. L. Larson (Eds.), *The Cutting Edge* (pp. 189–207). Carbondale, IL: S. Ill. Univ. Press.
Howell, J. M. & Frost, P. J. (1989). A laboratory study of charismatic leadership. *Organizational Behavior and Human Decision Processes*, 43(2), 243–269.
John, O. P. & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In: L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research, Vol. 2*, (pp. 102–138). New York, NY: Guilford Press.
Judge, T. A. & Piccolo, R. F. (2004). Transformational and Transactional Leadership: A Meta-Analytic Test of Their Relative Validity. *Journal of Applied Psychology*, 89(5), 755–768.

Kraus, N. (2015). *Deutlich sprechen lernen in 7 Tagen*. Nico Kraus Verlag.
Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
Mayew, W. J., Parsons, C. A. & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior*, 34(4), 243–248.
McAleer, P., Todorov, A. & Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLoS ONE*, 9(3): e90779.
Michalsky, J. (2016). Perception of pitch scaling in rising intonation. On the relevance of f_0 median and speaking rate in German. In: *Proceedings of P&P 12*, Munich, Germany.
Michalsky, J., Kordsmeyer, T., Niebuhr, O. & Penke, L. (2019). Prosodic correlates of dominance and self-assurance. Acoustic cues to testosterone related personality states of male speakers. In: *Proc. 1st International Seminar on the Foundations of Speech*.
Mixdorff, H., Niebuhr, O. & Hönemann, A. (2018). Model-based prosodic analysis of charismatic speech. In: *Proc. Speech Prosody 2018*.
Mortensen, K. W. (2011). *The Laws of Charisma: How to Captivate, Inspire, and Influence for Maximum Success*. New York: Amacom.
Niebuhr, O., Voße, J. & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic prosodic analysis of Steve Jobs tone of voice. *Computers and Human Behavior*, 64, 366–382.
Niebuhr, O., Tegtmeier, S. & Brem A. (2017). Advancing research and practice in entrepreneurship through speech analysis – from descriptive rhetorical terms to phonetically informed acoustic charisma metrics. *Journal of Speech Sciences*, 6, 3–26.
Niebuhr, O. (2017). Clear Speech – Mere Speech? How segmental and prosodic speech reduction shape the impression that speakers create on listeners. In: *Proc. Interspeech 2017*, 894–898.
Niebuhr, O., Thumm, J. & Michalsky, J. (2018a). Shapes and timing in charismatic speech – Evidence from sounds and melodies. In: *Proc. Speech Prosody 2018*, 582–586.
Niebuhr, O., Skarnitzl, R., and Tylečková, L. (2018b). The acoustic fingerprint of a charismatic voice – Initial evidence from correlations between long-term spectral features and listener ratings. In: *Proc. Speech Prosody 2018*, 359–363.
Niebuhr, O. & Gonzalez, S. (2019). Do sound segments contribute to sounding charismatic? Evidence from acoustic vowel space analyses of Steve Jobs and Mark Zuckerberg. *International Journal of Acoustics and Vibration*, 24, 343–355.
Niebuhr, O. & Michalsky, J. (2019). Computer-generated speaker charisma and its effects on human actions in a car-navigation system experimen – Or how Steve Jobs' tone of voice can take you anywhere. *Lecture Notes in Computer Science 11620*, 375–390.
Niebuhr, O. & Fischer, K. (2019). Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. In: *Proc. Interspeech 2019*.
Niebuhr, O. & Tegtmeier, S. (2019). Virtual-reality as a digital learning tool in entrepreneurship – How virtual environments help entrepreneurs give more charismatic investor pitches. In: R. Baierl, J. Behrens & A. Brem (Eds.), *Digital Entrepreneurship: Interfaces Between Digital Technologies and Entrepreneurship*. Berlin: Springer Nature.
Niebuhr, O. & Wrzeszcz, S. (2019). A woman's gotta do what a woman's gotta do, and a man's gotta say what a man's gotta say – Sex-specific differences in the production and perception of persuasive power. In: *Proc. 13th International Pragmatics Association Conference*.
Niebuhr, O. & Skarnitzl, R. (2019). Measuring a speaker's acoustic correlates of pitch – but which? A contrastive analysis based on perceived speaker charisma. In: *Proceedings of 19th International Congress of Phonetic Sciences*.
Niebuhr, O., Tegtmeier, S. & Schweisfurth, T. (2019). Female speakers benefit more than male speakers from prosodic charisma training – A before-after analysis of 12-weeks and 4-h courses. *Frontiers in Communication*, 4, 12.
Ning, L. (2019). *What makes a speaker sound charismatic? A cross-cultural study based on acoustic-prosodic analysis*. MA thesis, Chair of Technology Management, University of Erlangen-Nuremberg, Germany.
Novák-Tót, E., Niebuhr, O. & Chen, A. (2017). A gender bias in the acoustic melodic features of charismatic speech? In: *Proc. Interspeech 2017*.

Pangambam, S. (2016). Let's Face It: Charisma Matters by John Antonakis. Retrieved from https://singjupost.com/lets-face-it-charisma-matters-by-john-antonakis-full-transcript/ (last access: 13 June, 2019).
Park, S., Shoemark, P. & Morency, L.-P. (2014). Toward crowd-sourcing micro-level behavior annotations: The challenges of interface, training, and generalization. In: *Proceedings of the 18th International Conference on Intelligent User Interfaces*.
Pentland, A. (2008). *Honest Signals – How They Shape Our World*. Cambridge: MIT Press.
Peters, J. (2015). *Charisma: How to Develop Personal Charisma and Leave that Lasting Impression on Everyone You Meet*. CreateSpace Independent Publishing Platform.
Polla, B., D'Antona, G., Bottinelli, R. & Reggiani, C. (2004). Respiratory muscle fibres: Specialisation and plasticity. *Thorax*, 59(9), 808–817.
Rosenberg, A. & Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Communication*, 51, 640–655.
Salomoni, S., van den Hoorn, W. & Hodges, P. (2016). Breathing and singing: Objective characterization of breathing patterns in classical singers. *PLoS ONE*, 11, e0155084.
Schegloff, E. A. (2010). Some other "Uh(m)"s. *Discourse Processes*, 47, 130–174.
Scherer, S., Layher, G., Kane, J., Neumann, H. & Campbell, N. (2012). An audiovisual political speech analysis incorporating eye-tracking and perception data. In: *Proc. LREC'12*.
Schjødt, H., Stodkilde-Jorgensen, A. W., Geertz, T. E. & Lund, A. (2010). The power of charisma-perceived charisma inhibits the frontal executive network of believers in intercessory prayer social cognitive and affective. *Neuroscience*, 6, 119–127.
Seet, J. (2013). Speak Clearly: Crisp Clear Pronunciation. Retrieved from http://sgskill.com/?page=course_calendar&id=1097&m=9&y=2013 (last access: 13 June, 2019).
Shamir, B. & Howell, J. M. (1999). Organizational and contextual influences on the emergence and effectiveness of charismatic leadership. *Leadership Quarterly*, 10, 257–283.
Shamir, B., Arthur, M. B. & House, R. J. (1994). The rhetoric of charismatic leadership: A theoretical extension, a case study, and implications for research. *The Leadership Quarterly*, 5, 25–42.
Sharma, S., Bottom, W. & Elfenbein, H. A. (2013). On the role of personality, cognitive ability, and emotional intelligence in predicting negotiation outcomes: A meta-analysis. *Organizational Psychology Review*, 3(4), 293–336.
Smith, R. R. (2010). Overcoming charisma. Forbes Magazine. Retrieved from https://www.forbes.com/2010/02/25/charisma-presence-communication-leadership-managing-speaking.html#4c4770716730 (last access 13 June, 2019).
Soorjoo, M. (2012). *Here's the Pitch: How to Pitch Your Business to Anyone, Get Funded, and Win Clients*. Hoboken: John Wiley & Sons.
Sopory, P. & Dillard, J. P. (2002). The persuasive effects of metaphor. A meta-analysis. *Human Communication Research*, 28, 382–419.
Speech and Voice (2019). How breathing can improve your voice. Retrieved from https://www.speechandvoice.com/speech-voice-improvement-tips/how-breathing-can-improve-voice/ (last access: 13 June, 2019).
Sprague, J., Stuart, D. & Bodary, D. (2013). *The Speaker's Handbook*. Boston: Wadsworth.
Strangert, E. & Gustafson, J. (2008). What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations. In: *Proc. Interspeech 2008*, 1688–1691.
Thorpe, C., Cala, S., Chapman, J. & Davis, P. (2001). Patterns of breath support in projection of the singing voice. *Journal of Voice*, 15, 86–104.
Touati, P. (1993). Prosodic aspects of political rhetoric. In: *Proc. ESCA Workshop on Prosody*, 168–171.
Towler A. J. (2003). Effects of charismatic influence training on attitudes, behavior, and performance. *Personnel Psychology*, 56, 363–381.
Towler, A., Arman, G., Quesnell, T. & Hoffman, L. (2014). How charismatic trainers inspire others to learn through positive affectivity. *Computers in Human Behavior*, 32, 221–228.
Tucker, R. C. (1968). The theory of charismatic leadership. *Daedalus*, 97, 731–756.
Volkmann, S. (2013). *Der kleine Stimmkompass. Lebendig sprechen – punktgenau landen. 21 Impulse für Haltung, Stimme, Körpersprache*. Silke Volkmann Verlag.
Weber, M. (1947). *The Theory of Social and Economic Organization*. New York: The Free Press of Glencoe.

Weber, M. (1968). *On Charisma and Institutional Building*. Chicago: Univ. Chicago Press.
Weninger, F., Krajewski, J., Batliner, A. & Schuller, B. (2012). The voice of leadership: Models and performances of automatic analysis in on-line speeches. *IEEE Transactions on Affective Computing*, 3, 496–508.
Włodarczak, M. & Heldner, M. (2016). Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking. In: *Proc. Interspeech 2016*, 510–514.

RESUMÉ

Charisma je složitým jevem, což se projevuje v množství mýtů, polopravd a nezodpovězených výzkumných otázek. Většina mýtů spojených s charismatem není bez kontroverze. Protože empirická zkoumání v posledních několika letech výrazně pokročila, vracejí se autoři tohoto příspěvku k deseti nejdůležitějším mýtům, které se týkají převážně, ale nikoli výhradně lingvistických a fonetických aspektů charismatu. K těm patří například interakce mezi verbálními a neverbálními jevy a mezi segmentálními a prozodickými informacemi, ale také role dýchání a základní hlasové frekvence ve vnímání charismatičnosti mluvčího. Výsledkem je značně rozmanitý obrázek. Některé z představených mýtů, včetně některých velmi starých, mohou být přijaty. Jiné je třeba ve světle odporujících empirických výsledků odmítnout. Postavení některých dalších mýtů zůstává nezodpovězeno. Při diskusích o tomto rozmanitém obrázku autoři poukazují na mezery ve výzkumu a řečové praxi a navrhují konkrétní směry, jimiž by se další výzkum měl ubírat.

Jan Michalsky
Chair of Technology Management
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
E-mail: jan.michalsky@uni-oldenburg.de

Oliver Niebuhr
Centre for Industrial Electronics
University of Southern Denmark, Sonderborg, Denmark
E-mail: olni@sdu.dk

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 57–75
PHILOLOGICA 2/ PHONETICA PRAGENSIA

THE DYNAMICS OF INDEXICAL INFORMATION IN SPEECH: CAN RECOGNIZABILITY BE CONTROLLED BY THE SPEAKER?

VOLKER DELLWO, ELISA PELLEGRINO, LEI HE, THAYABARAN KATHIRESAN

ABSTRACT

Human voices are individual and humans have elaborate skills in recognizing speakers by their voice, phenomena that are deeply rooted in the evolution of human behavior. To date, the mechanisms of speaker recognition are not well understood because of the high variability of the acoustic cues to a speaker's identity. We wondered what role the speaker plays in making his/her voice more or less well recognizable. While it is evident from the literature that humans can control vocal properties to enhance their intelligibility, it is unclear whether speakers can and/or do control vocal characteristics to be better recognizable and whether such control mechanisms play a role in the communication process. In this paper, we reviewed results from the literature supporting the view that speaker idiosyncratic information is dynamic and that humans have the ability to control how well they can be recognized. We suggest possible experimental setups by which the control over identity in voice can be tested and present pilot acoustic characteristics of speech that was produced to be either targeted at being (a) intelligible (clear speech) and (b) suitable for person recognition (identity marked speech). Results revealed that there is reason to believe that speakers apply different mechanisms when making their individuality identifiable as opposed to making their speech better understood. We discuss predictions that a control of recognizability and intelligibility has within major theories of speech perception.

Key words: indexical information, voice recognition, identity marked speech

1. Introduction

February 10th 2019: Eliza D. makes her way home through a dark subway when a masked attacker grabs her from behind and commands in a whispered, foreign-accented voice: "Give me your money, quick!". She pulls out her wallet and before she understands, the man disappears and leaves her with nothing but the memory of his voice. Months later, Eliza appears at court and identifies a suspect as her attacker based on his voice. Such scenes are common to law enforcement agencies around the world. In this

https://doi.org/10.14712/24646830.2019.18

particular case, the probability of Eliza performing a correct recognition of the suspect would be estimated as rather low, because of the short duration of the familiarization (Clifford, 1980; Kerstholt et al., 2004; Yarmey, 1995), the long time lag between familiarization and recognition (Papcun et al., 1989; Yarmey, 1995), and because the presence of a foreign accent is likely to have biased her decision (Ladefoged & Ladefoged, 1980; Stevenage et al., 2012; Yarmey, 1995). However, for Eliza this was not the first time that the recognition of an individual by his/her voice was crucial in her life, in fact, from the time she was born, there were many occasions when her survival depended on it (Kriengwatana et al., 2015; Petkov et al., 2009). She recognized her mother as a central caregiver before (Kisilevsky et al., 2003, 2009; Panneton Cooper et al., 1997) and after birth (Sullivan et al., 2011), and relied on being recognized by others to receive the right amount of attention (Locke, 2006). Eliza's remarkable voice recognition skills are an ability she shares with numerous animal species (e.g. Belin, 2006; Larranaga et al., 2015; Molnár et al., 2009; Perrodin et al., 2011, 2015). Her individual voice became part of her overall personality (e.g. McAleer et al., 2014), it supports her in building up and position herself in social groups (Schegloff, 1979), it contains information about her fertility (Fisher et al., 2011; Raj et al., 2010), it attracts the right mating partner (Bruckert et al., 2010; Collins, 2001; Collins & Missing, 2003) and contributes to the trust that others have in her (Belin et al., 2017; O'Connor & Barclay, 2017; Oleszkiewicz et al., 2017). Her voice supports listeners in paying attention to her in the environment of other voices (Johnstrude et al., 2013) and it contributes to her esthetic appearance in casual or artistic activities like singing (Doscher, 1993; Sundberg, 1977). Losing her vocal identity cues (Kurowski et al., 1996) or her ability to recognize voices (Roswandowitz et al., 2014) – for example as a result of neurological malfunction – can drive her into social isolation.

Given the significance of her voice for her social life and the consequences of a loss or change in voice identity, it is not surprising that Eliza became a frequent motive in many fictional scenarios, for example as the flower girl *Eliza Doolittle* in George Bernard Shaw's *Pygmalion* (Shaw, 1916). It is surprising, however, that theories of speech and language processing have typically treated the vocal information about her identity (henceforth: idiosyncratic cues) as information that is unwanted acoustic variability, some form of noise that needs to be cancelled out to arrive at the underlying linguistic communicative message (see discussion in Creel & Bregman, 2011). As a form of noise, idiosyncratic cues have typically been understood as static information that is given away rather involuntarily and is not under the control of the speaker. However, considering Eliza's capacity to encode an extremely rich and multidimensional amount of information in her voice, it seems implausible that she and other speakers have no control over this information. In the present article, we investigated to what degree idiosyncratic information is a by-product of the articulation process (section 2). We reviewed results from speech and speaker information processing to suggest possible control mechanisms of speakers over their idiosyncratic information (section 3), and provide reasons for why it is plausible that control mechanisms of idiosyncrasy should exist (section 4). We then provide an experimental framework and first empirical evidence revealing that idiosyncratic and linguistic information may be controlled differently, when either speaker identity or linguistic intelligibility is at stake (section 5). As a conclusion, we outline predictions that a control of idiosyncratic properties has on information processing in major theories (abstractionist and exemplar models) of speech perception (section 6).

2. How invariable is speaker idiosyncratic information?

Two types of information are typically distinguished in a speech signal: linguistic (content of the message, i.e. what is being said?) and indexical (who said what in which way? Abercrombie, 1967; Dellwo et al., 2007; Levi and Pisoni, 2007). Indexical information can be manifold. The examples about Eliza in the previous section reveal that it can serve as cues to recognize a speaker (speaker idiosyncratic information) and/or to interpret his/her situational state (speaker state information). The term 'indexical' was probably introduced by David Abercrombie (1967) to the phonetics community and goes back to the semiotic theory of Peirce (Peirce et al., 1965). In this theory, indexicality is information that specifies an object further in the context in which it occurs. For example, smoke can be indexical for the presence of a fire and, in analogy, strong de-nasalisation in speech can be indexical for the speaker suffering of a cold or a low voice can be indexical for a male gender. This usage suggests that indexical information is treated as a mere by-product of the speech production process, i.e. involuntary information without a motivated communicative intent. It consequently implies that indexical information is not controlled by the speaker.

This view might initially seem plausible for speaker idiosyncratic information, as it should support the recognition of a speaker, independent of any situational variability. Idiosyncratic information is often categorised in inborn and acquired information (Nolan, 1997), the former being a result of anatomic shapes on dimensions of the articulatory apparatus, the latter the result of acquired characteristics through exposition to particular phonetic/phonological realisations of a certain social and/or geographical environment. While acquired idiosyncrasies can to some degree be reacquired, the nature of inborn information might appear particularly static and involuntary as the anatomic dimensions of the vocal apparatus can not easily be changed and if, then only to some degree. For this reason, inborn information has been understood as a strong invariable cue to the identity of the speaker (Belin, 2006; Nolan, 1997), even though there is a general awareness that also the inborn characteristics can underlie considerable within-speaker variability. It is also well known that within-speaker variability in either inborn or acquired information, probably poses the strongest problem on most recognition scenarios. In experimental settings, this variability is referred to as 'session variability', i.e. within-speaker variability that occurs when speakers produce speech during different recording sessions between which their cues to speaker idiosyncrasy may vary naturally or as a result of environmental influences (Hansen & Hasan, 2015). Within-speaker session variability might occur from a complex interaction between speaker idiosyncratic and speaker state information (e.g. varying emotional states), it might also occur as a result of external influence (e.g. accommodation to background noise or convergence between speakers).

Within a session, little attention has been paid to the variability of idiosyncratic information. This is also true in formal speaker recognition domains. In speaker recognition technology, for example, the most recent approach – so called i-vectors or x-vectors (Dehak et al., 2011; Garcia-Romero & Espy-Wilson, 2011) – idiosyncratic information of the entire speaker is reduced to a vector of about 200 dimensions, irrespective of session variability. In forensic phonetics, a sub-field of phonetics concerned with idiosyncratic

information for the purpose of solving crime, a typical task is to decide whether a speech sample of a perpetrator (evidence) and a speech sample from a suspect (comparison) were produced by one and the same or different speakers (cf. discussion in Dellwo et al., 2018a). Also in such scenarios, the between session variability is often especially strong, as evidence and comparison recordings have typically been recorded under different speaker states and in different communicative situations (for example, shouting during a crime and relaxed telephone conversation during surveillance recording). The variability of a speaker within a session is typically not paid the same attention to.

The lack of attention to within-session idiosyncratic variability has recently been identified as some of the central problems in speaker recognition technology ("The speech signal is taken with uniformity", Sriram Ganapathy, personal communication & presentation at *Interspeech* conference 2018), thus a higher attention to selective detail within a session might enhance the recognition performance significantly. This view is supported by findings revealing that vowels and nasals are better suitable for automatic recognition compared to other consonants (Amino & Arai, 2009; Amino et al., 2009; Moez et al., 2016). Similar awareness is present in forensic speaker comparison, where vocal features such as fundamental (f_o) or formant frequency (F_1, F_2, etc.) characteristics are not seldom viewed as average statistics for a speaker in a session and are used to characterise this particular speaker (e.g. de Jong et al., 2007; Hudson et al., 2007). Here, the dynamics of formant characteristics have been pointed out to reveal a high amount of detail about the speaker at different points in time (He et al., 2019; McDougall & Nolan, 2007).

A novel view to within-person variability has been suggested by Burton et al. (2016) in the domain of facial identity cues. They argue that the acquisition of variability in a face is central to understanding how the face varies, which in return is central to the recognition process. It means, knowing more about the variability of a face helps a viewer to recognize this face under many different settings. This is highly plausible because obtaining data from a face under various different angles increases the probability that one can recognize the face under this particular position. For voices, this point of view has been taken up by Lavan et al. (2019). They argue that within-speaker variability in speech is an informative signal of individuality which means that obtaining a high amount of variability form vocalisations during an initialisation phase should support the speaker recognition process. This view is not readily in agreement with a forensic phonetic claim, which holds that variables best suitable for speaker recognition are those that offer high between-speaker variability and low within-speaker variability (Hansen & Hasan, 2015; Nolan, 1997). According to Burton et al. and Lavan et al., high within-speaker variability should provide necessary recognition information. While this approach seems highly relevant for understanding auditory speaker variability, the limitations for formal forensic scenarios are evident, as the amount of available speech data is typically too small to derive strong models about speakers' idiosyncratic cue variability. Regarding the question of the present article – whether speaker-specific information can be controlled – it seems plausible that the potential for control mechanisms is increased by an increased signal variability. Static information is typically of low entropy characterised by a low number of degrees of freedom for control. So if speakers have control over their vocal identity information, it seems plausible that they can control cues to the variability patterns of

individuality. In the following section, we discuss theoretical frameworks with which such a control might be reached.

3. Possible control mechanisms of acoustic identity cues in speech

There is strong evidence from the niche field of forensic phonetics, revealing that speakers can deliberately change individuality cues to disguise their voice with more or less success (Eriksson, 2010; Eriksson & Wretling, 1997) by a sometimes seemingly random manipulation of their cues to individuality or by imitating cues to the individuality of other speakers (De Figueiredo et al., 1996; Eriksson & Wretling, 1997; Hirson & Duckworth, 1993; Hove & Dellwo, 2014; Kitamura, 2008; Růžičková & Skarnitzl, 2017; Wagner & Köster, 1999). This means that speakers can hide information relevant to their identity and they have some intuitive consciousness about which of the inborn information (e.g. fundamental frequency) and the acquired information (e.g. regional accent) needs to be chosen for this. Speakers can also imitate or caricature other speakers' voices more or less convincingly (Jansen et al., 2001; Klewitz & Couper-Kuhlen, 1999). This requires an awareness of speakers towards the idiosyncratic characteristics that are crucial to other speakers' identity. Professional actors can typically well control their vocal identity in acting a fictional character and maintain this constantly, sometimes over a variety of characters. Good examples are the German writer and actor Marc-Uwe Kling reading from his own books and changing voices between different characters or the actress Melissa Rauch playing the character 'Bernadette' in the US TV Show 'Big Bang Theory' which is distinctly different from the actress' non-acted voice. In summary, speakers can conceal their identity and they can imitate the identity of others. This demonstrates that speakers have some control over indexical cues. But can they also control cues to make themselves more recognizable?

Idiosyncratic information is sometimes at places that have been found to be less relevant for the encoding of linguistic information as, for example, coarticulatory parts of the signal between two segments. This view, however, is problematic, as coarticulatory transitions not seldom contain important cues to parse the linguistic message and idiosyncratic information is often part of the same cues that encode linguistic meaning (e.g. formant frequencies might vary relatively between speakers which is a cue to linguistics and speaker alike). Here we argue that the cues to idiosyncrasy can most likely be found intertwined with linguistic cues (Creel & Bregman, 2011); possibly a binary distinction between the cues is not even sensible. We find that there are two phenomena that play a role in controlling idiosyncrasy, (a) the choices over segments or prosodic patterns as idiosyncratic categories as well as within-segment acoustic variability and (b) variability in speaking styles that might make more or less use of the segmental/prosodic control mechanisms.

3.1. Choices and realisation of segments

As mentioned above, vowels and nasals reveal better speaker recognition performance compared to other speech segments (Amino & Arai, 2009; Amino et al. 2009; Moez et al., 2016). It thus seems plausible that a selective choice of segmental features might support recognition. This could be reached by a selective choice of words in which segments revealing stronger idiosyncrasies occur. It is unclear, however, whether the careful and intricate planning of linguistic information would allow such choices to a considerable degree. Given that vowels contain a higher amount of speaker-specific information, a more applicable mechanism is a long clear realisation of vowels as opposed to reduction or elision. Reducing vowels to schwa or even consonants is a technique that is widely distributed throughout the world's languages, in particular in unstressed syllables. It should also be possible to change vowel qualities to contain more or less amounts of idiosyncratic information. Techniques to make voices more or less recognizable in vocalic utterances could be viewed in a very similar way as the production of clear speech that is targeted at a higher signal intelligibility (Smiljanic & Bradlow, 2008; Hazan et al., 2012). The question of whether clear-speech and production targeted at idiosyncrasy should underlie the same mechanisms is therefore discussed in section 5.

Idiosyncratic information may also be distributed differently over time. He & Dellwo (2017) showed that within-syllable temporal information leading to the syllable nucleus is less variable compared to the temporal information between nucleus and offsets. They relate this to a lesser amount of articulatory control during the final part of the syllable that may reveal more system specific movement resonances (e.g. jaw movements). Such findings could also be replicated for the temporal development of formant frequencies (He et al., 2019). It seems probable that speakers should be able to control such characteristics by enunciating syllables in a more or less controlled way. Such temporal differences should be more salient in speech that is casually produced compared to speech in which temporal properties of syllables are more controlled (e.g. infant or child directed speech).

Given the results from facial recognition (Burton et al., 2016), Dellwo et al. (2018b) investigated whether more information about the human vocal tract aids recognition. Facial variability is transmitted through the visual channel and vocal variability through the acoustic channel. If facial variability contains cues to identity in the visual signal, then vocal tract variability – in analogy – should contain cues to identity in the speech signal. Dellwo et al. tested this hypothesis by comparing vowels with a sweeping f_o to vowels with steady state f_o. The latter leads to a sweeping of all harmonics in the vocal tract. In acoustic terms, this means that any fine detail of the vocal tract transfer function is sampled over a small period of time, while a steady state f_o predominantly samples the characteristics of the transfer function at the harmonic peaks. Consequently, this means that swept f_o in vowels should contain more fine speaker-specific detail about the vocal tract anatomy. Computers and human listeners were trained in this experiment with sentence utterances. Speaker recognition performance was tested with vocalic utterances of the test speakers that were either steady-state at low level (lvlo), mid-level (lvmd) or high-level (lvhi) pitch or with a sweeping fundamental at falling (fall), falling-rising (fari) or rising (rise) pitch. Results showed that the computer model as well as human listeners performed significantly poorer in speaker recognition for vowels

with steady state compared to sweeping f_o, but the recognition performance for humans did not show such differences. The lack of an effect for humans was reasoned in a particular choice of a stimulus subset (15 speakers for computers compared to 4 speakers for humans), since humans cannot easily be tested on a large number of speakers while computers can. The analysis of the human data further revealed a high complexity as humans use multiple different time and frequency domain cues while machines rely predominantly on short term spectral information. Most importantly, humans pay strong attention to fundamental frequency which varies across low, mid and high tones and was often used as a cue to speakers' average f_o. In summary, there are plausible reasons to believe that particular realizations of vowels with more or less fundamental frequency variability contain more or less idiosyncratic speaker detail. While such effects still need to be shown for human listeners there is first evidence that computer recognition can profit from this variability.

3.2. Control of speaker-specific detail by controlling speaking styles

The control of segmental choices and the segmental realisations vary drastically with speaking styles. Some speaking styles contain more variability in f_o than others which is why it seems plausible that maintaining certain speaking styles can have positive effects on recognizability. In an experiment with charismatic speech typical for politicians, Rosenberg and Hirschberg (2005) found that recognition performance of speakers is related to voice charisma. Other research argued that distinctive voices have recognition advantages (Foulkes & Barron, 2000; Skuk & Schweinberger, 2013). The effect of the speaker's voice characteristics extends also to word recognition (Goldinger, 1996; Kraik & Kirsner, 1974). Recognition memory for words has been shown to be increased by voice congruence between study and test (Campeanu et al., 2014) which implies that producing a charismatic or distinctive voice in public speaking has certain advantages for the content of utterances to be remembered. In other words, speakers wishing to increase the probability that their identity is remembered in connection with their verbal conten – for example politicians in a debate advertising for their ideas – should maintain a stable charismatic voice. Given the findings in Dellwo et al. (2018b; cf. discussion in the previous section), speaking styles containing high degrees of fundamental frequency variability might be particularly prone to contain a large amount of idiosyncratic detail about the vocal-tract. Such a speaking style is, for example, infant directed speech (IDS) and there is first evidence that there is a recognition advantage when speaker-specific detail is acquired through IDS (Kathiresan et al., 2019). The fact that infants are often addressed in IDS might thus support their ability to acquire the mother's voice with a high amount of variability as this variability contains highly salient cues to the speaker (Burton et al., 2016; Lavan et al., 2019).

4. Why should speakers control their acoustic cues to identity in speech?

The reasons for controlling identity in speech can be manifold. One obvious reason might be when identity is at stake in a forensic investigation or when speakers intend to imitate others for artistic reasons or for identity fraud. While such situations are interesting and in need of scientific clarification, they are possibly far from being part of everyday social communicative situations. The reason for identity control is likely to be a much more integral part of voice communication. We argue that one of the prime reasons to control idiosyncrasy lies in the fact that the information about who is speaking is crucial for structuring and understanding the linguistic message in speech. The identity of the speaker also allows many assumptions about the structure and content of the utterance, which provide abundant information relevant for top-down processing. For example, a speaker using the words 'you know' very frequently will not need to pronounce these words very clearly for listeners to understand them. In dialogue processing, the absence of voice identity cues might make the dialogue ambiguous at best. The following dialogue utterances (left) might have been carried out by two speakers (middle) or by three speakers (right):

Possible dialogue utterances:	**Interpretation I:**	**Interpretation II:**
How much is this?	**Buyer A:** How much is this?	**Buyer A:** How much is this?
Let's say three dollars.	**Seller:** Let's say three dollars.	**Seller:** Let's say three dollars.
Oh, that's expensive.	**Buyer A:** Oh, that's expensive.	**Buyer A:** Oh, that's expensive.
What about two?	**Seller:** What about two	**Buyer B:** What about two?
OK, let's call it a deal!	**Buyer A:** OK, let's call it a deal!	**Seller:** OK, let's call it a deal!

Without voice processing abilities a listener could only make informed guesses about the speakers, e.g. based on the linguistic structure or cues to turn-taking. This means, the lack of speaker information makes a sensible processing of the dialogue impossible, in particular since there are several possible ways in which it could be read. In interpretation I (middle), it is most likely that A bought the item from the seller, in dialogue B it seems more plausible that B bought the item. Assuming that this dialogue was part of a radio play where speakers are not visible, listeners rely exclusively on vocal cues to identity for the correct processing.

Some circumstances make the present example very particular. In voice recognition, two major tasks are typically distinguished, first, the recognition of familiar voices and second, the discrimination of unfamiliar voices (cf. Stevenage, 2017 for a review). Both tasks might seem highly related but there is strong neurological evidence that they are separate processes (Belin & Zatorre, 2003; Latinus et al., 2011; von Kriegstein & Giraud, 2004; von Kriegstein et al., 2005) and it seems natural that the ability to discriminate should precede recognition but there is strong counter evidence (cf. discussion in Kreimann & Sidtis, 2011). The recognition of familiar voices requires previous exposure to a speaker during which other identity related features (e.g. name or face) are brought into relation with voice. In the discrimination of unfamiliar voices, the identity of the speak-

er is irrelevant, it only requires the ability to tell one voice from another. Phonagnostic listeners – i.e. listeners with impaired voice processing abilities (van Lancker et al., 1988; Roswandowitz et al., 2014) – also provide evidence for the view that voice recognition and discrimination are separate processes, as they are typically impaired in recognition but less in discrimination.

In the present example, it seems that voice recognition and discrimination are no longer easy to separate. To understand the dialogue, it is first of all essential to discriminate between voices to perceive the change in speaker. When arriving at the boundary between the third and the fourth utterance, discrimination is no longer sufficient. The listener will have to be able to remember, whether the voice from the fourth utterance is the same as the voice from the third utterance or not. This can only be solved by recognizing voices with which the listener had just been familiarized (henceforth: just-familiar voices). It requires that the listener has already created an abstract representation of the speaker the first time he/she listens to an utterance for each and every speaker in the dialogue. For the speakers – in return – it means that if they have the intention to be processed correctly in the dialogue they will have to find strategies to make themselves more recognizable, for example, by marking their voice more distinctive, i.e. use individuality cues to be better recognizable in the dialogue. While visible cues in natural dialogue situations might heavily support speaker recognition, in particular of just-familiar voices, there are numerous situations in which the visual attention of a listener is not directed towards each speaker, thus it must be assumed that audible cues play an equally important role.

The recognition of just-familiar voices will increase in difficulty with an increasing number of speakers. Interestingly, recent animal studies showed that indexical properties are related to population sizes: smaller populations have less need to distinguish themselves from each other compared to larger populations in voice recognizing animal populations (Pollard & Blumstein, 2011). While Pollard and Blumstein showed differences in idiosyncratic characteristics of unrelated populations of different animal species, such findings give rise to the idea that within populations the need for individualisation might increase with increasing numbers of participants, in particular in humans. Thus the need for individualisation in a dialogue situation as described above is even stronger with higher numbers of participants to maximise the chance that just-familiar voices can be reliably used for the processing of the dialogue. Such situations might occur in families with larger numbers of offspring and gives rise to the hypothesis that children growing up in larger families or environments with numerous peers (e.g. in an orphanage) should develop higher idiosyncratic, possibly more charismatic voices compared to children growing up individually. In analogy, children in smaller classrooms might be less idiosyncratic compared to children in larger classroom environments. Additionally, it might be that extrovert children in classrooms develop a particular amount of idiosyncrasy to make themselves more distinct and recognizable from their peers. Such situations might also occur situationally, e.g. in debates with varying numbers of speaker, in particular in the absence of visual cues. Politicians debating in a radio programme, for example, might produce voices in a particularly idiosyncratic way when they are debating with a larger number of others as opposed to being interviewed on their own or debating with one single peer.

5. An experimental design to study control mechanisms of acoustic cues to identity

Linguistic information is known to be highly dynamic. Speakers can choose to a high degree which words they use, otherwise encodings of messages would be problematic. To increase the success of linguistic information encoding, it has been demonstrated repeatedly that speakers can use mechanisms to make speech more intelligible, for example, under adverse listening conditions. This leads to a speaking style referred to as 'clear speech' (Hazan et al., 2012; Smiljanic and Bradlow, 2008). Clear speech is characterised by hyper-articulated segmental and prosodic characteristics. There is strong evidence showing that clear speech is more intelligible compared to so called conversational speech (Hazan & Baker, 2011) and that speakers can rapidly adapt their vocalisations to the particular needs of the listener (Burnham et al. 2002; Hansen & Hasan, 2015; Kemper et. al., 1998; Knoll et al. 2015). This means that speakers are aware of canonical acoustic forms that are essential to encode linguistic information and that they can control and adapt them depending on the situation. It seems to be the case that such control mechanisms could be identical to mechanisms described in (3) that make speech containing more or less idiosyncratic information. For this reason, we wanted to know whether speakers use identical mechanisms in increasing acoustic information to their identity when it is at stake (henceforth: identity marked speech) or when intelligibility is at stake (clear speech). We tested this with a mock speaker and speech recognition system. Speakers were asked to train either a speech or a voice recognition system by providing read utterances. They would hence need to test the system by reading a sentence from a screen and the system would either identify them (voice recognition) or recognize the linguistic message of the sentence on the screen (speech recognition). The system would randomly respond with an error to make the participant try to enunciate the utterance differently to obtain a higher speech or voice recognition success respectively. In the case of speech recognition, we expected typical clear speech realisations, for voice recognition it was unclear whether the identity marked speech that speakers would apply differs systematically from clear speech to make themselves better recognized.

5.1. Method

We recorded two male speakers at three different occasions. First, speakers were told that they would be recorded to train a speech technology system that we were developing. Speakers read 7 sentences into the system. Second, speakers were explained that part of the system was a speech recognizer which has problems recognizing speech correctly. Speakers would read sentences repeatedly into the system and the system would make them repeat sentences between 3 and 5 times before it would respond with the correct answer. Third, speakers were told that another part of the development was a speaker recognition system. Again, they read the sentences repeatedly until the system recognized them. The order of the experiment was balanced between the two speakers (i.e. one speaker carried out the speaker recognition first, the other vice versa). For the analysis we used the last repetition of the productions (i.e. n=42: 2 speakers * 3 styles [training,

voice recognition, speech recognition] * 7 sentences). We carried out an acoustic analysis of the speech recordings in which we obtained measures of the total utterance durations, source signal characteristics (f_o mean and standard deviation) and vocal tract resonance characteristics (long-term F_1 and F_2 as well as F_1 standard deviation). Because of the small number of tokens (n=42) we refrained from using significance tests and based the analysis on a descriptive inspection of the variables analysed.

5.2. Results and discussion

Inspection of the data (Fig. 1) showed the typical acoustic differences between clear and conversational speech (here: read speech), for example, longer total duration of the utterance (i.e. slower speech rate in terms of syllable/seconds) and higher f_o. The standard deviation of F_1 is lower in clear speech, indicating a more stable formant frequency. For identity marked speech the f_o was higher than in the read training speech but lower than in clear speech, as was the total utterance duration (Fig. 2 top left and centre). From this result, it could be concluded that clear speech was just a stronger form of identity marked speech. However, looking at f_o variability (top right), we observed that this had a tendency to be higher than both read and clear speech. In fact, f_o variability in clear speech was comparatively low compared to its high f_o mean. This again confirms a typical low variability in some prosodic variables in clear speech. Looking at average long-time formants 1 and 2 (bottom left and center), we found that F_2 was comparatively high in identity marked speech, while formant variability of F_1 (standard deviation; bottom right) was lowest of all styles. This suggests that overall the vocal tract might have been shortened in identity marked speech compared to clear and read speech, leading to higher average long-term formants. An auditory analysis of the results additionally revealed that coarticulation in identity marked speech was stronger than in clear speech, where individual sounds were better identifiable as segments and which was rhythmically more staccato-like, putting emphasis on individual vowels. Such effects are difficult to quantify acoustically but it is plausible that speakers might want to maintain their coarticulation in ID marked speech as it contains rich information about individual articulation.

The tendency in identity marked speech to have a higher f_o variability might also be related to a possible mechanism by which individuals produce a larger amount of f_o variability to increase the information about their vocal tract characteristics (see section 3.1; Dellwo et al., 2018b).

In summary, the study provides first evidence of the acoustic characteristics of clear and identity marked speech based on a novel method that directly contrasts the two speaking styles in a human-computer interaction task. Given the small amount of data obtained thus far, it is difficult to draw safe conclusions but the data supports the view that speakers apply different techniques in counter acting situations in which their identity is at stake as opposed to situations in which they are not understood. The results motivate larger systematic studies to better understand the differences. It is unclear what influence the human-computer interaction can have on the realisation of the styles and whether human-human interaction would lead to similar results. It also seems plausible to involve participants in human-human interaction, e.g. over the telephone where in one case they are not being understood and in another case not recognized. The strong

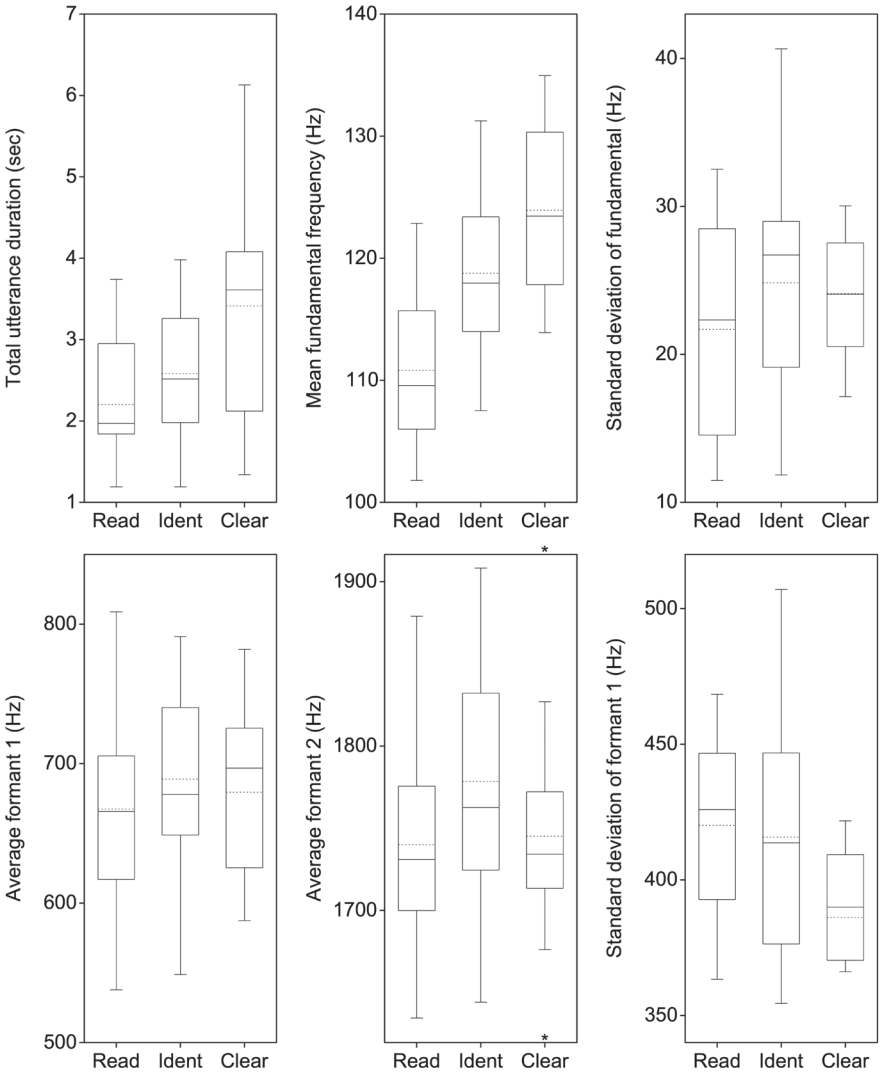

Figure 1. Distributions of six acoustic variables for each read, identity marked and clear speech: duration of utterance, fundamental frequency (f_0) mean, f_0 standard deviation, mean first formant (F_1), mean second formant (F_2) and F_1 standard deviation. All acoustic variables were calculated for each sentence utterance (n=7 in each box-plot).

advantage of the human-computer interaction is that it provides a plausible scenario in which speakers produce utterances of identical lexical content and structure by reading them. In human-human interaction, speakers would most likely have a spontaneous interaction on the telephone. This in return introduces a large amount of variability between utterances that needs to be counterbalanced by larger numbers of recordings and conversations.

6. An experimental framework for studying the dynamics of indexical information

In the previous section we saw that clear and identity marked speech are likely characterised by different acoustic features. These characteristics should help the intelligibility of the signal in the case of clear speech and they should support recognition of speakers in the case of identity marked speech. The effects of clear speech on intelligibility has been demonstrated repeatedly in the literature but the effects of identity marked speech on voice recognizability are unknown. Future research will show whether the mechanisms applied in situations in which speakers are not recognized can actually improve this situation. This could be tested by recognition experiments with humans and/or computers and the hypothesis would be that identity marked speech should lead to higher recognition rates of the speaker under conditions in which a listener has been trained on identity marked speech but possibly also under any recognition condition. Such experiments are interesting in respect to major theories of speech perception, which are probably divided by exactly the role of linguistic and indexical information in the speech signal: On one end of the scale are abstractionist theories (McClelland & Elman, 1986; Norris, 1994) which are mainly based on distinctions between language internal and external information (de Saussure, 1916: langue and parole; Chomsky, 1965: competence and performance). In these theories, indexical information is typically viewed as obstructing information (noise) that needs to be factored out of the signal to arrive at the abstract underlying linguistic forms (e.g. phonemes, words, utterances). Many phonetic theories are in line with this, viewing indexical information as a by-product of the articulation process which is an obstacle that listeners need to overcome to process the linguistic content (e.g. Fant, 1975; Liberman et al., 1967; Liberman & Mattingly, 1985). Vowels, for example, show varying formant frequencies depending on the length of the vocal tract (e.g. Peterson & Barney, 1952; Stevens, 1998) and it is argued that listeners need to normalize such speaker variability to arrive at the abstract vowel category (Adank et al., 2004). On the other end of the scale are exemplar models (Johnson, 1997) arguing that individual exemplars of speech are stored in human memory and aid linguistic processing such as recognizing phonological categories, syllables, words, etc. Thus, familiarity with a speaker's voice has a positive impact on linguistic processing which is typically measured in terms of speech processing abilities such as intelligibility. The hypothesis is that increased familiarity with a speaker leads to increased speech intelligibility (Creel & Tumlin, 2011; Nygaard et al., 1994; Theodore & Miller, 2010; Theodore et al., 2015). By now we know that there is a complex relationship between indexical and linguistic information. This relationship is also marked by studies showing that competence in a language enhances listeners' voice recognition ability (Bregman & Creel, 2014; Perrachione et al., 2015; for newborns: Fleming et al., 2014; Johnson et al., 2011; Perrachione et al., 2011). Thus, many models of speech recognition have been developed between the two poles of abstractionism and exemplarism and try to combine the advantages of each of the models (typically referred to as hybrid models, e.g. Kleinschmidt & Jaeger, 2015; see discussion in Smith, 2015).

Identity marked and clear speech allow different predictions regarding abstractionist and exemplar models of speech perception (see Fig. 2). In line with abstractionist models,

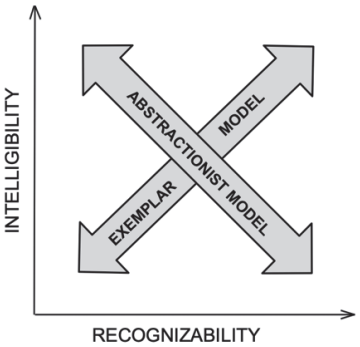

Figure 2. Predicted relationships about the intelligibility of speech and the recognizability of individuals in abstractionist and exemplar models.

it should be predicted that speakers become less recognizable with increasing intelligibility (i.e. increasing clarity) because, according to these models, speakers seem to suppress individual variability that obstructs intelligibility (Fig. 2, green arrow: negative correlation between intelligibility and recognizability). This would be in line with the findings under section 5 revealing that speakers might support two different acoustic modes for clear and identity marked speech. If the two were exclusive, it should be predicted that an increase in intelligibility automatically leads to a decrease in recognizability as speaker-specific information should be suppressed to warrant intelligibility. Given that individual variability is viewed as necessary to retrieve linguistic information from individual exemplars in exemplar models, it seems plausible that this relationship is reversed and that speakers become more recognizable with increasing intelligibility (Fig. 2, red arrow: positive correlation between intelligibility and recognizability). It will be interesting to test such predictions in the context of voice recognition experiments in future research.

ACKNOWLEDGEMENTS

Thanks to Sarah Lim for the graphical design of Fig. 2. The framework in this article will be funded by the Swiss National Science Foundation (grant number 100012_185399).

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: University of Chicago Press.

Adank, P., Smits, R. & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099–3107.

Amino, K. & Arai, T. (2009). Speaker-dependent characteristics of the nasals. *Forensic Science International*, 185, 21–28.

Amino, K., Sugawara, T. & Arai, T. (2006). Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties. *Acoustical Science and Technology*, 27, 233–235.

Belin, P. (2006). Voice processing in human and non-human primates. *Philos Trans R Soc Lond B Biol Sci.*, 361(1476), 2091–2107.

Belin, P., Boehme, B. & McAleer, P. (2017). The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS One*, 12(10), e0185651.
Belin, P. & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, 14(16), 2105–2109.
Bregman, M. R. & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, 130(1), 85–95.
Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H. et al. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116–120.
Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, 296(5572), 1435–1435.
Burton, A. M., Kramer, R. S., Ritchie, K. L. & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, 40(1), 202–223.
Campeanu, S., Craik, F. I. M. & Alain, C (2013). Voice congruency facilitates word recognition. *PLoS One*, 8(3): e58778.
Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, 4(4), 373–394.
Collins, S. A. (2001). Men's voices and women's choices. *Animal Behaviour*, 60(6), 773–780.
Collins, S. A. & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, 65(5), 997–1004.
Creel, S. C. & Bregman, M. R. (2011). How talker identity relates to language processing. *Linguistics and Language Compass*, 5(5), 190–204.
Creel, S. C. & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, 65(3), 264–285.
De Figueiredo, R. M. & de Souza Britto, H. (1996). A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, 3, 168–175.
de Saussure, F. (1916). *Cours de linguistique generale*. Laussane and Paris: Payot.
de Jong, G., McDougall, K., Hudson, T. & Nolan, F. (2007). The speaker-discriminating power of sounds undergoing historical change: A formant-based study. In: Proceedings of the 16th International Congress of Phonetic Sciences, 1813–1816.
Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P. & Ouellet, P. (2011). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4), 788–798.
Dellwo, V., Huckvale, M. & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In: Mueller, Ch. (Ed.). *Speaker Classification I: Fundamentals, Features, and Methods* (pp. 1–20). Berlin, Heidelberg: Springer.
Dellwo, V., French, P. & He, L. (2018a). Voice biometrics for forensic speaker recognition applications. In: Frühholz, S. & Belin, P. (Eds.), *The Oxford Handbook of Voice Perception* (pp. 777–798). Oxford: Oxford University Press.
Dellwo, V., Kathiresan, T., Pellegrino, E., Schwab, S. & Maurer, D. (2018b). Influences of fundamental oscillation on speaker identification in vocalic utterances by humans and computers. In: *Proceedings of Interspeech 2018*, 3795–3799.
Doscher, B. (1993). *The Functional Unity of the Singing Voice*. Scarecrow Press.
Fant, G. (1975). Non-uniform vowel normalization. *STL-QPSR*, 2–3/1975, 1–19.
Fischer, J., Semple, S., Fickenscher, G., Jürgens, R., Kruse, E., Heistermann, M. et al. (2011). Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PLoS One*, 6(9), e24490.
Fleming, D., Giordano, B. L., Caldara, R. & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38), 13795–13798.
Eriksson, A. (2010). The disguised voice: imitating accents or speech styles and impersonating individuals. In: Llamas, C. & Watt, D. (Eds.), *Language and Identities*. Edinburgh: Edinburgh University Press.
Eriksson, A. & Wretling, P. (1997). How flexible is the human voice? – A case study of mimicry. In: *Proceedings of Eurospeech 1997*, 1043–1046.

Foulkes, P. & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic Linguistics*, 7, 180–198.
Garcia-Romero, D. & Espy-Wilson, C. Y. (2011). Analysis of i-vector length normalization in speaker recognition systems. In: *Proceedings of Interspeech 2011*.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183.
Hansen, J. H. L. & Hasan, T. (2015). Speaker recognition by machines and humans: A tutorial review. *IEEE Signal Processing Magazine*, 32(6), 74–99.
Hazan, V. & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130(4), 2139–2152.
Hazan, V., Grynpas, J. & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions? *Journal of the Acoustical Society of America*, 132(5), EL371–EL377.
He, L. & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language and the Law*, 23(2), 243–273.
He, L. & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. *Journal of the Acoustical Society of America*, 141(5): EL488–EL494.
He, L., Zhang, Y. & Dellwo, V. (2019). Between-speaker variability and temporal organization of the first formant. *Journal of the Acoustical Society of America*, 145(3): EL209–EL214.
Hirson, A. & Duckworth, M. (1993). Glottal fry and voice disguise: a case study in forensic phonetics. *Journal of Biomedical Engineering*, 15(3), 193–200.
Hove, I. & Dellwo, V. (2014). The effects of voice disguise on f0 and on the formants. In: *Proceedings of IAFPA 2014*.
Hudson, T., de Jong, G., McDougall, K., Harrison, P. & Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, 1809–1812.
Jansen, W., Gregory, M. L. & Brenier, J. M. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. In: *ISCA tutorial and research workshop (ITRW) on prosody in speech recognition and understanding*.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In: Johnson, K. & Mullennix, J. W. (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.
Johnson, E. K., Westrek, E., Nazzi, T. & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011.
Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P. & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, 24(10), 1995–2004.
Kathiresan, T., Dilley, L., Townsend, S., Shi, R., Daum, M., Arjmandi, M. & Dellwo, V. (2019). Infant-directed speech enhances recognizability of individual mothers' voices. *Journal of the Acoustical Society of America*, 145(3), 1766.
Kemper, S., Finter-Urczyk, A., Ferrell, P., Harden, T. & Billington, C. (1998). Using elderspeak with older adults. *Discourse Processes*, 25(1), 55–73.
Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. J. & Broeders A. P. A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology*, 18(3), 327–336.
Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H., Zhang, K., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, 14(3), 220–224.
Kisilevsky, B., Hains, S., Brown, C., Lee, C., Cowperthwaite, B. & Stutzman, S. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, 32, 59–71.
Kitamura, T. (2008). Acoustic analysis of imitated voice produced by a professional impersonator. In: *Proceedings of Interspeech 2008*, 813–816.
Kleinschmidt, D. F. & Florian Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203.

Klewitz, G. & Couper-Kuhlen, E. (1999). Quote-unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics*, 9(4), 459–485.
Knoll, M. A., Johnstone, M. & Blakely, C. (2015). Can you hear me? Acoustic modifications in speech directed to foreigners and hearing-impaired people. In: *Proceedings of Interspeech 2015*, 2987–2990.
Craik, F. I. M. & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 26(2), 274–284.
Kriengwatana, B., Escudero, P. & ten Cate, C. (2015). Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Frontiers in Psychology*, 5, 1543.
Kreiman, J. & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Hoboken: John Wiley & Sons.
Kurowski, K. M., Blumstein, S. E. & Alexander, M. (1996). The Foreign Accent Syndrome: A reconsideration. *Brain and Language*, 54(1), 1–25.
Ladefoged, P. & Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics*, 49, 43–89.
Larranaga, A., Bielza, C., Pongrácz, P., Faragó, T., Bálint, A. & Larranaga, P. (2015). Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs from barking. *Animal Cognition*, 18(2), 405–421.
Latinus, M. & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2, 1–12.
Lavan, N., Burton, M., Scott, S. K. & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*, 26(1), 90–102.
Levi, S. V. & Pisoni, D. B. (2007). Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes. In: T. M. Lowrey (Ed.), *Psycholinguistics Phenomena in Marketing Communications* (pp. 203–219). Mahwah: Lawrence Erlbaum.
Liberman, A. M., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
Locke, J. L. (2006). Parental selection of vocal behavior: Crying, cooing, babbling, and the evolution of language. *Human Nature*, 17(2), 155–168.
McAleer, P., Todorov, A. & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS One*, 9(3): e90779.
McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
McDougall, K. & Nolan, F. (2007). Discrimination of speakers using the formant dynamics of /u:/ in British English. In: Proceedings of the 16th International Congress of Phonetic Sciences, 1825–1828.
Moez, A., Bonastre, J. F., Kheder, W. B., Rossato, S. & Kahn, J. (2016). Phonetic content impact on forensic voice comparison. In: *IEEE Spoken Language Technology Workshop (SLT)*.
Molnár, C., Pongrácz, P., Faragó, T., Dóka, A. & Miklósi, Á. (2009). Dogs discriminate between barks: the effect of context and identity of the caller. *Behavioural Processes*, 82(2), 198–201.
Nolan, F. (1997). Speaker recognition and forensic phonetics. In: W. Hardcastle & J. Laver (Eds.), *A Handbook of Phonetic Science*. Oxford: Blackwell.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234.
Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46.
O'Connor, J. J. & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, 38(4), 506–512.
Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K. & Sorokowska, A. (2017). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin & Review*, 24(3), 856–862.
Panneton Cooper, R., Abraham, J., Berman, S. & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, 20(4), 477–488.

Papcun, G., Kreiman, J. & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America*, 85(2), 913–925.
Peirce, C. S., Hartshorne, C., Weiss, P. & Burks, A. W. (1965). *Collected papers of Charles Sanders Peirce*. Cambridge, Mass: Belknap.
Perrachione, T. K., Del Tufo, S. N. & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, 333(July), 595.
Perrachione, T. K., Dougherty, S. C., McLaughlin, D. E. & Lember, R. A. (2015). The effects of speech perception and speech comprehension on talker identification. In: *Proceedings of ICPhS 2015*.
Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Current Biology*, 21(16), 1408–1415.
Perrodin, C., Kayser, C., Abel, T. J., Logothetis, N. K. & Petkov, C. I. (2015). Who is that? Brain networks and mechanisms for identifying individuals. *Trends in Cognitive Sciences*, 19(12), 783–796.
Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184.
Petkov, C. I., Logothetis, N. K. & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *The Neuroscientist*, 15(5), 419–429.
Pollard, K. A. & Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, 21(5), 413–417.
Raj, A., Gupta, B., Chowdhury, A. & Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *Journal of Voice*, 24(3), 363–368.
Rosenberg, A. & Hirschberg, J. (2005). *Acoustic/Prosodic and lexical correlates of charismatic speech*. Columbia University: Academic Commons.
Roswandowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S. & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology*, 24(19), 2348–2353.
Růžičková, A. & Skarnitzl, R. (2017). Voice disguise strategies in Czech male speakers. *Acta Universitatis Carolinae – Philologica 3*, 19–34.
Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In: Psathas, G. (Ed.), *Everyday Language: Studies in Ethnomethodology* (pp. 23–78). New York: Irvington Publishers.
Shaw, G. B. (1916). *Pygmalion*. New York: Brentano.
Skuk, V. G. & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, 296, 131–140.
Smiljanić, R. & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *Journal of the Acoustical Society of America*, 124(5), 3171–3182.
Smith, R. (2015). Perception of speaker-specific phonetic detail. In: Fuchs, S., Pape, D., Petrone, C. & Perrier, P (Eds.), *Individual Differences in Speech Production and Perception* (pp. 11–38). Frankfurt a. M.: Peter Lang.
Stevenage, S. V., Clarke, G. & McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology*, 24(6), 647–653.
Stevenage, S. V. (2017). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia*, 31(116), 162–178.
Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
Sullivan, R., Perry, R., Sloan, A., Kleinhaus, K. & Burtchen, N. (2011). Infant bonding and attachment to the caregiver: Insights from basic and clinical science. *Clinics in Perinatology*, 38, 643–655.
Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, 236(3), 82–91.
Theodore, R. M. & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *Journal of the Acoustical Society of America*, 128(4), 2090–2099.
Theodore, R. M., Blumstein, S. E. & Luthra, S. (2015). Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, Perception, and Psychophysics*, 77(5), 1674–1684.
Van Lancker, D. R., Cummings, J. L., Kreiman, J. & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, 24(2), 195–209.
Von Kriegstein, K & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, 22(2), 948–955.

Von Kriegstein, K., Kleinschmidt, A. & Giraud, A. L. (2005). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex*, 16(9), 1314–1322.
Wagner, I. & Köster, O. (1999). Perceptual recognition of familiar voices using falsetto as a type of voice disguise. In: Proceedings of the 14th International Congress of Phonetic Sciences, 1381–1385.
Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, 1(4), 792–816.

RESUMÉ

Každý člověk má jiný hlas a lidé mají propracované schopnosti, jak mluvčí rozpoznávat po hlasu. Jedná se o jev, který je hluboce zakořeněn ve vývoji lidského chování. Mechanismy rozpoznávání mluvčích dodnes jsou dobře pochopeny, a to především kvůli vysoké míře variability akustických vodítek k individualitě mluvčího. Příspěvek se zaměřuje na otázku, jakou roli hraje mluvčí, když svůj hlas dělá více či méně rozpoznatelný. Zatímco je z literatury evidentní, že mluvčí jsou schopni ovládat vlastnosti svého hlasu za účelem snadnější srozumitelnosti, není zřejmé, jestli jsou mluvčí schopni tyto vlastnosti ovládat za účelem snadnější rozpoznatelnosti a jestli to opravdu dělají. Otázkou také je, jestli takové ovládací mechanismy hrají nějakou roli v komunikačním procesu. Článek shrnuje výsledky dosavadních studií, které podporují názor, že idiosynkratické mluvčí jsou dynamické povahy a že lidé dovedou ovládat, nakolik bude jejich hlas rozpoznatelný. Autoři naznačují možné podoby experimentálního výzkumu, které by umožnily ovládání hlasové identity ověřit, a představují pilotní studii akustických vlastností řeči, která byla produkována s cílem být (a) srozumitelná (zřetelná řeč) nebo (b) vhodná pro rozpoznání mluvčího (řeč obsahující vodítka k identitě). Výsledky podporují názor, že když mluvčí chtějí, aby byla jejich identita dobře rozpoznatelná, využívají odlišných mechanismů ve srovnání se situací, kdy chtějí, aby jejich řeči bylo dobře rozumět. Autoři diskutují předpovědi, které z ovládání rozpoznatelnosti a srozumitelnosti vyplývají v rámci nejvýznamnějších teorií percepce řeči.

Volker Dellwo, Elisa Pellegrino, Lei He, Thayabaran Kathiresan
Phonetics & Speech Sciences Group
Institute of Computational Linguistics
University of Zurich, Switzerland
E-mail: volker.dellwo@uzh.ch

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 77–95
PHILOLOGICA 2/ PHONETICA PRAGENSIA

SPECTRAL AND TEMPORAL CHARACTERISTICS OF CZECH VOWELS IN SPONTANEOUS SPEECH

NIKOLA PAILLEREAU and KATEŘINA CHLÁDKOVÁ

ABSTRACT

This paper provides a comprehensive account of spectral and durational characteristics of Czech monophthongal vowels. It improves on the existing literature (that almost exclusively focused on read speech) in that it examines vowels in spontaneous speech recorded from 10 men and 10 women, who were recruited from the general population not restricted to students or media reporters (which were the populations used in previous studies). The present material thus represents a relatively naturalistic data set. The acoustical analyses of vowel spectral properties are not limited to only the first and the second formant (F1 and F2) but include also higher formants. Duration normalized for word length as well as long/short duration ratios are compared across all vowel qualities. In line with previous acoustic data on Czech high front vowels, the present results confirm that the phonologically short /ɪ/ is realized with a higher F1 than the phonologically long /iː/. The results further demonstrate that the mid front /ɛ/ and /ɛː/ are realized with a relatively high F1 and are numerically even closer to the low /a/ and /aː/ than to the other mid vowel quality, the back /o/ and /oː/. A novel finding is that short back vowels /o/ and /u/ have a higher F2 than their long counterparts: this slight fronting is likely attributable to the spontaneous style of speech as well as to the mostly coronal context in which the vowels were embedded. In contrary to recent literature that reported extremely low long/short ratios in high vowels our findings show that duration marks the phonological length distinctions consistently across all five vowel pairs: long vowels are on average 1.76 times longer than short vowels. The study concludes with a discussion of the implications that the vowel acoustic properties may have on the way the Czech vocalic system is transcribed.

Key words: vowels, Czech, vowel formants, vowel duration, spontaneous speech, phonological transcription

1. Introduction

Each of the world's languages contrasts its vowels by their spectral quality, that is, by a set of frequency components called formants which are the resonant frequencies of the vocal tract (Fant 1960). Vowels are typically described in terms of the first and the second formant (F1 and F2), the former being roughly correlated to the vertical position of the

https://doi.org/10.14712/24646830.2019.19

tongue and to jaw opening and the latter roughly to the horizontal position of the tongue and to lip settings (Crothers, 1978).

Although vowel descriptions commonly refer to a two-dimensional space with F1 plotted on the vertical and F2 on the horizontal axis, studies show that at least in some languages higher formants, especially the third and the fourth formant (F3 and F4), may serve as a main cue to vowel identity. F1 and F2 suffice to describe vowels whose dominant energies are located below 1000 Hz and whose higher formant frequencies are consequently weakened and become perceptually non-salient; these are typically back vowels such as /u/ or /o/ (Vaissière, 2011). However, when the energy is concentrated in higher frequencies, F3 and F4 can come to play a major role: this is especially the case of languages contrasting front rounded and unrounded vowels, such as French and Swedish, where F3 is roughly corelated to labiality (Fant, 1969; Vaissière, 2009). Higher formants alone might even differentiate vowel contrasts that had been traditionally understood as F1-based: in that respect, some native speakers of French do not distinguish their native /i/ and /e/ in terms of F1 and F2 but instead in terms of F3 and F4 (Kamiyama, 2011). Moreover, higher formants are pertinent in a cross-linguistic comparison of vowel spectra: while the acoustic target of French /i/ is to make F3 as high as possible such that it comes close to F4, thus making the F3/F4 zone perceptually most salient, the acoustic target of English /i/ is to make F2 and F3 come close together (Gendrot et al., 2008). In most languages, the realization of the phoneme /i/ indeed aims at maximal F2, but the "French" F3-F4 pattern is not uncommon and has been observed also in some speakers of English (Flemming, 2019). Since higher formants such as F3 and F4, and the distance between them, have been shown to cue vowel identity in at least some languages, it is desirable to include these higher formants in acoustic description of front vowels cross-linguistically.

1.1. The Czech vowel system

Czech vowel phonemes are distinguished by their spectral properties and by their duration. Czech has been described as contrasting 5 monophthongal vowel qualities, namely, [i]-like, [e]-like, [a]-like, [o]-like, and [u]-like, each of which occurs as short and long. To capture both the spectral and durational properties of the Czech vowels, the 10 monophthongal phonemes are by many recent authors transcribed as /iː ɪ ɛː ɛ aː a oː o uː u/ (Dankovičová, 1997; Podlipský et al., 2009; Chládková et al., 2009; Šimáčková et al., 2012; Paillereau, 2016; Skarnitzl et al., 2016; Chládková et al., 2019).

Although the monophthongal vowel inventory of Czech is symmetrical phonologically by differentiating the high front /iː ɪ/ from the high back /uː u/, and the mid front /ɛː ɛ/ from the mid back /oː o/, phonetically the mid front vowels are consistently realized with much higher F1 values than the mid back vowels (Skarnitzl & Volín, 2012; Šimáčková et al., 2012; Paillereau, 2016; Chládková et al., 2019). Besides the phonetic 'lowness' of the front mid vowel, what is perhaps the most intriguing feature of the Czech vowel system is the realization of vowel quantity contrasts.

The short-long phoneme distinction within each of the five phonological vowel qualities has been typically realized primarily by duration (Chlumský, 1928). Yet, the phonological length contrast within the high front vowel pair is consistently realized through

spectral properties, with the short member having a higher F1 (and a lower F2) than the long one, as captured in the commonly employed transcription /i:/ versus /ɪ/. The spectral distinction in the high front short-long vowel pair was observed already by the early Czech phoneticians (e.g. Frinta, 1909; Hála, 1962) who, however, did not consider it significant enough to be captured in the transcription (Frinta, 1925). The spectral differentiation of /i:/-/ɪ/ has been objectively confirmed by a number of recent acoustic measurements (Skarnitzl & Volín, 2012; Šimáčková et al., 2012; Paillereau, 2016; Chládková et al., 2019). Spectral differentiation of a phonological length contrast, comparable to that attested in /i:/-/ɪ/, has not (yet) been found for the high back vowels, although some note a potential trend in that respect (either explicitly as Skarnitzl & Volín, 2012, or implicitly by transcribing the vowels as /u:/ /ʊ/ in Duběda, 2005). Czechs not only realize the phonological length contrast between /i:/ and /ɪ/ through spectral differences when speaking but they also rely on spectral cues when listening. Two recent speech perception experiments report a strongly spectrally-guided perceptual differentiation of the long-short /i:/ /ɪ/ contrast and, at the same time, show that the extent to which spectrum cues the long-short contrast in the high back /u:/ /ʊ/ is smaller (Podlipský et al., in press; Paillereau & Skarnitzl, 2019).

About a century ago, (stressed) phonologically long vowels were measured as being twice as long as the (stressed) short ones (Chlumský, 1928). Only ten years ago, then, an analysis of vowels produced by 6 speakers reported strikingly smaller durational ratios, especially for the high front and high back vowel pairs: the long phoneme being only 1.3 times longer than the short one for the high front vowels (originally reported in Podlipský et al., 2009, subsequently referred to in Skarnitzl, 2012; Skarnitzl & Volín, 2012; Skarnitzl et al., 2016). The comparison of the early 1928 and the later 2009 measurement might seem to indicate a diachronic trend whereby the declining durational difference come to be supplemented, or perhaps even overtaken, by a more pronounced spectral difference in order to maintain the contrast (see also a similar proposal by Šimáčková et al., 2012). This proposal remains a speculation, partially due to the limited number of speakers in the 2009-sample and the difference in speech style between Chlumský's study of spontaneous speech and the Podlipský's et al. study of read speech.

1.2. Aims of the present study

The aim of our study is to provide a thorough acoustic analysis of Czech monophthongal vowels from spontaneous speech. Spontaneous production may better represent natural speech realization than recordings of read material, the latter being the focus of most recent studies. Our population are non-students, which is another improvement on previous studies that recruited students or professional media presenters (both of which are rather specific populations unlikely representative of the average speaker of Czech).

Vowels are analysed here in terms of vowel formants and duration. Our objectives are as follows. Firstly, we assess and compare the spectral F1 and F2 properties of all 10 monophthongs to show whether and to what extent short-long contrasts are differentiated by spectrum (being specifically interested in the spectral distinction within high front and high back vowels), and whether the F1 of front mid vowels is more close to

that of the back mid vowels or to that of the low vowel /a/. Secondly, we aim to find out whether, in spontaneous speech, durational ratios of long to short vowels are comparable to those reported for read speech in Podlipský et al. (2009). The ratios in spontaneous speech could be smaller, which would indicate that the importance of duration in Czech speech is indeed declining (in line with what the divergent results between old and new studies suggest). On the contrary, the long/short ratios could as well be larger than previously reported which would indicate that in spontaneous speech (in which vowel spectral qualities are in general reduced as compared to read, careful speech) duration reliably cues vowel distinctions. Thirdly, we analyse and report the F3 and F4 and test whether the psychoacoustic distances between the higher formants help differentiate amongst the four front vowels (which is what has been found in e.g. French). Finally, in relation to the vowels' acoustic characteristics, we discuss the IPA symbols that had been and could be used in the phonemic transcription of Czech vowels.

2. Method

2.1. Speakers

Ten male and ten female speakers who have been living in the Prague region for at least 5 years and who did not have any noticeable regional accent were recruited for the purpose of the study. Male speakers were aged between 27 and 48 years (mean = 34.6, s.d. = 5) and female speakers between 25 and 34 years (mean = 29.6, s.d. = 2.1). They were healthy individuals with no hearing or speech impairments and were paid for their participation.

2.2. Recording procedure

Speakers were instructed to spontaneously comment on 20 objects that were given at their disposal. The 20 objects had been carefully chosen so that their names would contain all Czech monophthongal vowels /ɪ iː ɛ ɛː a aː o oː u uː/ in a word-initial, i.e. stressed, syllable. The vowels were embedded in a controlled consonantal context (as far as this was possible with object names): preceding consonants were mainly bilabials and following consonants mainly alveolars. The speakers were instructed to mention the name of each object at least twice when talking about it. To ensure that the objects would be named consistently across participants, and in a non-diminutive form (which would alter the number of syllables in a word), all the objects had a sticker with their name written on it. The production task was mainly a monologue but when speakers were running out of ideas, the experimenter engaged in a conversation about the objects. The 20 words from which vowels were segmented and analysed are listed in Table 1. Recordings were made in a sound-treated booth using a head-mounted condenser microphone AKG C520 and an Edirol UA 25 sound card connected to a PC running the Audacity software (version 2.3.0. retrieved from http://audacity.sourceforge.net). The material was digitized at a 44.1-kHz sampling frequency and 16-bit quantization.

Table 1. Words (orthographic, phonemic, translation) containing the target vowels

vowel	words orthographic	words phonemic	English translation
ɪ	miska, pytel	/mɪska/, /pɪtɛl/	bowl, sack
i:	víla, sítko	/vi:la/, /si:tko/	fairy, sieve
ɛ	meloun, metla	/mɛloun/, /mɛtla/	melon, whisk
ɛ:	léto, pérko	/lɛ:to/, /pɛ:rko/	summer, (bail) spring
a	balón, maska	/balo:n/, /maska/	ball, mask
a:	šátek, páska	/ʃa:tɛk/, /pa:ska/	scarf, tape
o	bota, kostka	/bota/, /kostka/	shoe, cube
o:	tóny, glóbus	/to:nɪ/, /glo:bus/	tones, globe
u	dudlík, husa	/dudli:k/, /husa/	pacifier, goose
u:	hůlka, kůra	/hu:lka/, /ku:ra/	wand, crust

2.3. Acoustical analyses

Word and vowel onsets and offsets were marked and labelled using Praat (Boersma and Weenink, 2018). A vowel token was included in the analysis if the target word form did not change in the number of syllables (suffix alternations not resulting in syllable-count change were accepted), and if the word was not mispronounced. Word onsets and offsets were marked as the onsets and offsets of the first and last segment, respectively, aligned to zero crossings of the waveform. Vowel onsets and offsets were marked on the basis of both the spectrogram and the waveform: the vowel interval had to contain visible energy in a broad-band spectrogram and visible formants (especially F2), and its first and the last waveform-period had to have a similar shape as the token's medial periods.

Vowel formants were measured by the optimized ceiling method (Escudero et al., 2009; Chládková et al., 2011) which searched for such a formant ceiling that yielded minimal variation in the measured F1, F2, and F3 values, per vowel category and per speaker. With the optimal ceiling settings, values of the first four formants were measured over the entire vowel portion with a Gaussian-like window centered at vowel midpoint, using the Burg algorithm implemented in Praat (Boersma and Weenink, 2018). Tokens for which the analysis yielded unlikely values (e.g., /a/-tokens measured with /u/-like low F1 and low F2 values) were reanalysed manually. The final set contained 1386 vowel tokens (133 occurrences of /ɪ/, 153 of /i:/, 130 of /ɛ/, 143 of /ɛ:/, 136 of /a/, 149 of /a:/, 152 of /o/, 119 of /o:/, 135 of /u/, 136 of /u:/), of which 692 were uttered by women and 694 by men.

2.4. Statistical analyses

Formant values measured in Hz were transformed to ERB using the Praat *hertzToErb()* function that implements the formula:

$$y = 11.17 \ln\left(\frac{x + 312}{x + 14680}\right) + 43$$

where x is the formant value in Hertz.

Vowel duration measured in ms was normalized for total word duration using the formula:

$$y = \mathrm{a}\,\frac{x_V}{x_W}$$

where x_V is a token's vowel duration in seconds, x_W is the same token's word duration in seconds, and $a = 0.5$ which is the rounded word duration average across all 1386 words in the data set.

Statistical analyses were performed in R (R Core team, 2008), using packages *lmerTest* (Kuznetsova et al., 2017) and *emmeans* (Lenth et al., 2018). The ERB-transformed F1 and F2 and the normalized duration were each submitted to a linear mixed-effects model with vowel length, vowel quality, and sex as fixed factors with orthogonal contrasts that were specified uniquely in each of the three models as follows. For F1 we tested *i* vs. *u*, *i* vs. *o*, *e* vs. *a*, and *e* vs. *o*; for F2 we tested *e* vs. *i*, *o* vs. *u*, *a* vs. *e*, and *a* vs. *o*; for duration we compared each of *i*, *a*, *o*, and *u* to *e* as the reference category (note that here and in the following sections, we use vowel orthographic symbols in italics to denote one of the five phonological vowel qualities collapsing across the short-long phonemes of that vowel quality). Speaker was entered as a random factor with per-vowel quality and per-vowel length random slopes.

Another two mixed-effects models were run to test the higher-formant characteristics of the four front vowels: one for the F3-F2 difference and one for the F4-F3 difference (in ERB). Vowel and sex were fixed factors (with the following orthogonal contrasts for vowel /iː/ vs. /ɪ/, /iː/ vs. /ɛː/, and /ɪ/ vs. /ɛ/), including speaker as a random factor. A last model was run to test long/short duration ratios across the five vowel qualities. Long/short ratios were computed separately for each vowel quality per speaker from the normalized duration values. Sex and vowel quality were fixed factors, testing the following 4 orthogonal contrasts: *a* vs *iu, a* vs *eo, i* vs *u,* and *e* vs *o*; speaker was entered as a random factor.

3. Results

Figure 1 shows the 10 Czech monophthongs in an ERB-scaled F1-F2 space separately for women and men. Figure 2 visualizes the vowels' spectral characteristics from F1 through F4, pooled across sexes. Table 2 then lists F1 and F2 values in Hertz, and Table 3 gives the front vowels' F3 and F4 values, and their psychoacoustic distances from F2 and F3, respectively. Table 4 shows raw and normalized vowel durations and the long/short ratios.

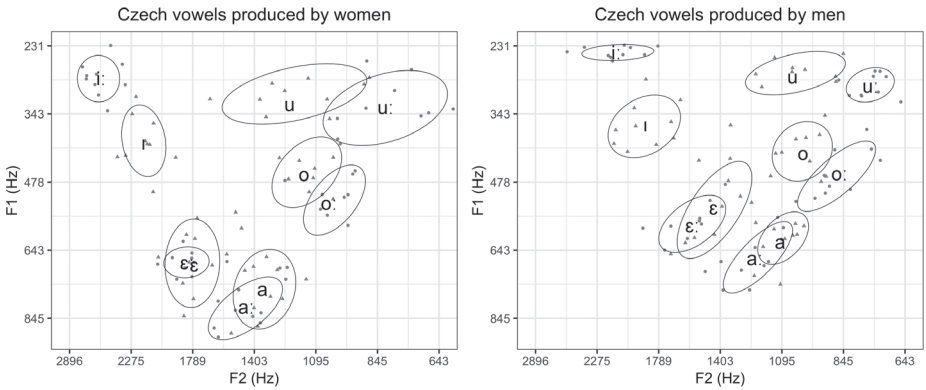

Figure 1. Czech vowels plotted in the F1-F2 plane, symbols show averages over 10 speakers per sex, ellipses cover one standard deviation from the mean, i.e. 68 % of the data. Grey symbols depict per-speaker median values (on which the group-averaging was applied); circles = long vowels, triangles = short vowels. The axes are scaled in ERB (with a 2-ERB distance between neighbouring axis labels), values are shown in Hz.

Table 2. First- and second-formant values (in Hz) and 95% confidence intervals (estimated with the *emmeans* R package) of the 10 Czech monophthongs, for 10 women and 10 men.

vowel	women				men			
	F1		F2		F1		F2	
	mean	95% c.i.	mean	95% c.i.	mean	95% c.i.	mean	95% c.i.
i:	287	268–308	2504	2382–2632	235	217–253	2157	2052–2267
ɪ	411	386–438	2177	2064–2296	347	324–371	1876	1778–1978
ɛ:	671	636–707	1825	1731–1924	583	552–616	1571	1490–1657
ɛ	650	613–688	1726	1630–1827	564	531–598	1485	1402–1573
a:	784	745–825	1436	1362–1513	685	650–722	1232	1168–1300
a	733	694–773	1322	1249–1399	639	604–675	1133	1069–1200
o:	529	497–563	1024	966–1085	455	426–485	872	821–925
o	474	442–507	1161	1095–1230	404	376–434	992	934–1052
u:	341	319–364	851	787–919	283	263–304	720	663–780
u	330	307–355	1221	1134–1313	274	252–296	1044	969–1125

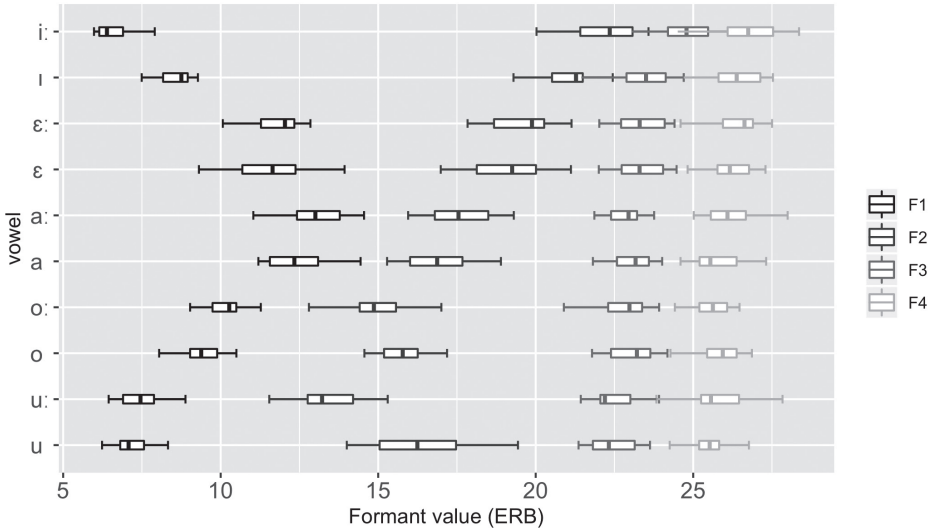

Figure 2. The F1–F4 values of the 10 Czech monophthongs, pooled across sexes. Boxes range from the 25% to the 75% percentile, vertical lines mark the 50% percentile, whiskers represent 1.5 times the interquartile range shown by the boxes.

Table 3. Third- and fourth-formant values (in Hz and ERB) of the 10 Czech monophthongs, the difference F3-F2 and F4-F3 (in ERB), averaged across sexes.

vowel	F3				F3-F2 (ERB)	
	Hz		ERB			
	mean	95% c.i.	mean	95% c.i.	mean	95% c.i.
i:	3202	3105–3301	24.82	24.57–25.08	2.65	2.37–2.92
ɪ	2688	2608–2770	23.38	23.13–23.63	2.37	2.10–2.65
ɛ:	2668	2588–2750	23.32	23.07–23.57	3.78	3.50–4.05
ɛ	2675	2595–2757	23.34	23.09–23.59	4.26	3.99–4.54
	F4				F4-F3 (ERB)	
	Hz		ERB			
	mean	95% c.i.	mean	95% c.i.	mean	95% c.i.
i:	4058	3940–4181	26.74	26.50–26.97	1.91	1.69–2.14
ɪ	3887	3775–4004	26.40	26.16–26.63	3.01	2.79–3.23
ɛ:	3893	3780–4009	26.41	26.17–26.64	3.08	2.86–3.31
ɛ	3785	3676–3898	26.18	25.95–26.42	2.84	2.62–3.06

3.1. Vowels formants

For F1, we found a main effect of sex confirming the anatomically conditioned sex difference in vowels having a larger F1 in women than in men (by on average 1 ERB, $t[22.6] = 6.277$, $p = 2 \times 10^{-6}$). As for the vowel quality contrasts, all of our 4 comparisons of vowel qualities turned out significant implying that *i* has an overall larger F1 than *u* and smaller F1 than *o* (by 4.9 and 9.4 ERB, respectively), and that *e* has a larger F1 than *o* and a smaller F1 than *a* (by 9.4 and 5.7 ERB, respectively; all *p*s < .001). Importantly, significant interactions with vowel length showed that the vowel quality comparisons are differentially modulated by vowel length: the *i* vs. *o* difference is 4 times larger for the long vowels than for the short vowels, being 3.6 and 0.9 ERB respectively, and the *i* vs. *u* difference is in different directions for short than for long vowels, short /ɪ/ having larger F1 than short /u/ by 1.3 ERB and long /i:/ having smaller F1 than long /u:/ by 0.9 ERB. As for short-long comparisons within each vowel quality, the estimated means and confidence intervals (see also Figure 2) show that for three vowel qualities the short and long members differ significantly in their F1: /ɪ/ has a larger F1 than /i:/ by 2 ERB, /a/ has s smaller F1 than /a:/ by 0.5 ERB, and /o/ has a smaller F1 than /o:/ by 0.7 ERB.

For F2, we again found a main effect of sex showing that vowels have overall larger F2 in women than in men (by on average 1.2 ERB, $t[21.2] = 6.443$, $p = 2 \times 10^{-6}$). Also, short vowels were found to have an overall larger F2 than long vowels (by on average 0.3 ERB, $t[58.6] = 3.947$, $p = 2 \times 10^{-4}$). All main effects of vowel quality as well as all interactions of vowel quality and vowel length came out as significant, we thus directly turn to the pairwise comparisons of estimated means. Comparisons of vowel qualities detected a significant between-vowel difference for all pairs except for short /u/ versus short /o/ (and short /u/ versus short /a/, which however was not a planned comparison in our design). The comparison of F2 between short and long members within each vowel quality reveal that for *i*, *e*, and *a* the long member has a higher F2 than the short member (by 1.2, 0.5, and 0.7 ERB), while for *u* and *o* it is the short member that has a higher F2 than the long one (by 2.8 and 1 ERB, respectively).

As for the higher formants, the analyses showed that the F3-F2 distance is larger in /i:/ than in /ɛ:/ by 1.02 ERB, and larger in /ɪ/ than in /ɛ/ by 2 ERB ($t[60] = -5.8$, and -11.4, respectively, both *p*s < .001). The F4-F3 distance is smaller in /i:/ than in /ɪ/ by 0.85 ERB, and smaller in /i:/ than in /ɛ:/ by 0.74 ERB ($t[60] = 5.5$, and -5.5, respectively, both *p*s < .001).

3.2. Vowel duration

For duration, the intercept was estimated as 0.096 norm s ($t[19.9] = 50.709$, $p < 2 \times 10^{-16}$), meaning that the average duration of vowels in our data set was 0.096 normalized seconds (that is, the mean vowel duration was 96 milliseconds in an average 500-ms-long word). There was a main effect of vowel length confirming that long vowels have overall larger duration than short vowels (by on average 0.051 norm s, $t[20.1] = -16.296$, $p = 5 \times 10^{-13}$). Furthermore, vowels produced by men were slightly longer than vowels produced by women (by on average 0.006 norm s, $t[28.4] = -2.342$, $p = .026$).

The main effect of vowel quality was significant for the *e-i* and for the *e-a* comparison suggesting that *e* is longer than *i* and shorter than *a* (by on average 0.030 and 0.008 norm s, respectively, both *p*s < 0.05). As vowel quality interacted with vowel length for three out of the four vowel contrasts (*e-i, e-o* and marginally for *e-u*), we turn to the inspection of the estimated means to unpack the interactions (involving both the planned and unplanned comparisons). Correcting alpha for all of the 20 individual comparisons, the data reveal that amongst long vowels, /aː/, /ɛː/, /oː/, and /uː/ are significantly longer than /iː/ by about 0.030 norm s (a similar but nonsignificant trend is seen in /ɛː/ and /oː/ tending to be longer than /uː/, by about 0.012 norm s). Amongst the short vowels, /a/, /ɛ/, and /u/ are trending towards being longer than /ɪ/ by about 0.014 norm s, reaching significance only for the /ɛ/-/ɪ/ comparison. As for the long-short comparisons within vowel qualities, all turned out significant implying that duration distinguishes a short and a long member in all 5 vowel pairs.

The model for duration ratios yields an intercept of 1.76, implying that long vowels are on average 1.76 times longer than short vowels ($t[80] = 20$, $p < 2 \times 10^{-16}$). The analysis further reveals that the long/short ratio in high vowels (*i* and *u)* is smaller than the ratio in the low vowel *a* (by 0.32, $t[80] = 2.978$, $p = .0038$) which in turn is smaller than the ratio in the mid vowels (*e* and *o*; by 0.35, $t[80] = -3.341$, $p = .0013$). The long/short ratio being the largest in mid vowels seems to be driven mainly by the large long/short ratio in *o* which significantly outweighs the long/short ratio in the other mid vowel quality *e* (by 0.19, $t[80] = 2.064$, $p = .042$); see also Table 4.

Table 4. Raw and word-length normalized duration of the 10 Czech monophthongs, and long/short duration ratios, averaged across sexes.

vowel	Raw duration (s) mean and 95% c.i.		Normalized duration (norm s) mean and 95% c.i.		Long/Short ratio
ɪ	0.052	0.046–0.058	0.061	0.054–0.067	1.66 (1.49–1.83)
iː	0.090	0.082–0.097	0.098	0.090–0.106	
ɛ	0.069	0.062–0.075	0.076	0.071–0.081	1.78 (1.61–1.95)
ɛː	0.125	0.114–0.136	0.133	0.126–0.140	
a	0.072	0.066–0.077	0.075	0.070–0.080	1.73 (1.56–1.90)
aː	0.138	0.128–0.147	0.126	0.119–0.133	
o	0.059	0.053–0.065	0.067	0.060–0.075	1.97 (1.80–2.14)
oː	0.137	0.126–0.147	0.132	0.122–0.143	
u	0.069	0.063–0.075	0.074	0.069–0.079	1.65 (1.48–1.82)
uː	0.104	0.093–0.114	0.119	0.111–0.128	

4. Discussion

In this study we recorded the spontaneous speech of 20 speakers representative of the general Czech-speaking population (who use the standard variety of Czech spoken in the central Bohemian area) and analysed the vowels occurring in the initial, stressed, syllable of disyllabic content words (nouns). We performed acoustical and statistical analyses of the vowels' spectral properties, namely, F1 and F2 in all 10 monophthongs, and F3 and F4 in the four front vowels, and on duration, namely, vowel duration normalized for word duration, and long/short duration ratios.

4.1. Acoustic characteristics of Czech monophthongs

The results showed that the high front vowel pair is reliably distinguished by F1: the long /iː/ has a smaller F1 than the short /ɪ/, by 2 ERB, a difference which by far exceeds the just noticeable difference for formants (which is 0.2 ERB for [ɪ]-like vowels, Kewley-Port, 1995).The significant lowering of the short /ɪ/ in the vowel space is further documented by this vowel being, in terms of F1, four times closer to the short mid back /o/ than the long /iː/ is to the long mid /oː/. This F1 distinction between /ɪ/ and /iː/ is in line with previous acoustic measurements of vowels from read speech (Skarnitzl & Volín, 2012; Šimáčková et al., 2012; Paillereau, 2016) and matches the impressionistic observations of spontaneous speech from the 20th century (Frinta, 1909, 1924; Beneš, 1943; Chlumský, 1928; Hála, 1955; note that Hála, 1941, 1962 noticed an openess not only of the short but also of the long front high vowel).

The data further showed an asymmetry across the mid vowels. The front /ɛ/ and /ɛː/ are realized with higher F1 than the back /o/ and /oː/. This disentanglement between front and back vowels is further strengthened by the front (phonologically) mid vowels being more similar in F1 to the *low* /a/ and /aː/ than to the other mid vowel quality, the back /o/ and /oː/. The front-back asymmetry could be explained in terms of Lindblom's Adaptive Dispersion Theory (Liljencrants & Lindblom, 1972) which argues that the (changes in) individual vowel qualities are determined by the entire system of vocalic contrasts. Thus, in order to maximize the perceptual contrast between short /ɪ/ (which is realized with much higher F1 than the long /iː/) and the front /ɛ/ and /ɛː/, the F1 of the front mid vowels aims at high(er) F1 values. In the back part of the vowel system, no evidence is found for a lowering of the short high /u/ and there is thus no reason for the mid /o/ to be pushed towards higher F1 values.

In terms of F2, the long vowels had more peripheral values than their short counterparts. Interestingly, however, this effect for the back vowels was more than twice as large as that for the front vowels indicating a significant fronting of the short /o/ and /u/. The apparent fronting of the short back vowels possibly had two interrelated causes. Firstly, most of the post-vocalic consonants were coronals that notoriously cause rising of back vowels' F2 (Stevens & House, 1963), and due to the short vowels' inherent shortness the coarticulatory effects of flanking consonants affect a larger proportion of the vowel than is the case for inherently long vowels. Secondly, due to a generally less careful articulation in spontaneous (as compared to read) speech, the back vowels for which speakers aimed at only short duration underwent target undershoot not reaching the peripheral, low, F2

values representative of phonological backness. To what extent it was the consonantal context or the spontaneous speech style that lead to the fronting of the short back vowels remains a question open for future research.

Curiously, our data revealed that long low vowel /aː/ has a slightly higher F2 than the short low /a/. Although the perceptual reality of the 0.7-ERB difference is questionable, a fronting of the long /aː/ has been mentioned previously by Skarnitzl & Volín (2012) and reported by Paillereau (2016) for speakers of the regional Pilsner dialect of Czech.

Results on higher formants showed that F3 and F4 are converging in the long /iː/ more so that they do in the short /ɪ/ (and in the short /ɛ/). This finding is interesting from a cross-linguistic perspective: the F4-F3 difference that we found for Czech /iː/ resembles that of the French (prepalatal) /i/ that had been thought to exhibit a cross-linguistically unique pattern of F3-F4 focalization (Gendrot et al. 2008, Vaissière 2011).

Table 5. Third- and fourth-formant values (ERB) of /i/ in 8 languages (Table 1 from Gendrot et al. 2008) and of the Czech /iː/ and /ɪ/ (the present data, in bold), as well as the difference F4-F3 (in ERB), averaged across sexes.

	F3	F4	F4-F3
French	24.15	25.92	1.77
Czech /iː/	**24.82**	**26.74**	**1.91**
Arabic	23.62	25.59	1.97
Mandarin	24.12	26.29	2.17
Spanish	23.96	26.27	2.31
Italian	23.67	26.16	2.49
English	23.05	25.79	2.74
German	23.40	26.23	2.83
Czech /ɪ/	**23.38**	**26.40**	**3.01**
Portuguese	23.05	26.13	3.08

Table 5 gives an overview of F3 and F4 values, and their psychoacoustic distance (in ERB), that had been previously reported for 8 languages by Gendrot et al. (2008) along with the currently measured values for Czech /iː/ and /ɪ/. It is seen that while the focalization is numerically smallest in the French /i/, Czech /iː/ appears to be more focalized than the /i/ in the 7 remaining languages (and at the same time seems to have the highest F3 and F4 values of the entire sample). Investigation of higher formants may be beneficial not only from cross-linguistic perspective but also cross-dialectally. The F1-F2 difference between /iː/ and /ɪ/ that we report here holds for Bohemian varieties of Czech and its extent is reportedly smaller in Moravian varieties (Šimáčková et al., 2012): future studies could investigate whether there are (also) any dialectal differences in the extent to which higher formants cue the distinction between the short and the long high front vowel.

We found that duration reliably distinguishes between the short and the long phoneme across all five vowel qualities. Amongst long vowels /iː/ was the shortest and since a similar trend was seen also in the short vowel set, the apparent shortness of /iː/ did

not lead to an exclusively smallest long/short ratio for the /i:/-/ɪ/ vowel pair. Long/short ratios of the high front and high back vowels were the smallest, followed by an intermediate long/short ratio for the low vowel quality and the largest ratio for the mid vowels. Crucially, however, the /i:/-/ɪ/ ratio measured here, i.e. 1.66, was much larger than the /i:/-/ɪ/ ratio of 1.29 reported by Podlipský et al. (2009) (and by comparing our lower confidence bound 1.49 to the mean of Podlipský et al., this difference was most likely significant). The methodological differences between ours and Podlipský's et al. study lying in the speech style (spontaneous vs read, respectively), population (general public vs news reporters, respectively), and in the number of participants (20 vs 6, respectively) suggest that the data from the current study may reflect Czech vowel durations more veridically than the data reported in the 2009 study.

Apart from the disparate finding for /i:/-/ɪ/, the long/short ratios for the remaining 4 vowel pairs resemble the ratios reported for these vowel pairs by Podlipský et al. The average long/short ratio in our spontaneous speech material was 1.76 which is smaller than the long/short ratio of 2 reported by Chlumský (1928), and except /o:/ none of the long vowels comes close to potentially being twice as long as the short one (with the highest upper confidence bounds of 1.9, 1.95, and 2.14 for /a:/-/a/, /ɛ:/-/ɛ/, and /o:/-/o/, respectively). It thus appears that duration ratios between long and short vowels may have become reduced over the past century. However, further research is needed that would assess and directly compare vowel durations across speech styles to resolve the conflict between ours and Podlipský et al. (2009) study with respect to the /i:/-/ɪ/ ratio.

As a final note on duration, we found that the long/short ratio was the largest for /o:/ vs. /o/, an effect which most likely stems from the fact that the long /o:/ is not a genuine Czech phoneme; it has come to the language with recent borrowings, and occurs only in a small set of relatively infrequently used words (Ludvíková & Kraus, 1966; Podlipský et al., 2009; Šimáčková et al., 2012). Because there is a link between item frequency and prototypicality of articulation (e.g. Aylett and Turk, 2006), the infrequent long vowel /o:/ may be realized as a hyperarticulated, unnaturally produced speech segment.

4.2. On the phonological notation

As noted in the Introduction, across authors and across studies there seems to be an inconsistency in how Czech vowels are transcribed phonemically. One, and nowadays probably the most frequently used, approach to transcribing Czech vowels is phonetically motivated and thus depicts both the length and the quality distinction in the high front vowels by transcribing them as /i:/ and /ɪ/ and also depicts the significant lowering of the mid front vowels – in contrast to the mid back vowels – by transcribing them as /ɛ(:)/ and /o(:)/, respectively. The phonetically motivated transcription has been used across acoustic vowel studies (including the present one) as well as in phonological descriptions of Czech (Dankovičová, 1997; Podlipský et al., 2009; Chládková et al., 2009; Šimáčková et al., 2012; Paillereau, 2016; Skarnitzl et al., 2016; Chládková et al., 2019).

The other approach to transcribing Czech vowels seems to be formally motivated such that it aims to capture the phonological symmetry of the system omitting some of the (relevant) phonetic information, which results in /i: i e: e a: a o: o u: u/ and has been used

by Bičan (2013) and Palková (1997), both of whom make explicit notes on the phonetic deviations violating the symmetry. Yet other recent authors' symbol use seems to be motivated both phonetically and phonologically resulting in a somewhat inconsistent description. For instance, transcribing the monophthongal phonemes as /i: ɪ ɛ: ɛ a: a ɔ: ɔ u: ʊ/, Duběda (2005) captures the actual phonetic realization of the front vowels but, at the same time, attempts to instantiate a front-back symmetry by using distinct symbols for the short versus the back high back vowel, and by transcribing the mid back vowel as an open /ɔ(:)/. The rather ambiguous choice to realign the back vowels to conform to the phonetically-grounded realizations of the front vowels has not been, to the best of our knowledge, supported by any acoustic or perceptual studies (although early Czech phoneticians did note a lowering of /o/ in the contemporary speech, see below).

Most of the earlier authors were, too, aware of the vowels' unique phonetic realizations but purposefully referred to the system as symmetrical with their goal being to *prescribe* how Czech speakers should realize vowels wishing to prevent the actually observed, disfavored open realizations (mostly pertaining to lowering of the front mid vowel; e.g. Hála, 1941, 1962 and Beneš, 1943, but see also Borovičková & Maláč, 1967 who describe the realizations of /i/ and /i:/ as spectrally similar). Frinta (1909, 1924) was one of the few early authors using phonetically motivated symbols aiming to *describe* the Czech phonemes as they are realized by an average speaker of Czech (and not to prescribe how the vowels should be pronounced). On the basis of impressionistic observations, Frinta (1909, 1924) used /ɛ/ and /ɔ/ to capture the lowering of the mid vowels and used /i/ and /i:/ for the high front vowels noting a spectral difference between them but not considering it large enough to be captured in the transcription.

In the present study that is aimed as a description of the spectral and durational characteristics of Czech vowels, we employed the transcription /i: ɪ ɛ: ɛ a: a o: o u: u/ capturing the significant spectral distinction within the high front vowel pair and the lowering of the mid front vowels. The present data do not support the use of /ʊ/ for the short high back vowel as we did not detect an F1 difference between the short and the long high back vowels (not detecting any F1 difference between /u/ and /u:/ of course does not mean that the difference may not exist but it does justify not introducing the use of two different symbols for those two vowels). We also keep transcribing the mid back vowels as /o(:)/ to depict the significant asymmetry in the F1 of front versus back mid vowels.

The variations in phonemic symbol use are apparent not only between authors but also between studies by the same authors who transcribe the Czech mid front vowel as /e/ in some cases (Skarnitzl & Volín, 2012) but as /ɛ/ in others (Podlipský, Skarnitzl & Volín, 2009; Skarnitzl, Šturm & Volín, 2016). Firstly, as Wells (2001) pointed out, the choice of IPA symbols can be adapted according to the audience that one and the same author may aim at with different studies. The above described inconsistency does not seem, however, to be due to different audiences that the authors aim a – all of them reporting on acoustic (and perceptual) properties of vowels. It rather demonstrates a general difficulty to transcribe mid vowels in a language that has only 3 degrees of vowel height with an IPA chart that was designed on the basis of French, English and German vowel inventories (Grammont, 1933), all of which contrast 4 degrees of height, and thus contrast also /e/ and /ɛ/. The mid front vowel of languages with 3 degrees of vowel height is then mostly transcribed as /e/ (to what extent that symbol reflects the true phonetic realization of

this vowel is not discussed here) which may support the occasional tendency to use that symbol also for Czech (e.g. by Nicolaidis, 2003; Lengeris & Hazan, 2010 in Greek; Fox et al., 1995; Cervera et al., 2001; Chládková et al., 2011 in Spanish; Hirata & Tsukada, 2009; Niimi et al., 1994; Kamiyama & Vaissière, 2009; Hirayama, 2003 in Japanese; Jones, 1953; Padgett, 2004; Lyakso et al., 2009 in Russian).

To conclude on the phonemic transcription motivated by acoustic results, it should be noted that even though the aim is to render the phonemic transcription as explicit as possible (i.e. truthfully reflecting the phonetic reality), different diacritics rendering any possible phonetic detail are still avoided. For instance, Šimáčková et al. (2012) employed two different length marks [ː] and [.] to capture the different durations of the long high front vowel across two major dialects of Czech. Although we found here that the long high front vowel is shorter than the other long vowels, the durations of the long vowels are larger than the durations of the short vowel across all five vowel pairs; therefore, we represent the long phoneme by appending /ː/ to the vowel symbol throughout for all the five Czech length contrasts. The long/short ratio reported here does not seem to be exceptionally small for the high front vowel pair, instead it seems to gradually decrease from mid to low and to high vowels. This could be understood as a physiologically conditioned duration-ratio phenomenon causing long vowels at the periphery of the vowel space to be sustained for a shorter amount of time than long vowels closer to the central part of the vowel space.

We should note here that other languages, too, lack a consensus on the phonological transcription of vowels. To name what is perhaps the most widely known instance, in order to transcribe lax/tense vowels in British English three main types of transcriptions have been used: quantitative transcription (using the same vowel symbol and appending a length mark, e.g. Palmer, 1920; Jones, 1932), qualitative transcription (using different symbols and no length mark, e.g. Ladefoged & Broadbent, 1957), or quantitative-qualitative transcription (using both, e.g. Cruttenden, 2014 and most contemporary authors). Another example is that of Japanese, in which the inconsistency concerns the phonemic notation of the back high vowel; many authors use /u/ (Hirata & Tsukada, 2009; Niimi et al., 1994; Kamiyama & Vaissière, 2009; Hirayama, 2003) but it is also possible to find the symbol /ɯ/ (Lambacher et al., 2005), which reflects the unrounded phonetic realization of the vowel.

There have been debates on the correctness of the different notations. According to some authors, phonemic symbols should correspond to the most frequent allophones and only those differences which cannot be expressed in terms of phonological rules should be made explicit by using a specific phonemic symbol (Duchet, 1992). According to this point of view, marking vowel length in tense/lax English vowel pairs would be a redundant information, because it can be inferred. On the contrary, it is accepted that the different aforementioned notations were formed according to IPA principles and thus are all scientifically correct (Wells, 2011). It is then up to each author to decide which notation she or he will use. The choice should be determined by what is being the message and to what type of readers the study is addressed.

We adhere to the idea that different phonemic notations are acceptable and "right" in their own sense. This is why – despite having adopted a phonetically-motivated transcription in the current work – the authors of this paper are not strictly opposing a transcription of Czech vowels that would employ the symbols /i iː e eː a aː o oː u uː/ if the focus is

on a formal description of the system and if the study is not targeting an audience who might try to pronounce the vowels according to the notation (e.g. speakers with speech disorders undergoing formal training or learners of Czech as a second language). After all, for the language-learning child, the only important information that she extracts from the phonetic environment might be that there are ten different clusters or perhaps categories for vowels in the ambient language (and the scientist may arbitrarily choose to transcribe them in a non-IPA based alphabet as ❥△❦ ❇ ◎ ✷ ✓ ☛ ❀ ◇) and perhaps the child sooner or later figures out that those ten discrete units are in fact a combination of, for instance, five times two category levels (such as ● ○ ✪ ☆ ◂ ◃ ☛ ☞ ■ □). While we still know little about how and when the developing child structures the phonetic vowel space in particular ways, the linguist has the knowledge, a particular aim, and the choice of how to appropriately convey their message. Crucially, whether an author's main aim is to reflect the phonetic reality, or whether it is to formalize and simplify, the approach she or he takes should be consistent and applied across all units of the system.

5. Conclusions

The present paper contributes a thorough spectral and durational characteristics of Czech vowels. Twenty speakers representative of the general, standard-Czech speaking population were recorded while spontaneously producing speech. Analyses of their vowels revealed that the mid front vowels are significantly lowered in the vowel space, appearing less distant in their F1 from the low vowels than from the mid back vowels. Confirming previous studies, the short high front vowel was found to be spectrally distinct from its long counterpart, namely, lowered along the F1 dimension. No such F1 differences were detected in the /u/-/u:/ vowel pair, which, instead revealed a significant difference in F2 with the short phoneme being fronter than the long one (and similarly for the /o/-/o:/ contrast). Whether this F2 distinction between short and long phonemes in back vowels is a feature of spontaneous speech or whether it is due to the consonantal context occurring in the present study remains to be shown in future work. Our data demonstrated that in spontaneous speech duration reliably distinguishes between short and long phonemes across all vowel pairs, including /i:/ vs /ɪ/, which runs contrary to some recent speculations that the short-long contrast in high front vowels may no longer be (primarily) cued by duration (Šimáčková et al., 2012). The study concluded with a discussion of whether and how phonological transcription can best reflect an author's goal and help the reader understand the linguist's message.

ACKNOWLEDGMENTS

This work was funded by an internal grant from Charles University PRIMUS/17/HUM19 and by Czech Science Foundation grant no. 18-01799S. The authors thank Kristýna Hrdličková, Zuzana Oceláková, Martina Černá, and Radka Klimičková for help with data collection and annotation.

REFERENCES

Borovičková, B. & Maláč, V. (1967). *The Spectral Analysis of Czech Sound Combinations*. Praha: Academia.

Beneš, J. (1943). Porušování základní barvy českých samohlásek. *Naše řeč*, 27(3), 64–65.

Bičan, A. (2013). *Phonotactics of Czech*. Frankfurt am Main: Peter Lang.

Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer. Version 6.0.40*. Retrieved from www.praat.org.

Cervera, T., Miralles, J. L. & Gonzalez-Alvarez, J. (2001). Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *Journal of Speech, Language, and Hearing Research*, 44, 988–996.

Chládková, K., Boersma, P. & Podlipský, V. J. (2009). Online formant shifting as a function of F0. *Proceedings of Interspeech 2009*, 464–467.

Chládková, K., Escudero, P. & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *Journal of the Acoustical Society of America*, 130(1), 416–428.

Chládková, K., Černá, M., Paillereau, N., Skarnitzl, R. & Oceláková, Z. (2019). Prenatal infant-directed speech: vowels and voice quality. In: *Proceedings of ICPhS 2019*, 1525-1529.

Chlumský, J. (1928). *Česká kvantita, melodie a přízvuk*. Praha: Česká akademie věd a umění.

Crothers, J. (1978). Typology and universals of vowel systems. In: J. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language*, Vol. 2: *Phonology* (pp. 93-152). Standford: Standford University Press.

Cruttenden, A. (2014). *Gimson's Pronunciation of English*. London and New York: Routledge.

Dankovičová, J. (1997). Czech. *Journal of the International Phonetic Association*, 27, 77–80.

Duběda, T. (2005). *Jazyky a jejich zvuky: univerzálie a typologie ve fonetice a fonologii*. Praha: Karolinum.

Duchet, J. L. (1992). *La Phonologie*. Paris: P.U.F.

Escudero, P., Boersma, P., Rauber, A. S. & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America*, 126, 1379–1393.

Fant, G. (1960). *The Acoustic Theory of Speech Production*. The Hague: Mouton.

Fant, G. (1969). Formant frequencies of Swedish vowels. In: *Speech Transmission Laboratory, Quarterly Progress and Status Report 3* (pp. 94–99). Stockholm: Royal Institute of Technology.

Frinta, A. (1909). *Novočeská výslovnost: pokus o soustavnou fonetiku jazyka českého*. Praha: Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.

Frinta, A. (1925). *A Czech phonetic reader*. London: University of London Press.

Flemming, E. (2019). Implications of [i] vowels for the theory of vowel inventories. Presented at *LSA Annual Meeting*, NYC.

Fox, R. A., Flege, J. E. & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, 97, 2540–2550.

Gendrot, C., Adda-Decker, M. & Vaissière, J. (2008). Les voyelles /i/ et /y/ du français: focalisation et variations formantiques. In: *Proceedings of JEP*, 205–208.

Grammont, M. (1933). Traité pratique de prononciation française. Paris: Delagrave.

Hála, B. (1941). *Akustická podstata samohlásek*. Praha: Česká akademie věd a umění.

Hála, B. (1955). *Výslovnost* spisovné češtiny: její zásady a pravidla. Praha: Nakladatelství Československé Akademie Věd.

Hála, B. (1962). Uvedení do fonetiky češtiny na obecně fonetickém základě. Praha: ČSAV.

Hirata, Y. & Tsukada, K. (2009). Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica*, 66, 129–149.

Hirayama, M. (2003). Contrast in Japanese vowels. *Toronto Working Papers in Linguistics*, 20, 115–132.

Jones, D. (1932). *An Outline of English Phonetics*. Leipzig: B. G. Teubner.

Jones, L. G. (1953). The vowels of English and Russian: An acoustic comparison. *Word*, 9(4), 354–361.

Kamiyama, T. & Vaissière, J. (2009). Perception and production of French close and close-mid rounded vowels by Japanese-speaking learners. *Acquisition et interaction en langue étrangère Aile... Lia*, 2, 9–41.

Kamiyama, T. (2011). Pronunciation of French vowels by Japanese speakers learning French as a foreign language: Back and front rounded vowels /u y ø/. *Phonological Studies: Phonological Society of Japan*, 97–108.
Kewley-Port, D. (1995). Thresholds for formant frequency discrimination of vowels in consonantal context. *Journal of the Acoustical Society of America*, 97(5), 3139–3146.
Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
Ladefoged, P. & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A. & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227–247.
Lengeris, A. & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America*, 128(6), 3757–3768.
Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. (2018). *Estimated marginal means, aka least-squares means*. R package version 1.2.4. Retrieved from https://cran.r-project.org/package/emmeans.
Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862.
Ludvíková, M. & Kraus, J. (1966). Kvantitativní vlastnosti soustavy českých fonémů. *Slovo a slovesnost*, 27, 334–344.
Lyakso, E., Frolova, O. & Grigorev, A. (2009). The acoustic characteristics of Russian vowels in children of 6 and 7 years of age. In: *Proceedings of Interspeech 2009*, 1739–1742.
Nicolaidis, K. (2003). Acoustic variability of vowels in Greek spontaneous speech. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, 3221–3224.
Niimi, S., Kumada, M. & Niitsu, M. (1994). Functions of tongue-related muscles during production of the five Japanese vowels. *Ann. Bull. R. I. L. P. Univ. Tokyo*, 28, 33–40.
Padgett, J. (2004). Russian vowel reduction and dispersion theory. *Phonological Studies*, 7, 81–96.
Paillereau, N. (2016). 'Identical' vowels in L1 and L2? Criteria and implications for L2 phonetic teaching and learning. In: Liszka, S. A., Leclercq, P., Tellier, M. & Daniel, G. (Eds.), *EUROSLA Yearbook 2016* (pp. 144–178). Amsterdam: John Benjamins.
Paillereau, N. & Skarnitzl, R. (2019). An acoustic-perceptual study on Czech monophthongs. In: T. Radeva-Bork & P. Kosta (Eds.), *Current developments in Slavic Linguistics. Twenty years after (based on selected papers from FDSL11)* (pp. 453–466). Berlin: Peter Lang.
Palková, Z. (1997). *Fonetika a fonologie češtiny*. Praha: Karolinum.
Palmer, H. E. (1920). *A First Course of English Phonetics*. Cambridge: W. Heffer and Sons Ltd.
Podlipský, V. J., Skarnitzl, R. & Volín, J. (2009). High front vowels in Czech: A contrast in quantity or quality? In: *Proceedings of Interspeech 2009*, 132–135.
Podlipský, V. J., Chládková, K. & Šimáčková, Š. (in press). Spectrum as a perceptual cue to vowel length in Czech, a quantity language. *Journal of the Acoustical Society of America: Express Letters*.
R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org.
Skarnitzl, R. & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18, 7–11.
Skarnitzl, R. (2012). Dvojí *i* v české výslovnosti. *Naše řeč*, 95, 141–153.
Skarnitzl, R., Šturm, P. & Volín, J. (2016). Zvuková báze řečové komunikace: fonetický a fonologický popis řeči. Praha: Karolinum.
Stevens, K. N. & House, A. S. (1963). Perturbations of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research*, 6, 111–128.
Šimáčková, Š., Podlipský, V. J. & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association*, 42, 225–232.
Vaissière, J. (2009). Articulatory modeling and the definition of acoustic-perceptual targets for reference vowels. *The Chineses Phonetics Journal*, 2, 22–33.

Vaissière, J. (2011). On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. In: *Proceedings of ICPhS 2011*, 52–59.
Wells, J. (2001). IPA transcription systems for English. Retrieved from https://www.phon.ucl.ac.uk/home/wells/ipa-english-uni.htm (last accessed on May 16, 2019)

RESUMÉ

Článek se zaměřuje na akustickou analýzu českého vokalického systému a popisuje spektrální a délkové charakteristiky 10 českých monoftongů. Oproti dosavadní literatuře jsou studovány hlásky z projevu spontánního a sesbíraného jak od mužských tak od ženských mluvčích, kteří byli rekrutováni ze široké veřejnosti (nejedná se tedy pouze o studenty nebo reportéry médií, jejichž čtený projev byl předmětem zkoumání v předchozích studiích). Spektrální analýza zahrnuje první čtyři formanty (F1 až F4). Analýza hláskové délky se zaměřuje na poměrné trvání dlouhých vůči krátkým vokálům a srovnává trvání samohlásek normalizované pro délku slov. Výsledky potvrzují významný spektrální rozdíl u vysokých předních samohlásek, kde je fonologicky krátké /ɪ/ realizováno s vyšším F1 než dlouhé /iː/. Přední střední vokály /ɛ/ a /ɛː/ jsou realizovány s relativně vysokým F1, kvalitativně jsou tak dokonce blíže k nízkým samohláskám /a/ a /aː/ než k zadním středním samohláskám /o/ a /oː/. Novým zjištěním je, že krátké zadní samohlásky /o/ a /u/ mají vyšší F2 než jejich dlouhé protějšky: tento mírný posun dopředu lze pravděpodobně připsat spontánnímu stylu řeči a také převážně koronálnímu kontextu, ve kterém se samohlásky objevovaly. Na rozdíl od moderní literatury, která uvádí velmi malý poměr trvání dlouhých vůči krátkým vysokým vokálům, naše výsledky ukazují, že trvání konzistentně rozlišuje fonologickou délku napříč všemi pěti vokalickými páry: dlouhé samohlásky jsou v průměru 1,76krát delší než samohlásky krátké. Diskuze je uzavřena zamyšlením se nad vztahem akustických vlastností hlásek a jejich fonologickou transkripcí.

Nikola Paillereau
Kateřina Chládková
E-mail: nikola.paillereau@ff.cuni.cz

Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic

Institute of Psychology, Czech Academy of Sciences
Prague, Czech Republic

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 97–108
PHILOLOGICA 2/ PHONETICA PRAGENSIA

THE RELATIONS BETWEEN PHONOTACTICS AND SPEECH RHYTHM IN CZECH

ELIŠKA CHURAŇOVÁ

ABSTRACT

The main objective of this study is to explore the relationships between the phonotactic structure of the Czech stress-group and rhythm of speech. Three most frequent consonantal-vocalic (CV) structures of Czech two-syllable stress-groups were selected for the purpose of this study: CVCV, CVCCV, and CCVCV. In an auditory experiment, which contained the mutual comparison of stress-groups or the comparison of a stress-group and a low-frequency shadow of a stress-group, the respondents established how similar the rhythmic pattern of each couple of stress-groups sounded.
The results indicate that the position of a consonantal cluster within the stress-group is the strongest phonotactic factor in perception of the rhythmic similarity. The number of consonants within a consonantal cluster and the presence of a long vowel in both stress-groups are considered weaker factors for perceiving the rhythmic similarity by the respondents. Possibilities for a follow-up research are proposed for the factors that did not reach statistical significance, i.e., the difference in sonority or voicing of consonants.

Key words: speech rhythm, stress-group, phonotactics, consonantal-vocalic structure, Czech

1. Introduction

All types of natural rhythmic behaviour show one essential characteristic, which is periodical repetition of certain patterns. These patterns consist of alternating of contrasts which are perceived as regular by the listeners. Rhythmic processes are generally easier and more stable than non-rhythmic and, therefore, they are also more eligible. The concept of stability of rhythmical actions was explored, for instance, on a synchronization of movements with metronome pulses: The experiments revealed that from a certain frequency of clicks, the anti-phase synchronization changed to in-phase (Kelso, 1995; Repp, 2005). The importance of the rhythmical events is shown even on the brain processing level (Grossberg, 2003).

Rhythmic behaviour is observed in many areas of nature. The manifestations of rhythm are embodied in the physiology of living creatures: the heartbeat, breathing, chewing, etc. Human speech is also one of the types of behaviour in which rhythmic

https://doi.org/10.14712/24646830.2019.20

aspects take place. The contrasts creating the impression of a regular rhythm are lower-level alternations such as of consonants and vowels, and higher-level alternations like stressed and non-stressed syllables.

The assumption that speech rhythm is related to the phonological structure of a given language appeared in the middle of the 20th century. Languages were classified into two (and later in three) groups according to their rhythmical characteristics (Pike, 1945; Abercrombie, 1967). In the so-called "stress-timed" languages the intervals between stresses and in "syllable-timed" the intervals between nuclei of the syllables were assumed to be equal in duration. It was expected that the latter languages, in contrast to the former, allowed for complex consonant clusters and tended to reduce vowels in unstressed positions (Dauer, 1983). This idea reassumed in many studies trying to confirm its conclusions through measuring consonantal and vocalic intervals in both classes of languages (e.g., Grabe & Low, 2002; Ramus et al., 1999). Despite all the effort, the languages could not be divided in the above mentioned groups unambiguously. Some languages, like Czech, had characteristics of both categories to various extents (Dankovičová & Dellwo, 2007). It is possible that these rhythmical features of languages create a continuum rather than separate groups. The weak point of the studies measuring duration of speech intervals is that these measures said nothing about the nature of rhythm itself. In the research of the rhythm of speech it is necessary that the listener is taken into account. Rhythm as such rests in the *impression* of isochrony (Lehiste, 1977, 1979), not in the objective equality of the intervals between contrasting events.

Nevertheless, the fact that the interval-measuring models seem unsubstantiated does not imply that consonantal-vocalic structure has no meaning in perception of rhythm. Šturm & Volín (2016) proved that number and type of consonants in a consonantal cluster had some influence on perception of rhythm in Czech, but the exact relationship between rhythm of speech and phonology of the Czech language remains unresolved.

The carrier of rhythm – the smallest entity, on which speech rhythm manifests itself – is presumably the stress-group, in Czech with fixed stress on the first syllable. Within this Western Slavonic language many types of the consonantal-vocalic (CV) structure of stress-groups exist (CV, CCV, CVCV, CCVC etc.). In Czech, the most frequent phonotactic patterns in stress-groups are CVCV (e.g., [budɛ]), CVCCV (e.g., [nɛzna:]), CVCVCV (e.g., [bofiati:]), CCVCV (e.g., [stoji:]) and CVCVC (e.g., [potok]) (Churaňová, 2013). The aim of this study is to explore, whether these different phonotactic structures carry also distinct rhythmical patterns from the viewpoint of speech perception.

2. Method

2.1. Material

Fourteen texts of radio news broadcasts were selected for the purposes of the main experiment. They were read by professional native speakers of standard Czech (7 male and 7 female), without any dialect or slang features. The recordings were taken from the Prague Phonetic Corpus (Skarnitzl, 2010). Twelve of the recordings were used in previous study of the current author (Churaňová, 2013); two recordings were added to that

material. All recordings were subsequently processed in the programme Praat (Boersma & Weenink, 2018). The individual texts ranged between 234 and 603 stress-groups; the total sample consisted of 6216 stress-groups.

The duration of the material was 51.85 minutes in total. The recordings were divided into breath-groups, and TextGrid objects were accordingly created for annotation. Individual breath-groups were then divided into phones and words using the Prague Labeller algorithm (Pollák et al., 2007). Word boundaries were marked automatically, stress-groups were segmented manually (see Churaňová, 2012: 81 for further details). It was also necessary to correct all transcription errors caused by automatic annotation. All temporal and phonological data required (e.g., the duration of each stress-group and phone, the consonants and vowels in each stress-group) were gained by using an ad-hoc Praat script.

The analysis of the material revealed that the most frequent consonantal-vocalic patterns of stress-groups are CVCV (6.7% from all the stress-groups), CVCCV (5.3%), CVCVCV (4.4%), CCVCV (4%), and CVCVC (3.5%). These results agree with the findings of Churaňová (2012, 2013). Since two-syllable stress-groups are typical for the Czech language, three most frequent two-syllable stress-groups (CVCV, CVCCV, CCVCV) were selected for the present experiment as possible items in the auditory perception test.

332 different types of stress-groups respecting the three consonantal-vocalic patterns were found in the material. The candidates for items in the auditory experiment were estimated with regard to balanced frequency of the word(s) constituting the stress-groups with certain phonotactic patterns. For the final candidates only the correctly pronounced stress groups with duration limited by ±1 standard deviation from the average duration of all stress-groups with the same vocalic quantity were selected. The stress-groups with neutral intonation were preferred.

32 phonotactic patterns (with voicing/sonority characteristics of each consonant considered) which met the above-mentioned conditions were selected for the auditory experiment. The patterns were divided into pairs according to a certain phonological or phonotactic characteristic which distinguished one pattern from the other (e.g., a pattern consisting of a voiced obstruent, short vowel, voiceless obstruent and a short vowel differs from its counterpart which consisted of a voiceless obstruent, short vowel, voiceless obstruent and a short vowel only in voicing of consonants). The pairs established four groups according to the characteristic in which the patterns in the pair differed from each other: voicing of obstruents, number of segments in a consonant cluster, position of a consonant cluster and sonority of consonants.

2.2. Experimental design

An item in the auditory perception test contained the comparison of two naturally spoken two-syllable stress-groups or the comparison of a naturally spoken stress-group and a low-frequency shadow of a stress-group. The low-frequency shadow of a stress-group was obtained using a 400 Hz low-pass filter. This method was used to allow the listener to abstract from the segmental content of the words, and focus only on assumed rhythmic factors. The intensity of filtered items was increased by 5 dB to achieve perceptual comparability – because of the range of the hearing field, the items with higher

frequencies are perceived as louder (Palková, 1994: 95–96). Some items, including both fully pronounced stress-groups and low-frequency shadows, were duplicated and manually smoothed using PSOLA (Pitch Synchronous Overlap and Add) technique, but this manipulation was required not to interfere with the impression of naturalness of speech. Items with smoothed intonation were added to eventually verify previously discovered findings that the variable or higher *F*0 contributes to the perception of sound duration as longer than duration of the same sound with balanced fundamental frequency (Donovan & Darwin, 1979; Brigner, 1988; Cumming, 2011; Šimko et al., 2015; Dawson et al., 2017).

Within one item, a silent pause of 750 ms was inserted between the stress-groups or a stress-group and the low-frequency shadow of a stress-group; a second pause lasting 1.5 seconds was added after the second stress-group or the shadow of a stress-group followed by a desensitisation sound and a silent pause (1.5 s). One test item lasted about 7 seconds.

A total of 210 items were pseudo-randomly sorted three times for three variants of the auditory perception test and then divided into three blocks of 70 items. The tests, together with the forms, were passed to 40 respondents, native speakers of Czech. The listeners then estimated on a five-point scale if the sounds in each item were rhythmically almost identical (= 1), rather similar (= 2), neither similar nor different (= 3), rather different (= 4), completely different (= 5).

3. Results

3.1. Features of the experimental design

Some of the items contained a stress-group and its own low-frequency shadow (see Section 2.2). These items were included in the test to give the respondents a reference point of rhythmic similarity. Since these items were specific (identical rhythm and speaker), all the general analyses in this chapter were performed both on the evaluation of the entire set of items and on the evaluation of the set of items that did not contain rhythmically identical stress-groups within an item.

There was a significant effect of the presence of identical stress-groups within an item (t-test: $t(221) = 11.9$; $p < 0.001$). While evaluations of items that contained different stress-groups (either in full form or as a low-frequency shadow), they were around the average of 2.88, the rating of rhythmically identical items was averaging at 1.64.

Another factor in the item evaluation was the presence of a low-frequency shadow. One item could contain either two different stress-groups (e.g., [bɪlɪ] and [budɛ]), or one stress-group and one low-frequency shadow of another or the same stress-group (e.g., the stress-group [mɛzɪ] and the shadow of the stress-group [ɲɪmɪ]). T-tests showed that listeners perceived the items with a low-frequency shadow as more similar than the others. However, this relationship was only reflected in the entire set of items: ($t(221) = -2.36$; $p < 0.05$). The question is whether rhythm perception was influenced by the presence or absence of segmental content of the second stress-group, or by a general evaluation of identical stress-groups within an item as significantly more rhythmically similar than others.

3.2. Phonotactic features

The results of t-tests of the items, which varied only by the **number of consonants**, showed a statistically significant effect of the presence vs. absence of a consonant cluster on the perception of rhythmic similarity. Stress-groups with a different number of consonants in a cluster within one item (e.g., CVCV and CCVCV – [fa:zɛ] and [sta:lɛ]) were rated as rhythmically less similar than the stress-groups in items that did not include this difference. The significance was more robust when the entire set of items was analysed, $t(221) = -2.56$; $p < 0.05$; in the case of omission of the items containing identical stress-groups only an insignificant trend of a similar direction was visible: $t(182) = -0.5$; $p \doteq 0.617$.

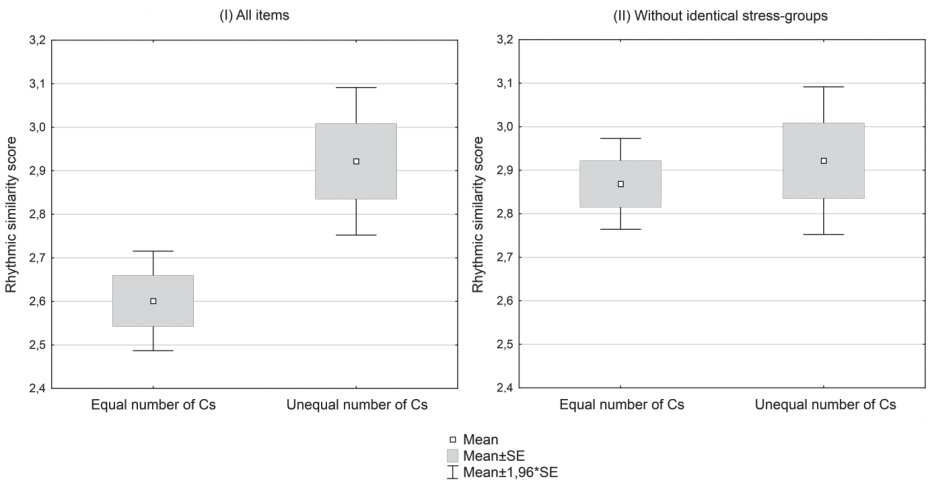

Figure 1 (I, II). A box plot of the influence of the difference in the number of consonants in the consonant cluster on the average evaluation of items. Figure I includes all data; Figure II excludes the items with rhythmically identical stress-groups. SE = standard error; rhythmic similarity score on y axis: 1 = very similar, 5 = very different.

The items consisting of stress-groups with different **position of a consonant cluster** (e.g., CVCCV and CCVCV – [volbɪ] and [vjɛlɪ]) were rated as less rhythmically similar than the items in which this difference did not appear. This effect was evident both in the analysis of all items ($t(221) = -5.96$, $p < 0.001$) and in the analysis of the set of items that did not include items contained a stress-group with its own low-frequency shadow: $t(182) = -4.54$; $p < 0.001$. Compared to the previous variable (the number of consonants in the consonant cluster), the effect of the consonant cluster position was clearly stronger.

Another feature with the presumed influence on the perception of rhythmic similarity of stress-groups was the **presence of a long vowel** in the compared stress-groups. If there was a long vowel in one item, it was always in the same syllable in both stress-groups compared (e.g., [sta:tɪ] and [svɛ:ɦo] with a long vowel in the first syllable or [budɛ] and [ɲɪmɪ] with all vowels short). If a long vowel was in just one of the stress-groups, it would

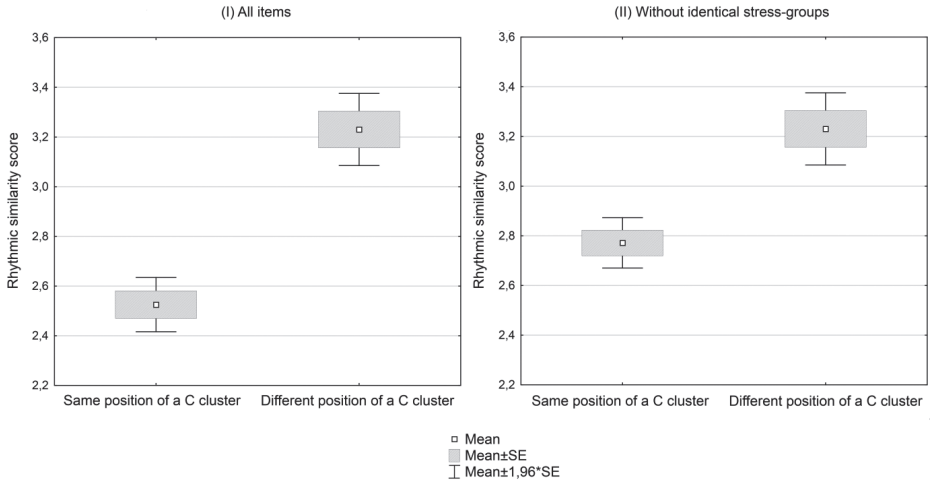

Figure 2 (I, II). A box plot of the influence of the difference in the position of a consonant cluster on the average evaluation of items. Figure I includes all data; Figure II excludes the items with rhythmically identical stress-groups.

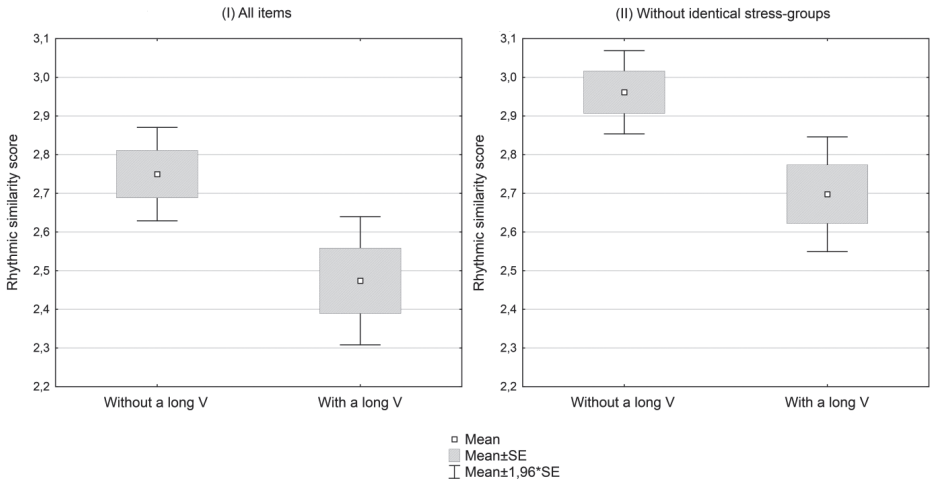

Figure 3 (I, II). A box plot of the influence of the presence of the long vowel in both stress-groups within one item on the average evaluation of items. Figure I includes all data; Figure II excludes the items with rhythmically identical stress-groups.

be difficult to trace another influence on the perception of the speech rhythm than the difference in the presence of a long vowel: the listener's attention would be fixed to the difference in the length of the vowel, and subtle effects on speech rhythm perception may therefore be suppressed.

According to the result of the t-test, stress-groups containing only short vowels were evaluated by listeners as less rhythmically similar than those with long vowels. This find-

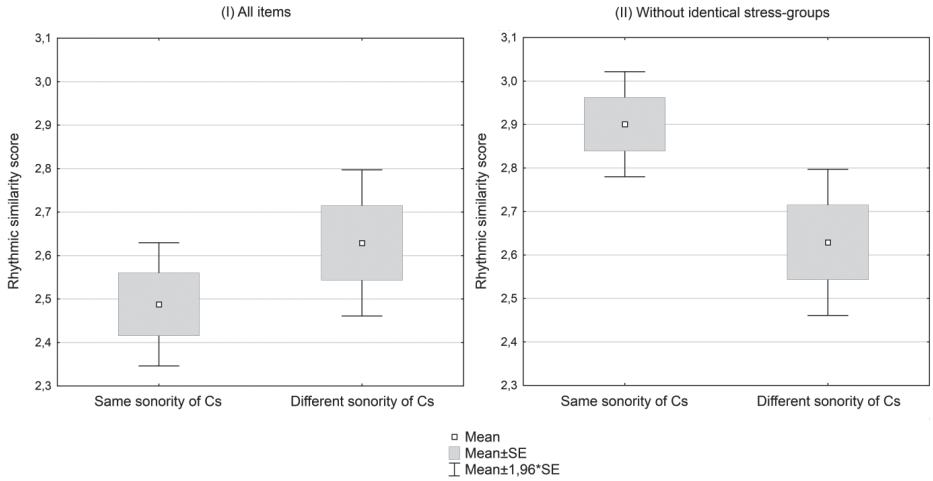

Figure 4 (I, II). A box plot of the influence of the difference in the sonority of voiced consonants on the average evaluation of items. Figure I includes all data; Figure II excludes the items with rhythmically identical stress-groups.

ing was confirmed both in the analysis of all items ($t(221) = -2.55$; $p < 0.05$) and for the set of items without those containing identical stress-groups: $t(182) = -2.72$; $p < 0.01$.

The analysis of the effect of the difference in **sonority of consonants** on the entire set of items showed an inconclusive trend when items containing consonants of different sonority (voiced obstruents and sonorants, e.g., [z**b**oru] and [v**j**ɛlɪ]) were evaluated less rhythmically similar than the other items: $t(173) = -1.17$; $p \doteq 0.244$. In the analysis that did not contain items with identical stress-groups, a reverse trend appeared – the items differing in sonority of consonants were rated rhythmically less different than others: $t(134) = 2.64$; $p < 0.01$.

This result may appear surprising, but considering the design of the experiment, the explanation is quite simple: after omitting the items with identical stress-groups, only items with distinctions in other features remained, some of which may be considered stronger than subtle differences in sonority of consonants (e.g., the number of consonants in a cluster, the position of a consonant cluster). To examine the effect of the difference in sonority, an experiment would be required focused specifically on this feature.

The analysis of the perception of rhythmic similarity also did not show any significant effect of the distinct **voicing characteristic** of obstruents within a pair of stress-groups in one item (e.g., [**s**o͡u to]–[ʒa:t͡sɪ] or [mɛ**z**ɪ]–[ja**k**o]). Trends were inconclusive both for the analysis of all items ($t(221) = 1.31$; $p \doteq 0.192$), and for the analysis excluding items with identical stress-groups ($t(182) = -0.88$; $p \doteq 0.381$). However, a clear conclusion should be drawn from studies focusing exclusively on this feature.

It is worth noting that the phonological voicing feature may be indicated by factors other than the presence of the fundamental frequency. For paired consonants, the duration of the voiceless phones is longer than the duration of their voiced counterparts; some languages (e.g., English) use temporal compensations: if a voiced consonant fol-

lows a vowel, the duration of the vowel is longer than the duration of a vowel preceding a voiceless consonant (Kent & Read, 2002). Although the influence of temporal compensations in Czech has not been systematically investigated, the results of Borovičková & Maláč (1967) and Machač (2006) suggested some tendencies in a similar direction. The end part of a vowel preceding a consonant appears to be a stronger clue in determining phonological voicing feature (e.g., Hogan & Rozsypal, 1980). Speakers may indicate voicing by the presence of a fundamental frequency during the occlusion, by duration of the occlusion, by intensity of the release burst or aspiration (as in English), but also by the frequency of the fundamental tone and the first formant: The results of Hogan & Rozsypal (1980) or Castleman & Diehl (1996) showed that the fundamental frequency and the first formant of the vowel-consonant divide were lower if the vowel was followed by a voiced stop, as opposed to the case when the vowel preceded a voiceless stop. All these features can be perceived and used by the listener whenever it is needed, but it is also possible that the individual features contributing to the perception of phonological voicing form a hierarchy in which temporal aspects are perceived only if other clues are absent or ambiguous (Kent & Read, 2002).

3.3. Suprasegmental and biological factors

The stress-groups in the perceptual test items should not contain any significant intonation differences, so that the melody changes would not interfere with the evaluation of the rhythmic similarity. However, some of the items were included in the test with the original intonation pattern, while others with *F*0 smoothed. T-tests did not reveal any evidence of the influence of smoothed or original intonation on the perception of stress-groups as rhythmically similar. The analysis performed on all items showed only the marginal significance of the variable ($t(221) = 1.91$; $p \doteq 0.057$), when smoothed items were on average rated rhythmically more similar than others. In addition, a set of items excluding items with identical stress-groups was also analysed ($t(182) = 1.51$; $p \doteq 0.133$), as well as only the items containing smoothed intonation along with their original counterparts, again with and without the items which included identical stress-groups ($t(52) = -0.23$; $p \doteq 0.822$; $t(38) = 0.01$; $p \doteq 0.99$ respectively).

Another feature examined was the possible influence of different sexes of the speakers within one item. A significant difference only showed in the analysis of the full set of items: $t(221) = 3.14$; $p < 0.01$; the stress-groups within items in which the sexes of the speakers matched were rated more rhythmically similar. A trend of the same direction emerged in the analysis without the items containing identical stress-groups, but it was statistically insignificant: $t(182) = 0.16$; $p \doteq 0.876$. It is therefore possible to conclude that the perception of rhythmic similarity was influenced by the individuality of the speaker rather than the gender.

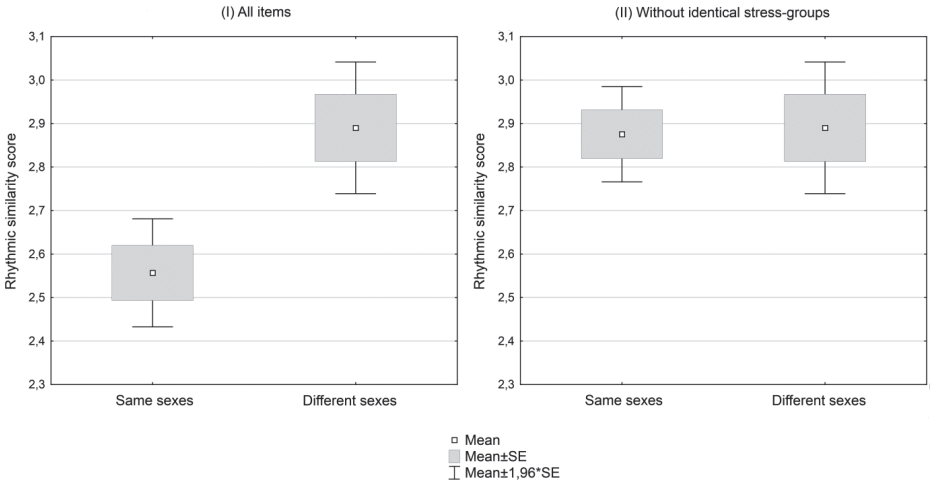

Figure 5 (I, II). A box plot of the influence of the difference in the sexes of the speakers within one item on the average evaluation of items. Figure I includes all data; Figure II excludes the items with rhythmically identical stress-groups.

Table 1. Comparison of the strength of individual influences on the perception of rhythmic similarity by comparison of the size of the test criterion *t*. Statistically significant results are bold.

feature	*t* (all items)	*t* (without items containing identical stress-groups)
match of the low-frequency shadow	**11.9**	
presence of the low-frequency shadow	**−2.36**	1.68
number of consonants in the cluster	**−2.56**	0.5
position of the consonant cluster	**−5.96**	**−4.54**
presence of a long vowel	**−2.55**	**−2.72**
difference in the sex of the speakers	**3.14**	0.16
difference in sonority of voiced consonants	−1.17	**2.64**
difference in voicing of obstruents	1.31	−0.88
smoothing of *F*0	1.91	1.51

4. Discussion

The main objective of the research was to find out the relations between the phonotactic structure of the Czech stress-group and the perceived rhythmicity of the speech unit. Since rhythm is a phenomenon that has its basis in listener's perception (Lehiste, 1977, 1979; Morton et al., 1976; Fletcher, 2010), an auditory assessment experiment was designed to establish this relationship. Šturm & Volín (2016) showed that the phonotactic

structure of the syllable played a significant role in perceiving the rhythm of Czech, and the experiment in the present study was to determine the degree of the influence of the individual phonotactic features on perception of speech rhythm.

The results showed that from the observed phonotactic factors the position of the consonant cluster in the stress-group had the strongest influence on evaluating rhythmic similarity. If the stress-groups within the item differed in this characteristic (i.e., the comparison of the stress-groups with the consonantal-vocalic patterns CCVCV and CVCCV), they were consistently evaluated by the listeners as rhythmically rather dissimilar. A slightly less strong but still relatively robust factor of speech rhythm perception was also the number of consonants in a consonant cluster – the stress-groups corresponding to the CVCV pattern were perceived by the listeners as rhythmically different from CCVCV or CVCCV units. Another important feature in evaluating rhythmic similarity was the quantity of a vowel. The respondents considered items in which a long vowel appeared in the same syllable in both stress-groups rhythmically more similar than the stress-groups with all vowels short. This trend also reached statistical significance in the evaluation of the experiment.

Although some factors proved to be almost or totally insignificant, it is impossible to state unequivocally that they do not participate in the perception of rhythmic similarity. As mentioned above, the difference in sonority of voiced consonants and in voicing of obstruents can be considered as more subtle a feature in the phonotactic structure of stress-groups, and if the significance of these contrasts is tested against items that contain differences in more robust factors (such as the position of a consonant cluster within a stress-group), the influence of the difference only in sonority or voicing of the consonants will not prove to be relevant. Whether factors such as contrast of voicing or sonority really do not participate in the perception of speech rhythm can be verified in future research focused solely on these two features. Such an experiment would include the pairs of stress-groups corresponding to the most common consonantal-vocalic patterns, which would differ either in voicing or in the sonority of consonants.

5. Conclusion

In terms of the phonotactic factors, the position of a consonant cluster had the strongest influence on the perception of rhythmic similarity of Czech stress-groups. The results showed that the number of consonants in each stress-group and the presence of a long vowel also contributed to the evaluation of stress-groups as rhythmically similar or different. The features as sonority of consonants and voicing of obstruents had only minimal or inconclusive effect on the perception of the rhythmic similarity of the stress-groups, but it is not possible to state without further examination that these are redundant for the perception of the speech rhythm. To capture the possible influence of these factors, a similar experiment would be required, this time only focused on the above-mentioned features.

ACKNOWLEDGEMENTS

The author would like to express her thanks to doc. PhDr. Jan Volín, Ph.D., the supervisor of her PhD thesis, from which this study arose. Thanks are also due to the 40 listeners who participated in the auditory perception experiment.

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer* [computer program], version 6.0.30. Retrieved from http://www.praat.org.

Borovičková, B. & Maláč, M. (1967). *The Spectral Analysis of Czech Sound Combinations.* Prague: Academia.

Brigner, W. L. (1988). Perceived duration as a function of pitch. *Perceptual and Motor Skills*, 67, 301–302.

Castleman, W. A. & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics, 24*(4), 383–398.

Cumming, R. (2011). The effect of dynamic fundamental frequency on the perception of duration. *Journal of Phonetics*, 39, 375–387.

Dankovičová, J. & Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. In: *Proceedings of the XVIth ICPhS*, 1241–1244. Saarbrücken: Organizing Committee.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of the International Phonetic Association,* 11, 51–62.

Dawson, C., Aalto, D., Šimko, J. & Vainio, M. (2017). The influence of fundamental frequency on perceived duration in spectrally comparable sounds. *PeerJ*, 5.

Donovan, A. & Darwin, C. J. (1979). The perceived rhythm of speech. In: Fischer-Jørgensen, E. et al. (Eds.), *Proceedings of IXth ICPhS*, 268–274. Copenhagen: University of Copenhagen.

Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In: Hardcastle, W., Laver, J. & Gibbon, F. (Eds.), *The Handbook of Phonetic Sciences*, 523–602. United Kingdom: Wiley-Blackwell Publishing.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In: Warner, N. & Gussenhoven, C. (Eds.), *Papers in laboratory phonology 7*, 515–546. Berlin: Mouton de Gruyter.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31(3–4), 423–445.

Hogan, J. T. & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America,* 67, 1764–1771.

Churaňová, E. (2012). *Fonotaktická osnova českého slova a mluvního taktu.* Master's thesis. Prague: Charles University.

Churaňová, E. (2013). The consonantal-vocalic structure of the Czech word and stress group. *AUC Philologica 1/2014, Phonetica Pragensia XIII*, 79–90.

Kelso, J. A. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior.* Cambridge: MIT Press.

Kent, R. D. & Read, C. (2002). *The acoustic analysis of speech*. Australia: Singular/Thomson Learning.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics,* 5, 253–263.

Lehiste, I. (1979). Temporal relations within speech units. In: Fischer-Jørgensen, E. et al. (Eds.). *Proceedings of IXth ICPhS*, 241–244. Copenhagen: University of Copenhagen.

Machač, P. (2006). *Temporální a spektrální struktura českých explozív*. PhD thesis. Prague: Charles University.

Morton, J., Marcus, M. & Frankish, C. (1976). Perceptual centres (P-centres). *Psychological Review*, 83, 405–408.

Palková, Z. (1994). *Fonetika a fonologie češtiny*. Prague: Karolinum.

Pike, K. L. (1945). *The intonation of American English*. University of Michigan Press: Ann Arbor.
Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-Based Phonetic Segmentation in Praat Environment. In: *Proceedings of the XIIth International Conference "Speech and computer – SPECOM 2007"*, 537–541.
Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292.
Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969–992.
Skarnitzl, R. (2010). Prague Phonetic Corpus: status report. *AUC Philologica 1/2009, Phonetica Pragensia XII*, 65–67.
Šimko, J., Aalto, D., Lippus, P., Włodarczak, M. & Vainio, M. (2015). Pitch, perceived duration and auditory biases: comparison among languages. In: *Proceedings of the 18th ICPhS*. Glasgow: University of Glasgow.
Šturm, P. & Volín, J. (2016). P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment. *Journal of Phonetics,* 55, 38–52.

RESUMÉ

Tato studie si klade za cíl prozkoumat vztahy mezi fonotaktickou stavbou českého mluvního taktu a řečovým rytmem. Pro účely této práce byly vybrány takty, jež odpovídaly třem nejčastějším dvouslabičným konsonanticko-vokalickým (CV) vzorcům v češtině: CVCV, CCVCV a CVCCV. V percepčním experimentu, který zahrnoval srovnávání dvou plně proslovených mluvních taktů nebo taktu a nízkofrekvenčního obrazu jiného či stejného taktu, posluchači určovali, nakolik jsou si takty různých i stejných vzorců rytmicky podobné.

Výsledky ukázaly, že nejsilnější vliv na vnímání rytmické podobnosti má z fonotaktických faktorů pozice souhláskového shluku v mluvním taktu. O něco méně silnými faktory pro percepci řečového rytmu byly počet souhlásek v konsonantickém shluku a přítomnost dlouhého vokálu v obou porovnávaných taktech. Pro subtilnější rysy, jejichž význam pro percepci rytmické podobnosti prokázán nebyl (např. rozdíl v sonoritě znělých souhlásek či znělosti obstruentů), byly navrženy možnosti dalšího zkoumání.

Eliška Churaňová
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: eliska.churanova@gmail.com

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 109–128
PHILOLOGICA 2/ PHONETICA PRAGENSIA

PHONETIC ASPECTS OF STRONGLY-ACCENTED CZECH SPEAKERS OF ENGLISH

RADEK SKARNITZL and JANA RUMLOVÁ

ABSTRACT

This paper contributes to the study of Czech-accented English by examining multiple pronunciation features, both segmental and prosodic, typically associated with or previously studied in Czech English. We analyzed ten female speakers who had been evaluated as having a strong accent in their English, using auditory and acoustic approaches. In the segmental domain, most of the analyzed speakers used Czech equivalents of the English open vowels /æ ɒ/ and tended to pronounce a velar plosive after a velar nasal. In the domain of connected speech, linking was very rare in our speakers, and their pitch range tended to be very flat. The results underscore the fact that the label "strong Czech accent" may, in different speakers, refer to different constellations of pronunciation features.

Key words: foreign accent, pronunciation, second language acquisition, L1 transfer, Czech English

1. Introduction

In the last few decades, English has become the dominant language of international communication, with more non-native speakers using English today than native ones (Crystal, 2002: 10). The inevitable outcome of English being used as an international language (EIL) by speakers of different origins and mother tongues (L1) is that one frequently encounters non-native, or foreign accents. In other words, one commonly hears English spoken with pronunciation patterns which deviate, in terms of their segmental or prosodic properties, from those found in the speech of native speakers.

Foreign-accented speech can be described in terms of several dimensions. The traditional approach focuses on the above-mentioned deviations from native-like pronunciation: *accentedness* refers to the overall strength of these deviations. It soon became clear, however, that not all pronunciation deviations from L1 are "made equal": their consequences for the success and smooth flow of the communication process vary widely. That is why other dimensions of accented speech have been proposed: *intelligibility* and *comprehensibility* have been shown to be only partially related to *accentedness* (Munro & Derwing, 1995). The authors demonstrated that even strongly accented speech can be fully intelligible; in other words, listeners may be able to understand the message completely. *Comprehensibility* refers to the subjective ease of processing of foreign-accented speech:

https://doi.org/10.14712/24646830.2019.21

while we may be able to understand a speaker's message, this may be only at the price of high cognitive effort. While *intelligibility* (an indicator of objective understanding) and *comprehensibility* (subjective understanding) are clearly related, they do represent different constructs (see also Derwing & Munro, 2009); it is interesting to point out that it has already been 70 years since these two constructs were treated jointly as *comfortable intelligibility* (Abercrombie, 1949: 120). The more accurate description of pronunciation constructs is associated with a re-evaluation of aims in pronunciation teaching: the earlier Nativeness Principle has been replaced by the Intelligibility Principle (Levis, 2005). As a result, researchers have been attempting to identify those features of pronunciation which have the greatest impact on intelligibility; an excellent recent survey of these endeavours can be found in Levis (2018).

This study will examine English as a foreign language (L2) pronounced by native speakers of Czech. However, its objective is not to examine a particular pronunciation feature with respect to intelligibility or comprehensibility. Rather, we aim to analyze the pronunciation of strongly-accented speakers of Czech English (Skarnitzl, Volín & Drenková, 2005) and identify which non-native features are most clearly associated with their speech. Naturally, the English pronounced by Czech speakers is not a new research objective. Nevertheless, previous studies have typically addressed one particular pronunciation feature or a group of features, as described in the following section. The aim of the present study is to provide a more global analysis of Czech English.

2. The study of Czech-accented English

In this section, we will briefly compare the sound patterns of English and Czech, focusing on those which are known to cause problems to Czech speakers of English, and introduce studies which have examined various aspects of Czech English. Segmental properties will be addressed first, followed by prosodic ones.

2.1. Vowels in Czech English

The English vocalic system is considerably more complex than the Czech one, as shown in the schematic comparison of the monophthongs of British English and Czech in Figure 1. There are two major differences between the two systems. First, vowel length is distinctive in Czech, and for three of the pairs (the non-high vowels) the quality of the short and long vowel is the same. There is a qualitative difference between the short and long high front vowels, and a similar difference is emerging in the high back vowels (Skarnitzl & Volín, 2012). In English, length is traditionally marked in the tense vowels but, in fact, length itself is not distinctive. Second, English has more vowels in its inventory than Czech, and the discrepancy is visible especially in the open region.

It is indeed the open region which causes most problems for Czech learners of English: while Czech has only one vowel pair /a/–/aː/ in the entire open region, there are four vowels occupying this space in English, /æ ʌ ɑː ɒ/. Notable among these is the open front vowel /æ/ which, to our knowledge, is the only one examined from the perspective of production or perception by Czech learners. A part of the title of Šimáčková's (2003)

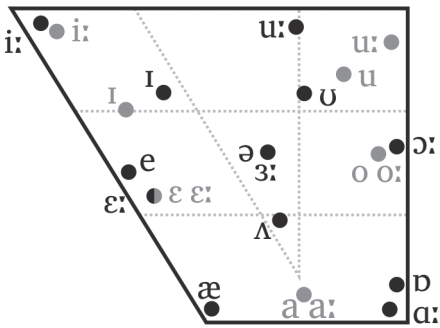

Figure 1. Schematic comparison of the British English (in black) and Czech (in grey) systems of monophthongs. Based on Hawkins and Midgley (2005) for English, and Skarnitzl and Volín (2012) for Czech.

study, "Engela's Eshes", reflects the most frequent realization of Czech speakers of English: the open front /æ/ is typically pronounced as a closer (mid to open-mid) vowel, [ɛ]. The contrast with the short /e/ is then achieved by means of duration, with the phrase *bad bed* typically pronounced [bɛːd bɛd] by Czech learners. Šimáčková found that Czech learners rely predominantly on duration when deciding between the English /æ/ and /e/, and that their production of the two vowels spectrally overlap. Similar results were obtained by Šimáčková and Podlipský (2018): even highly proficient Czech speakers of English used duration to contrast the two vowels, with vowel height being less reliable. Šturm and Skarnitzl (2011) studied perceptual aspects of the vowel /æ/ by two groups of listeners: students who had been instructed in the phonetics and phonology of English and naïve students with no such formal instruction. Their results show that the former group's judgements correlated with the openness of the vowels, as reflected in the value of their F1. In this sense, the instructed group's assessment of Czech speakers' renditions of /æ/ may be regarded as more accurate.

Although the perception or production of other English vowels by L1 Czech listeners have not been studied, we may predict a similar process of equivalence classification (Flege, 1987) with other English vowels in the open region. The short open /ɒ/ is likely to be qualitatively equated with the Czech mid /o/, and the pair /ʌ/–/ɑː/ with the Czech open central vowels /a/–/aː/, respectively. However, such pronunciation is likely to impact intelligibility and comprehensibility considerably less than in case of the open front vowel /æ/.

As shown in Figure 1, vowels in the non-high regions are not realized identically in the two languages. Nevertheless, the differences are of a phonetic rather than phonological nature (for instance, the present-day long /uː/ of British English is pronounced as nearly a central vowel, compared to the Czech back /uː/), and the impact on understanding will probably not be dramatic. The mid central vowel /ə/ will be addressed below in sections 2.3 and 2.4, since it constitutes more of a prosodic feature of English than a segmental one.

Finally, it remains to be pointed out that English and Czech also differ in their diphthongs: Czech has three diphthongs closing towards [u], while English has in total seven diphthongs which close towards [ʊ] and [ɪ] and also (in non-rhotic varieties) target the centre of the vocalic space, [ə]. However, in spite of these differences, English diphthongs do not seem to constitute a major problem for native speakers of Czech: they may con-

tribute to a speaker's accentedness, but most likely will not impede intelligibility or comprehensibility.

2.2. Consonants in Czech English

The English consonantal system as such is, in comparison with the vocalic one, not too complex. There are certain consonants, though, which are difficult for speakers of other languages. The dental fricatives, /θ/ and /ð/, are especially notorious. Although their functional load is rather low (i.e., they do not participate in many minimal pairs; see for example Derwing & Munro, 2015: 74f.) and their incorrect pronunciation did not negatively impact intelligibility (Munro & Derwing, 2006), there are several reasons why especially the voiced /ð/ should be an important sound for learners of English. It is the sixth most frequent phoneme in connected speech, 11 of the 100 most frequent English words (especially grammatical ones like *the*, *with*, *they*) contain /ð/, and some alternative pronunciations for both the voiced /ð/ and the voiceless /θ/ are stigmatized throughout the English speaking world (Brown, 2016).

Interestingly, to our knowledge, only various BA- or MA-level theses seem to have dealt with the pronunciation of the dental fricatives by Czech learners of English (e.g. Skarnitzl, 2001). The pronunciation of the voiceless /θ/ is typically reported as [f] or [s], less frequently as [t], while that of the voiced /ð/ is given as [d] or [z], rarely also as [d͡z].

Another English consonant whose difficulty is shared by speakers of more languages is the labiovelar approximant /w/. Based on informal observations, Czech speakers are known to realize this sound as a fricative [v] (e.g., *which* as [vɪt͡ʃ]), but they may also pronounce the English /v/ as an approximant [w] (e.g., *very* as [werɪ]).

Some consonants function differently in the system of the two languages. While both Czech and English have the velar nasal [ŋ], it has a distinctive, phonemic function in English (e.g., *sin* /sɪn/ vs. *sing* /sɪŋ/) but only appears in the context of place assimilation in Czech (e.g., *banka* [baŋka]). For that reason, Czech speakers of English often pronounce [ŋ] with a following plosive sound (e.g., *singing* [sɪŋgɪŋk]; see Skarnitzl, 2004).

Moreover, Czech and English have a different way of implementing the voicing contrast. In Czech, the property distinguishing between /p/ and /b/ or /s/ and /z/ is phonetic voicing. In contrast, English makes use of the tenseness contrast, which is salient especially in plosives: phonologically voiceless plosives are pronounced as aspirated in stressed positions (e.g., *Peter* [pʰi:tə]). Pospíšilová's (2011) analysis showed that even relatively advanced speakers, with no explicit instruction in the sound patterns of English, aspirate significantly less (i.e., produce shorter voice-onset-time values) than after having received instruction in phonetics and phonology.

Skarnitzl and Šturm (2017) focused on the assimilation of voicing in Czech (and also Slovak) speakers of English across the word boundary. They found that both more and less accented speakers tend to assimilate voiceless consonants to the following voiced one (e.g., *nice day* as [naɪz deɪ]) to a similar extent, but that the more accented group devoiced phonologically voiced consonants more in pre-sonorant contexts (e.g., *phase one* as [feɪs wʌn]).

Finally, English is rather untypical in that it preserves the voicing contrast also in the final position (e.g., *dock* /dɒk/ vs. *dog* /dɒg/); in Czech the voicing contrast is neutralized

(e.g., *spát* and *spád* will both be /spa:t/). In English, the contrast is not achieved through phonetic voicing but uses duration: the vowel will be significantly shorter before voiceless consonants (in *dock*) than before voiced ones (in *dog*). Not surprisingly, Skarnitzl and Šturm (2016) found that Czech speakers, who had a relatively strong accent in their English, did not exploit duration to cue this contrast.

2.3. Lexical stress in Czech English

The two languages whose interactions will be examined in this study differ in the realization of lexical stress. Czech is a language with stress fixed on the first syllable of the prosodic word and serving only a delimitative function, while stress is contrastive in English and stress placement rules are very complicated. The stressed syllable does not bear any marks of positive prominence in Czech (Skarnitzl, 2018); in fact, some studies suggest that the second syllable is frequently pronounced with higher fundamental frequency (f_0) than the stressed one (Palková & Volín, 2003; Volín, 2008). In English, lexical stress is manifested through longer duration, flatter spectral slope and also higher fundamental frequency (Eriksson & Heldner, 2015).

Learning the English stress patterns involves not only the placement and adequate acoustic realization of the stressed syllable but also, and perhaps more importantly, mastering the quality of the unstressed syllables. Unstressed syllables tend to be reduced in English; this reduction includes shorter duration, centralization towards the mid central vowel *schwa* /ə/ (as in *together* /təˈgeðə/), and steeper spectral slope. It is this aspect of English which has received most attention in studies of Czech speakers. Volín, Weingartová & Skarnitzl (2013) compared the spectral properties of *schwa* in native British and Czech speakers. While the Czech speakers' formant values did not significantly differ from the native speakers' pronunciation (in other words, vowel quality was comparable to a *schwa*), the Czech-accented *schwas* were still too prominent, as reflected in narrower formant bandwidths and flatter spectral slopes. In a follow-up study, more advanced Czech speakers of English were shown to approximate native durational and spectral patterns more than less advanced speakers (Weingartová, Poesová & Volín, 2014). Similar results were reported by Poesová and Weingartová (2018).

The reduced vowel *schwa* occurs not only in unstressed syllables of polysyllabic words but also in weak forms of grammatical words such as *and*, *for*, *that* or *were*. All this contributes to the characteristic rhythm of English, which will be addressed in the following section.

2.4. Aspects of Czech English related to rhythmic patterning

The temporal and qualitative reduction of unstressed vowels and unstressed grammatical words is a major factor which determines the nature of rhythmic patterning of English. Volín and Johaníková (2018) examined the normalized duration of selected grammatical words in their weak forms, as pronounced by L1 speakers of British English and Czech, and found that the Czech speakers of English pronounced these words as significantly longer (i.e., less temporally reduced).

The typical rhythm of English is facilitated by other factors apart from reduction which may be grouped under the heading connected speech processes. The function of these processes in English is "to promote the regularity of English rhythm by compressing syllables between stressed elements and facilitating their articulation" (Alameen & Levis, 2015: 161). Included among these processes are assimilations of articulation place and manner (e.g., *in bed* [ɪmˈbed], *in the* [ɪn̪ˈnə]), coalescence (*did you* [dɪʤə]), consonant-to-vowel and vowel-to-vowel linking (*make it* [meɪkˌɪt], *see it* [siː$_{(j)}$ɪt]), and elision (*did he* [dɪdˌi]).

In their analysis of weak-form word pronunciation by Czech and British speakers, Volín and Johaníková (2018) focused on these processes as well. They found that the Czech speakers linked grammatical words like *a*, *and*, *in*, *of* much less than the native speakers, rarely elided [h] in *have/has* or [r] in *from*. In an earlier study, Bissiri and Volín (2010) found that Czech speakers of English with a strong foreign accent glottalized (i.e., did not link) in more than 75% of all possible instances and that there was little difference within or across phrasal boundaries. Šimáčková, Podlipský and Kolářová (2014) examined linking in advanced speakers of Czech from Moravia (linking is more prevalent in this variety of Czech than in Bohemia) and found that linking occurred between 42 and 64% of the possible instances, with consonant-to-vowel linking being most frequent. In a related study, Šimáčková, Kolářová and Podlipský (2014) found that the tendency of Czech speakers to link increased at higher speech rates.

All the above-mentioned processes, including the reduction of unstressed syllables and words, contribute to the specific rhythm of the English language; it is therefore clear that rhythm is a true product of its phonological and phonetic patterns.

2.5. Intonation in Czech English

Intonational cues may fulfil a number of functions in languages, and languages tend to differ in this respect. This also applies to Czech and English, and the functions are determined, to a large extent, by the rather free word order in Czech and the rather fixed one in English. That is why English relies mainly on melodic cues when expressing prominence, while word order changes may be used alongside or even instead of melodic ones in Czech. In addition, the melody of speech appears to be more important for expressing pragmatic meanings in English (Wichmann, 2005). That may be the reason for the much wider pitch range in English than in Czech, as confirmed by Volín, Poesová and Weingartová (2015). The authors of the study compared Czech and British radio broadcasters and found that pitch range (specifically, the 80-percentile range) was 2 semitones narrower in L1 Czech than in L1 English for both male and female newsreaders. In a following step, native English and Czech non-professional speakers read the same sentences in English. While the pitch range of native British speakers was similar to that of the British newsreaders, it was by over 1 semitone narrower in the L2 speakers of English than in the L1 Czech newsreaders. The results of the study by Volín et al (2015) therefore do not support a straightforward transfer hypothesis, according to which one would predict values intermediate between (or identical to) those of Czech and English. With the pitch range in English as an L2 of Czech speakers even narrower than in L1 Czech, the authors suggest that there must be other factors at play, such as anxiety of the L2 speakers.

3. Method

For this study, we analyzed the speech of ten female speakers. They were native speakers of Czech, and their pronunciation in English was evaluated by three expert phoneticians as strongly accented. The speaker selection method can be justified by our previous study (Skarnitzl et al., 2005) which showed that native English speakers and proficient Czech speakers of English manifest very high correlations when judging the degree of foreign accent. The speakers were asked to read a standard BBC news bulletin; six different texts were used, with an average reading duration of 4 minutes. The recordings were obtained in the sound-treated recording studio of the Institute of Phonetics in Prague, at a sampling rate of 32 kHz and with 16-bit quantization, using the high-quality AKG C4500 B-BC condenser microphone. The speakers were given sufficient time for preparation.

As mentioned at the end of the Introduction, the aim of this exploratory study is to identify which of the features of Czech-accented English discussed in the previous section are most reliable. In other words, we are interested in finding out which of the features appear most frequently. The wide selection of the features necessarily affects the choice of the methodology: since vocalic, consonantal, as well as prosodic features will be analyzed here, no single way of analyzing them is possible. That is why both auditory and acoustic analyses are included in this study.

The pronunciation features examined by means of listening are listed in Table 1, along with the number of items for each feature; naturally, the numbers were constrained by the texts, but we aimed at analyzing at least 10 items per feature per speaker. Selection criteria for some of the features are listed in Table 1 as well. The last column provides details about how the individual features were evaluated. Three features were assessed in a binary way (present or absent). Some segmental features were assessed either in a ternary manner (with 2 corresponding to, for instance, a native-like open [æ], 0 to a completely Czech [ɐ/ɛ], and 1 to an intermediate realization), or we noted the specific realization (e.g., for the vowel *schwa*, we noted the vocalic quality the specific realization was closest to). For lexical stress, we noted the syllable which was stressed in the particular word. The auditory evaluations were entered into a special tier in Praat (Boersma & Weenink, 2018).

The only features which were analyzed acoustically in this study concern melodic patterning. As Czech English has been found to be very flat (see section 2.5), a measure of pitch range was a natural choice for analysis. For this purpose, we split the utterances into breath groups (portions of the speech signal between two intakes of breath). Values of f_0 were extracted using autocorrelation in Praat every 10 ms, the contour was smoothed by a 10-Hz filter, interpolated, and converted into the Praat PitchTier objects where the contours were carefully inspected and manually corrected to reduce the most salient measurement errors, especially octave jumps and spurious f_0 measurements in creaky phonation or voiceless portions of the signal. Finally, the curves were once again interpolated to approximate the perceived pitch contour. From these manually corrected f_0 objects, we calculated the 80-percentile range of each speaker (i.e., a range value where the lower and upper 10% values are ignored).

In addition, we analyzed the difference in mean f_0 in the stressed vowel and the vowel in the following syllable; as Czech speakers have been shown to pronounce the post-stressed syllable higher than the stressed one, the aim was to determine whether they

Table 1. Features analyzed by listening (see text).

	Feature	Items	Selection criteria	Scoring
V	æ	113		2 – 1 – 0
	ɒ	124		2 – 1 – 0
C	ŋ	94		specific realization
	θ	58		specific realization
	ð	159	include as many lexical words as possible	specific realization
	v	162		2 – 1 – 0
	w	212		2 – 1 – 0
	prevocalic ɹ	103	include word-initial and -medial items, as well as those following a plosive	specific realization
	aspiration in p, t, k	307	include stress on the first and other than first syllable, and words with preceding /s/	present – absent
	voicing assimilation	64		present – absent
prosody	lexical stress	342	aim for two-, three- and four-syllabic words with stress on another than the first syllable	stressed syllable
	ə	361	include word-initial, -medial, and -final, in lexical words only	specific realization
	linking	227	include linking to lexical and grammatical words	present – absent

transfer this tendency to their L2 English. The f_0 values were extracted from the manually corrected PitchTiers using a Praat script; only those words were used which were marked for lexical stress (see Table 1).

The auditory data were extracted from the evaluation tier using a Praat script and subsequently processed in the R programme (R Core Team, 2015). The PitchTiers were processed in the *rPraat* package (Bořil & Skarnitzl, 2016). All visualizations were performed in the *ggplot2* package (Wickham, 2009).

4. Results and discussion

The pronunciation of the analyzed speakers will be described in five sections, following the structure of the introduction.

4.1. Vowels

The two target open vowels of English – the front /æ/ and the back /ɒ/ – were expected to be realized as their Czech counterparts, the (open-)mid /ɛ/ and /o/, respectively. Figure 2 shows that this hypothesis is largely confirmed: only three realizations of /ɒ/ and one of /æ/, produced by three different speakers, were evaluated as target-like.

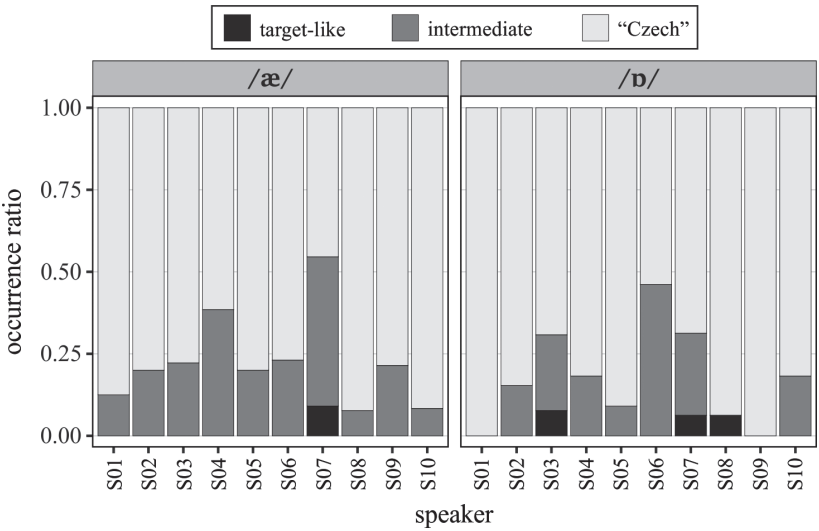

Figure 2. Proportion of target-like, intermediate, and "Czech"-like realizations of the vowels /æ/ and /ɒ/.

4.2. Consonants

We will first focus on those consonantal sounds which do not exist in Czech at all, starting with the dental fricatives. As the two **dental fricatives**, the voiceless /θ/ and the voiced /ð/, occur in different word types, they will be addressed separately. The realizations of the voiceless /θ/ by our speakers are illustrated in Figure 3. First of all, it is clear at first sight that there is much greater variability between speakers: while speakers S01, S08 and S10 pronounced all the target sounds as voiceless dental fricatives, speakers S04, S05 and S06 substituted more than 75% of /θ/-items by different consonants. The most frequent substitute was the plosive [t], with the exception of speaker S05 who used [s] more; a closer examination reveals that the [s] substitutions occur predominantly at the end of words like *death*, *month(s)* or *both*. It is noteworthy that [f], phonetically the closest candidate for a substitute of English /θ/, was only pronounced once by speaker S04 in the word *three*. The affricate [t͡s] was used in similar words (*three*, *thirty*), and the sequence [th] appeared in *authority*, *thousand* and *strengthening*. The voiced [ð] occurred in the context of regressive voicing assimilation, in the phrase *foot and mouth disease*.

The voiced dental fricative /ð/ was pronounced as [ð] in 33% cases and substituted by the plosive [d] in 58% cases. On the one hand, /ð/ is very frequent in grammatical words like *the*, *this* or *than*; on the other hand, it also occurs in lexical words like *father*, *southern* or *together*. As shown in Figure 4, the substitutions of the Czech speakers differ with respect to these categories: [d] occurs with higher frequency in the grammatical words, while [ð] is pronounced in about two thirds of the cases in lexical words. The preposition *with* is depicted separately in the figure, since it manifests specific substitutions [t], [s] and [θ] ([t] was also pronounced once in the words *gathering* and *the*).

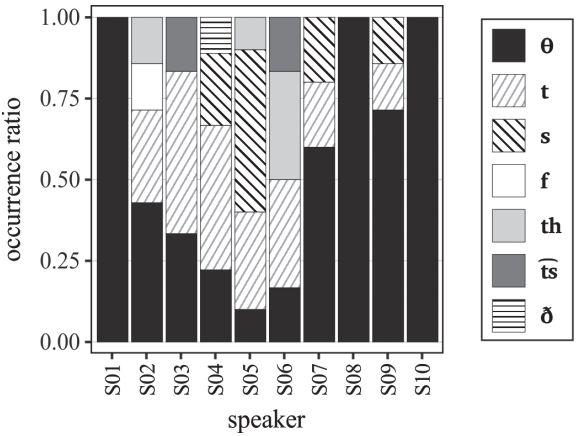

Figure 3. Proportion of realizations of the voiceless dental fricative /θ/.

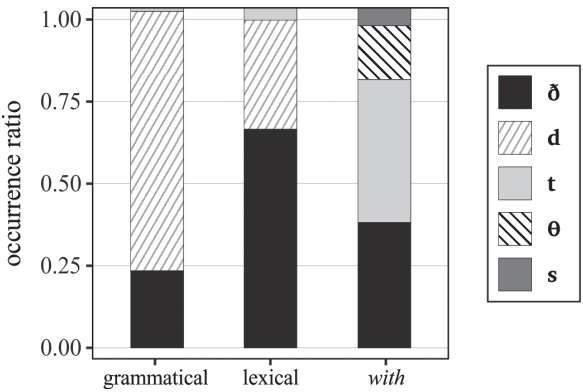

Figure 4. Proportion of realizations of the voiced dental fricative /ð/ according to the lexical class of the word; the preposition *with* is shown separately (see text).

The realizations of /ð/ are depicted for the individual speakers in Figure 5. It is interesting to compare these substitution patterns of /ð/ with those of /θ/ in Figure 3: with the exception of speaker S10, those who tended to pronounce /θ/ as a dental fricative did the same with /ð/, and those who substituted /θ/ by other consonants tended to substitute /ð/ also most. Based on our data, we can thus draw the conclusion that the voiced dental fricative /ð/ is more difficult for Czech speakers of English, especially in grammatical words.

Let us next turn to the English **labiovelar approximant /w/** and the **labiodental fricative /v/**. As mentioned in section 2.2, Czech speakers may pronounce both of them incorrectly. It is obvious from Figure 6 that this is indeed the case, albeit to a much smaller extent than in the case of the dental fricatives. Both /w/ and /v/ were pronounced correctly in over 70% of all items, but individual speakers differ considerably. It seems

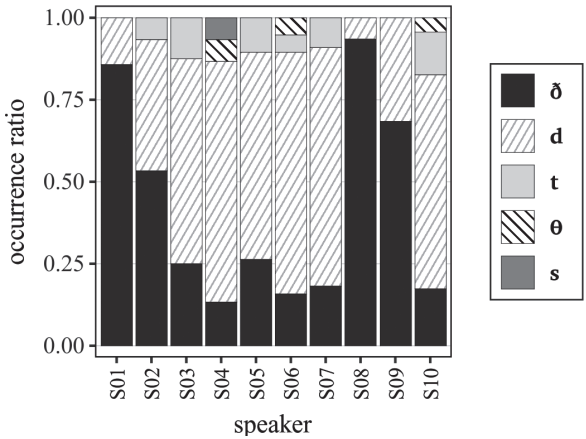

Figure 5. Proportion of realizations of the voiced dental fricative /ð/.

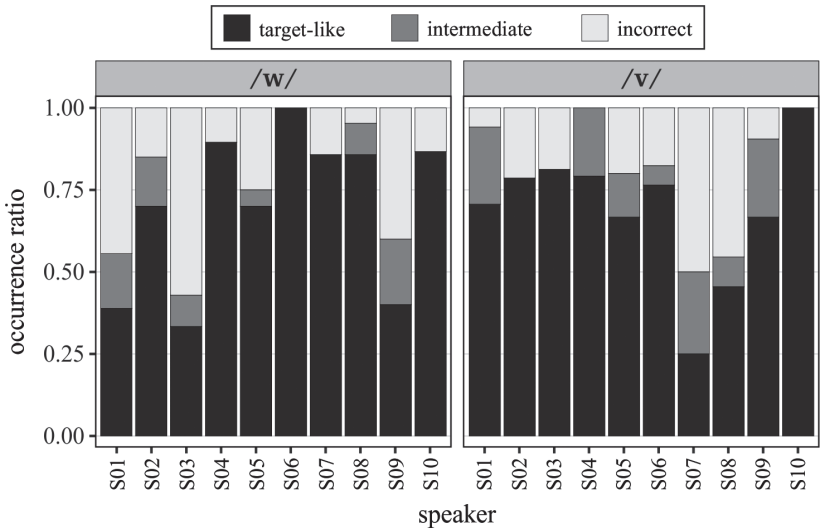

Figure 6. Proportion of target-like, intermediate, and incorrect realizations of /w/ and /v/.

that for some speakers (especially S07 and S08), there may even be partial neutralization, with sounds corresponding to both /w/ and /v/ approximating [ʍ]. In a more detailed analysis, we examined the effect of lexical class on the pronunciation of /w/; unlike in the case of /ð/, /w/ was pronounced more correctly in grammatical words like *was*, *with* or *will* than in lexical ones like *world*, *wide* or *twenty*.

In the following paragraphs, we will present results concerning those consonants which function differently in English, or which are realized with noticeable phonetic differences. The **velar nasal sound /ŋ/**, which functions as an allophonic variant of /n/

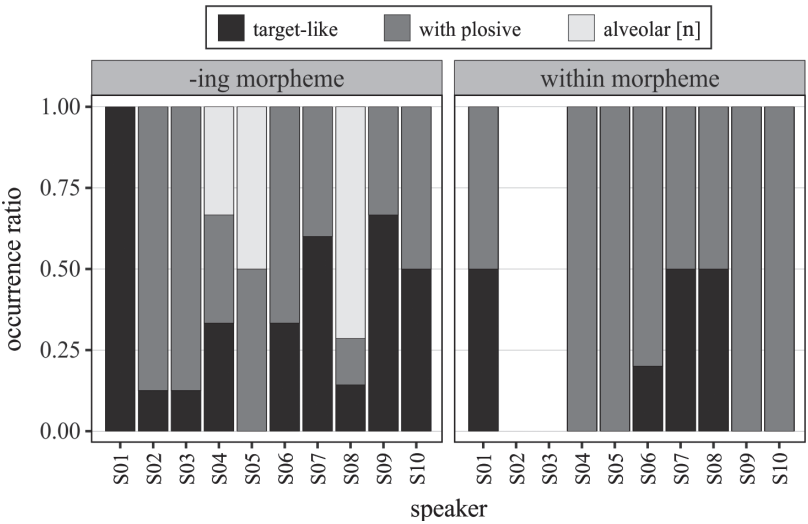

Figure 7. Proportion of realizations of /ŋ/ by the ten speakers in *–ing* morphemes and within morphemes.

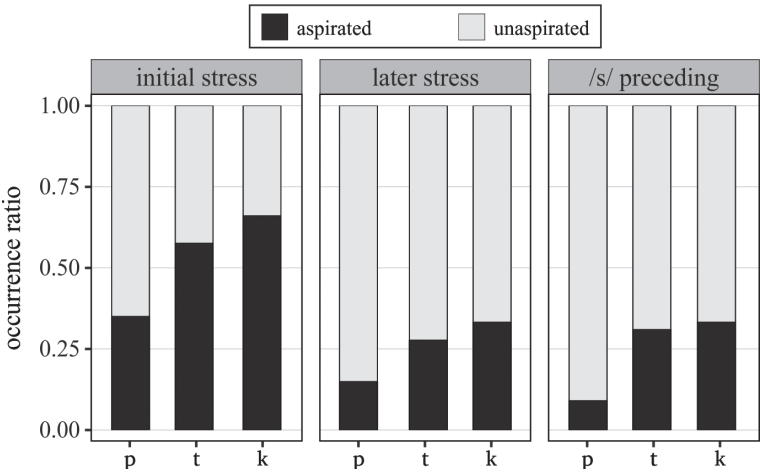

Figure 8. Proportion of aspirated and unaspirated voiceless plosives in stressed syllables at the beginning of the word, later in the word and following a /s/.

in Czech, was pronounced incorrectly in nearly 75% of the cases. As shown in Figure 7, the most frequent incorrect variant was one where the velar nasal [ŋ] was followed by a plosive [k] or [g]. Additionally, three speakers pronounced the alveolar [n] in words containing the progressive form; for speaker S08, this was in fact the most frequent realization of the *–ing* morphemes. Speakers S02 and S03 read the same text which featured no instance of [ŋ] within a morpheme.

We were also interested in the realization of the **/r/ sound**; standard British and American English has the postalveolar or retroflex approximant [ɹ] or [ɻ], while Czech uses the alveolar trill [r]. The trilled pronunciation may be a stereotypical part of the sound of Czech English; however, our results show surprisingly little substitution by a trilled [r]. This was most frequen – in 7 out of 45 cases – when /r/ was preceded by a plosive sound, as in the words *president*, *group* or *hundred*.

The different implementation of the voicing contrast in English and Czech is most salient in **aspiration**. Aspiration tends to be the strongest in the onset of stressed syllables, and that is why aspiration was assessed only in stressed syllables. In Figure 8, we distinguish three contexts: the stressed syllable also being the first syllable, stress on another than the first syllable, and the voiceless plosive preceded by /s/ (there is no aspiration with /s/ preceding the plosive). Not surprisingly, /t/ and /k/ were aspirated more than /p/ (*cf.* Cho & Ladefoged, 1999). More interestingly, the Czech speakers were more likely to aspirate at the beginning of the word (e.g., *parliament*, *territory*, *council*), in 53% of the cases, than when a later syllable was stressed (e.g., *impartial*, *attempt*, *become*), in only 23% of the cases. It is also noteworthy that in 23% of the cases, the speakers aspirated even when a /s/ preceded the voiceless plosive (e.g., *spokesman*, *street*, *escape*), in what may be regarded as overgeneralization.

When analyzing **regressive assimilation of voicing**, it was necessary to exclude all cases where the Czech speakers separated the words by a pause. Approximately 40% of word-final voiceless obstruents, both fricatives and plosives, were assimilated in their voicing when a voiced sound followed (e.g., *West Bank* pronounced [wezd beŋk]). The individual speakers differed in their tendency to assimilate voicing, with speaker S06 assimilating in over 80% of her items and speakers S01 and S03 not assimilating at all.

4.3. Stressed and unstressed syllables

This section will address stress placement in two-, three- and four-syllabic words where stress appears on another than the first syllable, and also the pronunciation of unstressed syllables where vowels correspond to the reduced vowel *schwa* in native English.

Lexical stress was misplaced to the first syllable of the word in about 50% of the cases, regardless of word length. Interestingly, as shown in Figure 9, stress was also misplaced to another (incorrect) syllable in several cases. The word *effort* was twice pronounced [ɪˈfɔːrt]; a similar change occurred in the word *injured*. As for longer words, stress was placed on the last syllable in the words *communiqué* and *communities*.

The **mid central vowel /ə/** was analyzed in longer, autosemantic words like *un<u>a</u>ccept<u>a</u>ble*, *mod<u>er</u>n* or *<u>o</u>ppon<u>e</u>nt*; that is why it is covered alongside lexical stress. As mentioned in Table 1, we were interested in the specific vocalic quality of the sounds which would be pronounced as *schwa* by native speakers. These realizations, as produced by the ten analyzed speakers, are summarized in Figure 10. In total, 37% of the items were pronounced with a mid central quality of a *schwa* [ə] or an *r*-coloured *schwa* [ɚ]; one must keep in mind, however, that this only refers to the quality of the vowel, not to the overall (absence of) prominence (*cf.* Volín et al., 2013 and other studies mentioned in section 2.3). Not surprisingly, the *schwa* vowels were frequently realized with what is known as spelling pronunciation. The most frequent substitutes were the mid front [e/ɛ] (in words

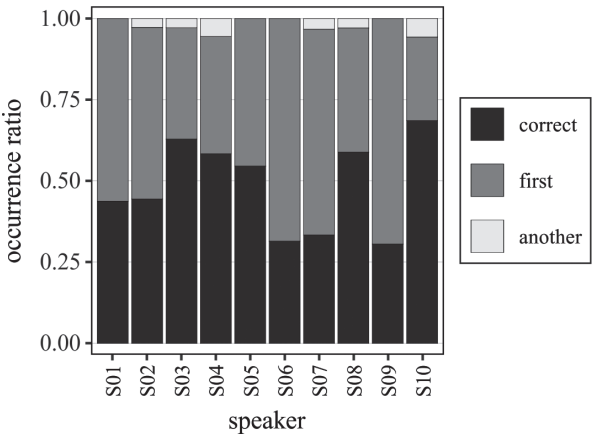

Figure 9. Proportion of words stressed correctly, incorrectly on the first syllable, and incorrectly on another syllable.

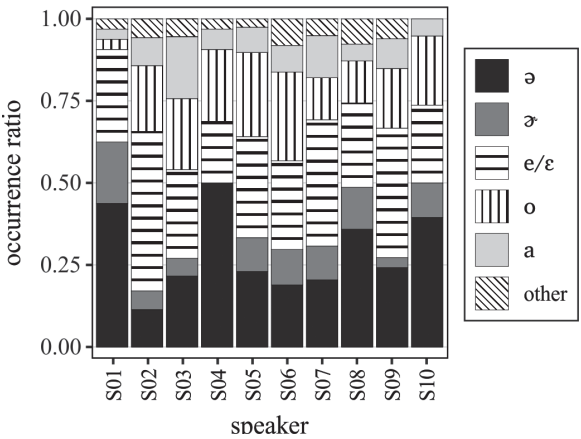

Figure 10. Realizations of vowels corresponding to *schwa*.

like *system*, *operate*, *concentrate* and also with 'a' spellings, for example *company*, *across*, *England*) and mid back [o] (e.g., *completely*, *official*, *ceremony*). An open vowel [a] was realized especially at word ends (e.g., *India*, *idea*, *data*) and also in words like *industry* or *successful*. The category labelled as "other" in Figure 10 includes [u] (*supplies*, *surprise*) and [ɪ] (*allegations*), but also long vowels or diphthongs like [ɔː] (*effort*), [ɔ͡ʊ] (*unanimously*, *protester*) or [e͡ɪ] (*affordable*, *cooperative*).

4.4. Rhythmic patterning

From the pronunciation features which contribute to the typical English rhythm, vowel reduction was addressed in the previous section. In this section, we focus on **linking and glottalization**. As can be seen in Figure 11, most of our speakers did not link words together much; the tendency was slightly higher when the vowel-initial word was grammatical (e.g., *millions of*, *save it*) than when it was lexical (e.g., *should allow*, *in effect*). Speaker S09 linked the most, 46% of the cases, speaker S04 linked in 41% of the cases, predominantly when the second word was grammatical. On the other hand, speaker S02 did not link in any of the 26 possible contexts in her text.

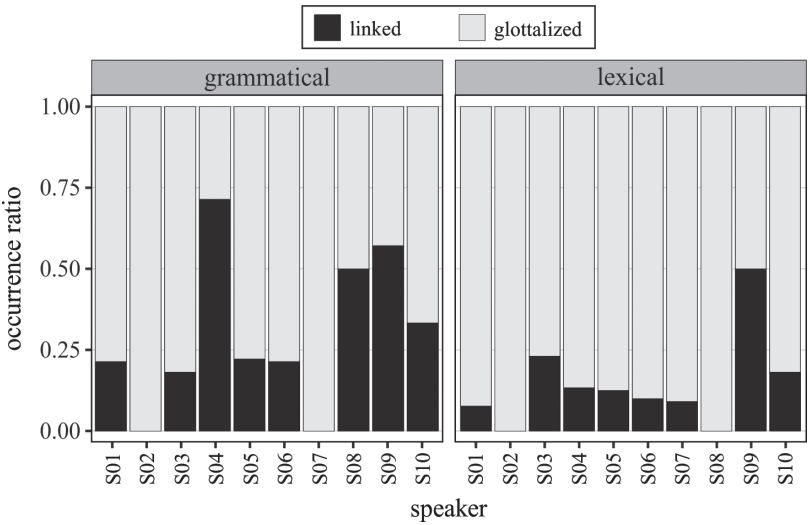

Figure 11. Proportion of linking and glottalization in grammatical and lexical words.

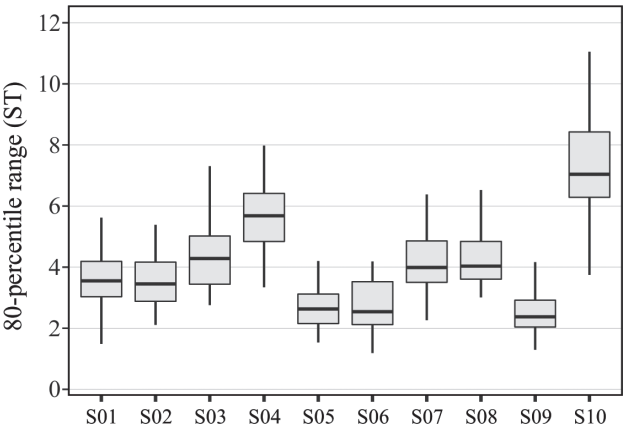

Figure 12. 80-percentile range in semitones in individual breath groups.

4.5. Melodic patterning

The first feature related to intonation analyzed in this study was **pitch range**. As shown in Figure 12, our speakers' pitch range, as evidenced by the 80-percentile range, was confirmed to be quite narrow; the only exception is speaker S10, whose intonation indeed strikes listeners as remarkably lively. Volín et al. (2015) reported their Czech English speakers' 80-percentile range around 4 semitones (ST), and we can see that the median value of most of the speakers in this study moves around the same value, with three speakers' median even below 3 ST.

Finally, we were interested in the **melodic step between the stressed and post-stressed syllable**. The difference in f_0 is illustrated in Figure 13. If we regard the <–0.5; 0.5> ST range as level, since just noticeable difference corresponds to approximately one half of a semitone (Klatt, 1973), it is clear that there are considerably more post-stress rises than there are falls. Only in speaker S09 can we see more falls than rises, but this speaker also displays a lot of "level" steps. To summarize, most of the Czech speakers analyzed in this study tended to pronounce the second syllable in a stress group as higher than the stressed one.

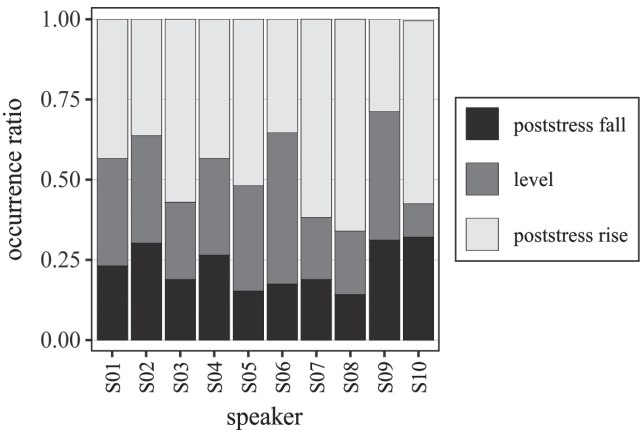

Figure 13. Occurrence of f_0 relationships between the stressed and post-stressed vowel.

5. General discussion and conclusion

The objective of this study was to investigate several pronunciation features which have either been shown by previous research or known based on observation to cause problems to Czech speakers of English. In what, to our best knowledge, is to date the most comprehensive analysis of Czech English pronunciation, we examined ten female speakers with a strong accent in their English, focusing on both segmental and prosodic features. The scope of the analyses, while being a decisive advantage, also constitutes one of the limitations of the current study: the pronunciation features were assessed using different approaches (auditory and acoustic), and the auditory analyses made use of dif-

ferent evaluation scales, as suitable for the particular pronunciation features. However, we are convinced that these drawbacks are outweighed by the benefit of the uniform and consistent approach to the auditory analyses.

The results of this study are summarized in Table 2; the table may serve as a schematic illustration, allowing one to compare the individual pronunciation features and speakers. In accordance with the previous displays, the darkest shade of grey corresponds to most target-like pronunciation. It should be noted that some of the analyses presented above did not include by-speaker display, so that the table contains details beyond those described in section 4.

Table 2. Schematic representation of the speakers' realization of the analyzed features. Dark represents mostly target-like pronunciation, light represents mostly "Czech-like" pronunciation, grey an intermediate step corresponding to inconsistent performance.

	vowels æ ɒ	θ ð	w v	ŋ	ɹ	aspiration	assimilation	stress place	schwa	linking	f_0 range	post-stress f_0
S01												
S02												
S03												
S04												
S05												
S06												
S07												
S08												
S09												
S10												

It is immediately apparent, for instance, that none of the speakers analyzed here pronounced the English /r/ consistently as an alveolar trill: in what we consider a surprising result, seven of the speakers mostly pronounced /r/ as an approximant. Similarly, the difference between /v/ and /w/ was not a problem for half of the speakers, with the other half being less consistent in their pronunciation. From the other end of the scale, we can see that the pronunciation of the two open vowels, /æ/ and /ɒ/, is inconsistent at best (in speakers S06 and S07) and Czech-like for most of the speakers. Similarly, most of the speakers failed to link most of the times, pronounce /ŋ/ without a following plosive, and their intonation range was very compressed in comparison with native speakers of English.

Turning to individual speakers, the table elegantly shows that although all of them have been evaluated as having a relatively strong Czech accent in their English, they differ in their pronunciation of the individual features. Speaker S07 is the only one who did not manifest target-like pronunciation of at least one of the features. All of the others

show a satisfactory performance in at least two pronunciation features, most frequently the above-mentioned /r/ and /w–v/. To summarize, it is clear that the label **strong Czech accent** may be "filled" in different ways, that it may refer to diverse constellations of pronunciation features. Of course, this is not surprising, as speech is a multidimensional phenomenon; in this study, we tried to provide a glimpse of those dimensions most associated with Czech English.

ACKNOWLEDGEMENTS

This study was supported by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

REFERENCES

Abercrombie, D. (1949). Teaching Pronunciation. *ELT Journal*, 3(5), 113–122.

Alameen, G. & Levis, J. M. (2015). Connected speech. In: Reed, M. & Levis, J. M. (Eds.), *The Handbook of English Pronunciation*, 159–174. Oxford: Wiley Blackwell.

Bissiri, M. & Volín, J. (2010). Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English. In: Vích, R. (Ed.), *Speech Processing*, 23–28. Praha: Institute of Photonics and Electronics, CAS.

Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer*, version 6.0.41. Retrieved August 6, 2018 from http://www.praat.org/.

Bořil, T. & Skarnitzl, R. (2016). Tools rPraat and mPraat: Interfacing phonetic analyses with signal processing. In: Sojka, P., Horák, A., Kopeček, I. & Pala, K. (Eds.), *Proceedings of the 19th International Conference on Text, Speech and Dialogue*, 367–374. Cham: Springer International Publishing.

Brown, A. (2016). Barriers to learning the English *th* sounds II: The relative importance of the two sounds. *Speak Out! (Journal of the IATEFL Pronunciation Special Interest Group)*, 54, 6–14.

Červinková Poesová, K. & Weingartová, L. (2018). Character of vowel reduction in Czech English. In: Volín, J. & Skarnitzl, R. (Eds.), *The Pronunciation of English by Speakers of Other Languages*, 96–116. Newcastle upon Tyne: Cambridge Scholars Publishing.

Cho, T. & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207–229.

Crystal, D. (2002). *The English Language: A Guided Tour of the Language*. 2nd ed. London: Penguin Books.

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490.

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing Company.

Eriksson, A. & Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style. In: *Proceedings of Interspeech 2015*, 41–45.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.

Hawkins, S. & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35(2), 183–199.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of speech perception. *Journal of the Acoustical Society of America*, 53, 8–16.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377.
Levis, J. M. (2018). Intelligibility, Oral Communication, and the Teaching of Pronunciation. Cambridge: Cambridge University Press.
Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97.
Munro, M. J. & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531.
Palková, Z. & Volín, J. (2003). The role of F0 contours in determining foot boundaries in Czech. In: *Proceedings of 15th International Congress of Phonetic Sciences*, 1783–1786.
Pospíšilová, A. (2011). Aspiration of English plosives in Czech students of English studies. Praha: Faculty of Arts, Charles University. (unpublished bachelor thesis)
R Core Team (2017). *R: A language and environment for statistical computing (version 3.3.2)*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.Rproject.org.
Skarnitzl, R. (2001). *Teaching and Learning the English Dental Fricatives in the Czech Environment*. Praha: Faculty of Education, Charles University. (unpublished diploma thesis)
Skarnitzl, R. (2004). Fonotaktické chování velární nazály v české angličtině. In: Duběda, T. (Ed.), *Konference česko-slovenské pobočky ISPhS 2004*, 75–83.
Skarnitzl, R. (2018). Fonetická realizace slovního přízvuku u delších slov v češtině. *Slovo a slovesnost*, 79, 199–216.
Skarnitzl, R. & Šturm, P. (2016). Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language*, 14, 1–14.
Skarnitzl, R. & Šturm, P. (2017). Voicing assimilation in Czech and Slovak speakers of English: Interactions of segmental context, language and strength of foreign accent. *Language and Speech*, 60, 427–453.
Skarnitzl, R. & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18, 7–11.
Skarnitzl, R., Volín, J. & Drenková, L. (2005). Tangibility of foreign accents in speech: The case of Czech English. In: Grmelová, A., Dušková, L. & Farrell, M. (Eds.), *2nd Prague Conference on Linguistics and Literary Studies Proceedings*, 11–20. Praha: PedF UK.
Šimáčková, Š. (2003). "Engela's Eshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics. In: *Proc. of 15th ICPhS*, 2293–2296.
Šimáčková, Š., Kolářová, K. & Podlipský, V. J. (2014). Tempo and connectedness of Czech-accented English speech. *Concordia Working Papers in Applied Linguistics*, 5, 667–677.
Šimáčková, Š. & Podlipský, V. J. (2018). Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility. *Research in Language*, 16(2), 169–191.
Šimáčková, Š., Podlipský, V. J. & Kolářová, K. (2014). Linking versus glottalization: (Dis)connectedness of Czech-accented English. *Concordia Working Papers in Applied Linguistics*, 5, 678–692.
Šturm, P. & Skarnitzl, R. (2011). The open front vowel /æ/ in the production and perception of Czech students of English. In: *Proceedings of Interspeech 2011*, 1161–1164.
Volín, J. (2008). Z intonace čtených zpravodajství: výška první slabiky v taktu. Čeština doma a ve světě, 2008 (1–2), 89–96.
Volín, J. & Johaníková, T. (2018). Weak structural words in British and Czech English. In: Volín, J. & Skarnitzl, R. (Eds.), *The Pronunciation of English by Speakers of Other Languages*, 181–195. Newcastle upon Tyne: Cambridge Scholars Publishing.
Volín, J., Poesová, K. & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, 13, 107–123.
Volín, J., Weingartová, L. & Skarnitzl, R. (2013). Spectral characteristics of schwa in Czech accented English. *Research in Language*, 11, 31–39.
Weingartová, L., Poesová, K. & Volín, J. (2014). Prominence contrasts in Czech English as a predictor of learner's proficiency. In: *Proceedings of Speech Prosody 2014*, 236–240.
Wichmann, A. (2005). The role of intonation in the expression of attitudinal meaning. *English Language and Linguistics*, 9(2), 229–253.
Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

RESUMÉ

Příspěvek se věnuje zvukové podobě angličtiny českých mluvčích, kteří ve své angličtině vykazují silný cizinecký přízvuk. Jako první studie se systematicky věnuje většímu množství výslovnostních segmentálních i prozodických rysů, které jsou s českou angličtinou spojovány nebo které již u českých mluvčích angličtiny byly zkoumány. Studie je založena na kombinace poslechových a akustických analýz deseti mluvčích se silným přízvukem. Výsledky ukazují, že v segmentální oblasti mluvčí téměř výhradně vyslovují namísto anglických otevřených samohlásek /æ ɒ/ jejich české středové ekvivalenty. Velární nazála bývá ve slovech chybně následována velární explozivou. V řeči analyzovaných mluvčích se jen zřídka vyskytovalo vázání slov a jejich intonační rozpětí je většinou velmi ploché. U některých dalších výslovnostních jevů, například u aspirace, realizace dentálních frikativ nebo v umístění lexikálního přízvuku, se však mluvčí liší. Výsledky tak zdůrazňují skutečnost, že „silný český přízvuk v angličtině" je označení, které může odpovídat různým konstelacím výslovnostních jevů.

Radek Skarnitzl
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: radek.skarnitzl@ff.cuni.cz

Jana Rumlová
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
Prague, Czech Republic

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 129–143
PHILOLOGICA 2/ PHONETICA PRAGENSIA

DIALECTAL DIFFERENCES IN VOICING ASSIMILATION PATTERNS: THE CASE OF MORAVIAN CZECH ENGLISH

PAVEL ŠTURM and LEA TYLEČKOVÁ

ABSTRACT

One challenge for the second language (L2) learner of English is to master a novel phonetic implementation of the voicing contrast, whereas another challenge is to learn how consonant sequences behave in connected speech. Learners of English coming from three different language backgrounds were tested; their native varieties were Bohemian Czech, Moravian Czech, and Slovak. The Moravian variety of Czech is more similar in voicing assimilation to the Slovak language than to the Bohemian variety of Czech. Percentage of phonetic voicing was measured in the L2 (i.e. English) word-final obstruents preceding three classes of sounds: voiceless and voiced obstruents, and sonorants. Bohemian and Moravian speakers exhibited different strategies in pre-sonorant contexts, following their native (variety-specific) assimilation rules.

Key words: voicing assimilation, transfer, Czech, dialect, L2 English

1. Introduction

As a prime example of a phonological process with clear phonetic grounding, assimilation is a frequent pattern recurring in many languages (Gordon, 2016: Chapter 5). Speakers tend to produce articulatory gestures with some degree of overlap, which may result in segmental changes and elisions. Assimilation can thus be viewed as adaptation of speech sounds to the immediate contex – one sound modifies some of its characteristics, so that the result is more similar to the conditioning segment. The influence is usually anticipatory/regressive (Farnetani & Recasens, 2010). For instance, the place of articulation of the nasal consonant in the Spanish indefinite article "un" is pronounced differently when preceding labial, dental, alveolar or velar consonants ([umˈbaso] "a glass", [un̪ˈd̪eðo] "a finger", [unˈlaɣo] "a lake", [uŋˈgato] "a cat"). In this case, the assimilation occurs online between words in connected speech, but it can also be lexicalized, as in the English prefixed word "impolite" /ˌɪmpəˈlaɪt/. However, the focus of assimilation is not restricted to the consonantal place dimension; we frequently encounter assimilation of voicing (both within words and between words) or, less frequently, manner of articulation. Sounds can even undergo complete assimilation, i.e., modification of all their features, creating articulatory geminates (e.g., the past participle Finnish suffix /nut/, in which the /n/ becomes identical to the preceding oral continuant sound; Gordon, 2016:

https://doi.org/10.14712/24646830.2019.22

125). Given the extensive use of assimilation in the world's languages and, more to the point, the natural tendency of speakers to simplify their production at various levels (Lindblom, 1983), it is not surprising that assimilation is readily transferred to a second language (Altenberg & Vago, 1983; Cebrian, 2000; Simon, 2010), creating high potential for L1 and L2 interaction.

The current article investigates the assimilation of voicing in L2 learners of English with various native languages and varieties. Although voicing assimilation is relevant in the English language for some morphemes (like the plural or the past tense), voicing changes across the word boundary are not allowed. For instance, "lake" is pronounced with a phonologically voiceless obstruent at the end irrespective of the following segment (/lêĭk pə͡ʊɪt/, /lêĭk bɑːd/). In other words, the assimilated form /lêĭg bɑːd/ is not to be expected.[1] In contrast, many Slavic languages do assimilate voicing across words extensively. In Czech, one would say /dost pɛɲɛs/ "enough money", but /dozd bodu:/ "enough points" (for voicing assimilation rules in Czech, see e.g. Palková, 1994: 329ff.). Consequently, voicing of the final obstruent in these languages can entirely be predicted based on the following context, the voicing distinction being neutralized, whereas in English, the distinction is maintained, albeit through other acoustic correlates than actual phonetic voicing (see Chen, 1970; Blevins, 2006).

Interestingly, the Czech language is not uniform in terms of voicing assimilation across the word boundary (see e.g. Palková, 1994: 329ff.), which is similar, for instance, to Polish, whose dialects may be classified into "voicing" and "devoicing" based on assimilatory activity before sonorants (Lew, 2002). There are two main varieties of Czech: Bohemian and Moravian (see Šimáčková, Podlipský & Chládková, 2012). In Moravian Czech, voicing is also assimilated before sonorants, giving rise to forms like /dozd masa/ "enough meat", which would be considered non-standard in Bohemian Czech (where /dost masa/ is pronounced). Thus, Moravian Czech patterns with Slovak in this respect (Pauliny, 1979: 152ff.; Bárkányi & Beňuš, 2015; Bárkányi & Kiss, 2015).

It is clear from the facts mentioned above that the assimilatory process is highly language- and even variety-specific, and therefore prone to interaction in L2 acquisition when the two languages have different assimilation systems. Restricting our attention to L2 English, we can point out several studies that all showed evidence of transferring L1 assimilation rules into L2 English production (Altenberg & Vago, 1983 for Hungarian; Rubach, 1984 and Lew, 2002 for Polish; Cebrian, 2000 for Catalan; Simon, 2010 for Dutch). In these studies, voiceless targets were typically realized as voiced when preceding voiced obstruents. Word-final voicing has also been studied in Czech English. The implementation of the voicing contrast was examined by Fejlová (2013) or by Skarnitzl and Šturm (2016), whereas the process of voicing assimilation itself was analysed by Skarnitzl and Poesová (2008), Kanioková (2011) and Skarnitzl and Šturm (2014, 2017). The latter work (Skarnitzl & Šturm, 2017) focused on voicing assimilation in L1 British English and in the L2 speech of Czech and Slovak speakers. An interesting finding was that whereas the two L2 groups showed comparable patterns before obstruents, the

[1] Traditional works on English phonetics (e.g. Cruttenden, 2014) do not admit cross-word assimilation, especially if a voiceless segment should become voiced. However, empirical data suggest that the situation is more complex. For example, Jansen (2004, 2007) found that word-initial [d] and [z] exerted some influence on the voicing of word-final [k] in native British English.

context before sonorant consonants yielded disparate patterns: the (Bohemian) Czech participants tended to produce voiceless pre-sonorant obstruents (*at*[t] *least*), but the Slovak speakers had a tendency to assimilate voicing to the following sonorant segment, producing a voiced sound (*at*[d] *least*). Crucially, this reflected the voicing assimilation rules of the respective L1 languages, which differ in the pre-sonorant context.

Importantly, Chládková and Podlipský (2011) showed that learning L2 speech sounds contrasts is not only language-specific, but also variety-specific. They examined the cross-language perception of Dutch vowels by speakers of the Bohemian and Moravian varieties of Czech. Their study revealed different perceptual assimilation patterns in the non-native Dutch high front vowel region that reflected the between-dialect acoustic differences in signalling the L1 Czech phonological length contrast in high front vowels.

In the present study, we investigate whether Bohemian and Moravian speakers exhibit variety-specific voicing assimilation patterns in L2 English. We can hypothesize that they will behave in a variety- rather than language-specific manner. If this is the case, then, Moravian Czech L2 English should be closer to Slovak L2 English (a different source language, but with similar assimilation rules) than to Bohemian Czech English (the same source language, but with dissimilar assimilation rules). A second research question will be connected to the general pronunciation competence of our L2 speakers, which might also influence the strength of L1-L2 transfer. We predict that more accented speakers will show higher rates of voicing assimilation in L2 English than less accented speakers.

2. Method

The current study presents new data but also uses acoustic data from our previous study (Skarnitzl & Šturm, 2017). The "new" dataset involves speakers of Moravian Czech L2 English (MorCZ), whereas the "previous" dataset was based on speakers of Bohemian Czech L2 English (BohCZ) and Slovak L2 English (SK), in addition to a control British English (BrE) L1 group not considered here. The recording and analysis were identical in both cases. To facilitate statistical comparisons between the groups, both datasets are merged into a single analysis.

2.1. Participants

12 female speakers of Moravian Czech were recorded with origin in various Moravian and Silesian regions, i.e., regions where pre-sonorant voicing assimilation occurs. Auditory observation of their productions confirmed that their Czech production had the "assimilating" characteristics typical of the Moravian variety of Czech. In contrast, the 12 Bohemian Czech speakers from the previous dataset did not show this type of voicing assimilation in their Czech production. The Slovak group also included 12 speakers. All the speakers were female and aged between 20 and 25 years.

The three learner groups comprised two types of speakers. Six speakers belonged to a "more-accented (ma)" group with a strong Czech/Slovak accent in their English and six to a "less-accented (la)" group that was almost near-native in their pronunciation of English. In the previous study, the speakers were selected from a larger corpus of L2

English based on agreement between three phoneticians. In addition, the speakers were evaluated by seven native speakers of English in terms of foreign accent strength on a 7-point scale. The results showed a clear separation between the "ma" and "la" groups; the control group (BrE) was evaluated similarly as the "la" group, but with a clear separation from the "ma" group.

For the current study, we recorded 6 Moravian Czech students of English philology in Prague whose pronunciation was considered near-native ("la"). For the "ma" group, three speakers were recorded in Prague immediately at the start of their English studies, and three more speakers who did not study English at all were recorded in Olomouc. Both groups were characterized by a strong Czech accent. Nevertheless, a perceptual evaluation was also conducted to substantiate the categorization of the Moravian speakers' accent strength into "ma" and "la" groups. 56 listeners (students of English studies, aged 18–68, with mean age 21.18 and sd 7.1) were asked to evaluate the speakers' foreign accent strength on a 7-point scale ranging from "strong foreign accent" (1) to "native-like" (7). The test included the 12 MorCZ speakers, 12 SK speakers and also 6 native speakers of English as controls and several L2 speakers of other languages as fillers. The listeners were assigned one of four test versions differing in presentation order, and evaluated 60 stimuli. The duration of the test was 11 minutes.

The results did not show significant differences between the four test versions (differing in trial orders). Figure 1 shows the evaluation of the speakers' accent strength as a function of learner group. There was a clear difference between the "ma" and "la" groups of both Moravian and Slovak speakers. The mean value of the "la" participants differed somewhat from the control native English group (Tukey contrasts: MorCZ by −0.96 with SE = 0.3, $z = -3.1$, $p < 0.05$; SK by −1.09 with SE = 0.3, $z = -3.5$, $p < 0.01$), whereas the "ma" participants differed substantially (MorCZ by −3.68 with SE = 0.3, $z = -11.7$, $p < 0.001$; SK by −3.88 with SE = 0.3, $z = -12.4$, $p < 0.001$). We can therefore conclude that our initial evaluation of the MorCZ speakers was correct.

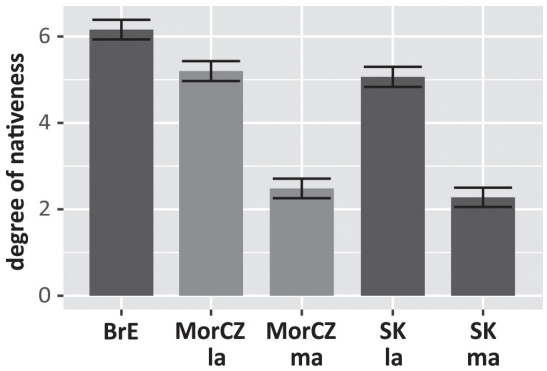

Figure 1. Evaluation of foreign accent strength (degree of nativeness) for an L1 British English (BrE) control group and for L2 groups of more ("ma") and less accented ("la") Moravian Czech (MorCZ) and Slovak (SK) speakers. For Bohemian Czech speakers, see Skarnitzl and Šturm (2017).

2.2. Material

After sufficient time for preparation, the participants were asked to read one of six BBC World Service news bulletins. In the MorCZ group, "ma" and "la" speakers were matched for the text version. They were recorded in sound-treated studios in Prague and Olomouc (16-bit, 32 kHz .wav files recorded with a condenser microphone). Each recording was approximately 4 minutes long and consisted of 450–500 words, depending on the text version. The speakers from Skarnitzl and Šturm (2017) were recorded in the Prague studio under the same conditions but, due to the corpus structure, the text versions were not matched between the "la" and "ma" groups, and a wider range of texts was used as well (16 versions).

The recordings were automatically segmented by means of P2FA forced alignment (Yuan & Lieberman, 2008), and the boundaries of the target speech sounds were manually adjusted based on the phonetically motivated recommendations for manual segmentation of the speech signal (Machač & Skarnitzl, 2009). The targets involved two consecutive phones – a word-final obstruent and the initial consonant of the following word (except for /h/). Therefore, there were three contexts in which the target sound occurred: before a voiceless (fortis) consonant, before a voiced (lenis) consonant, and before a sonorant consonant. The nature of the preceding speech sound was not controlled. Target sequences interrupted with a pause were excluded since we do not expect assimilation to occur in such cases. The presence of prosodic breaks was noted, as the assimilation rate may differ within and across prosodic boundaries (Mády & Bárkányi, 2015). The analysis was based on 947 tokens (MorCZ data) and 1952 tokens (SK and BohCZ data from Skarnitzl & Šturm, 2017). The structure of the data is shown in Table 1.

Table 1. Breakdown of the dataset according to learner group (Bohemian Czech, Moravian Czech, Slovak, la = less accented, ma = more accented) and assimilatory context (vl = voiceless, vd = voiced, son = sonorant).

Learner group	**vl-vl**	**vd-vl**	**vl-vd**	**vd-vd**	**vl-son**	**vd-son**	**Group total**
BohCZ_la	59	123	55	123	66	83	**509**
BohCZ_ma	58	127	42	127	85	60	499
MorCZ_la	57	118	67	109	74	76	**501**
MorCZ_ma	32	104	61	115	62	72	**446**
SK_la	101	110	57	120	73	60	**521**
SK_ma	99	66	50	91	60	57	**423**
Assim. context total	406	648	332	685	420	408	2899

2.3. Analysis

In order to assess voicing of the target sounds, we examined the presence or absence of the fundamental frequency (F0), which was extracted in all word-final obstruents in Praat (Boersma & Weenink, 2017) with the default setting for F0 extraction, as we were not interested in specific values. The degree of voicing was expressed as the *percentage of*

voicing. Voicing information was extracted every millisecond and the voicing ratio was computed for each target sound (i.e., how much of the consonant was produced with vocal fold vibration).

For statistical analysis we used the publicly available program R (R Core Team, 2019) and the associated R packages *lme4*, *effects* and *ggplot2* (Bates, Mächler, Bolker, & Walker, 2015; Fox, 2003; Wickham, 2009). Linear mixed-effects (LME) modelling was chosen because it is suitable for multiple observations from the same speaker or of the same item to resolve the non-independence of such observations. The fixed effects were ASSIMILATORY CONTEXT (vl-vl × vl-vd × vl-son × vd-vl × vd-vd × vd-son), LEARNER GROUP (BohCZ_la × BohCZ_ma × MorCZ_la × MorCZ_ma × SK_la × SK_ma), PROSODIC BREAK (yes × no), TARGET MANNER (fricative × stop) and LEXICAL STATUS (lexical × grammatical). The random effects were the intercepts for SUBJECT and WORD; by-subject random slopes were not added because the complex effect structure would lead to singularity in the random terms. Whether individual fixed effects/interactions were significant was evaluated by comparing the full model to a reduced model in which the factor in question was excluded, using likelihood ratio tests. In addition, pairwise comparisons were evaluated post hoc using Tukey contrasts from the *multcomp* package (Hothorn, Bretz & Westfall, 2008).

3. Results and discussion

Skarnitzl and Šturm (2017) reported some general phonetic and linguistic effects on voicing assimilations of Czech and Slovak learners of English. For instance, the presence of a prosodic break after the target obstruent led to a decrease in the amount of phonetic voicing, or word-final fricatives were on the whole articulated with less voicing than word-final stop consonants. We therefore included such factors in the current model by default and examined their interactions. PROSODIC BREAK interacted with LEXICAL STATUS ($\chi^2(1) = 6.9$, $p < 0.01$): grammatical words were associated with a higher degree of voicing than lexical words, but only in the absence of a prosodic break. Furthermore, the LEXICAL STATUS effect was restricted to stop consonants, as fricatives did not show any difference in voicing between lexical and grammatical words ($\chi^2(1) = 4.4$, $p < 0.05$). Finally, there was no significant interaction between PROSODIC BREAK and TARGET MANNER ($\chi^2(1) = 0.01$, $p = 0.89$).

The factors of greatest interest for the present research question are ASSIMILATORY CONTEXT and LEARNER GROUP. Crucially, there was a significant interaction between them ($\chi^2(25) = 255.6$, $p < 0.001$). The fixed effects estimates of the full model are given in Appendix A, but it is easier to evaluate the direction and size of the effects in the accompanying effects plots in Figure 2. The contexts *before word-initial voiceless consonants* (Fig. 2a) are included as baselines for determining the degree of phonetic voicing due to carryover from previous sounds. The results generally follow the expectations that there should be no significant differences between the varieties, given that all L1 backgrounds assimilate in these contexts. However, the less accented BohCZ speakers yielded a higher rate of voicing in the lenis-fortis (voiced-voiceless) context compared to the other groups, which might reflect a more English-like pronunciation (i.e., an attempt at approximation

to word-final devoicing, rather than to assimilation). Also, there was a general tendency in all groups to produce more phonetic voicing in the lenis targets compared to the fortis targets.

With regard to the *pre-sonorant contexts* (Fig. 2b), several aspects need to be mentioned. Although the BohCZ speakers did not assimilate word-final voiceless consonants before sonorants, treating the target sounds as if a voiceless consonant followed (compare the solid columns in Fig. 2b to 2a), the MorCZ group was associated with a significantly higher amount of phonetic voicing before sonorants. Consequently, the MorCZ variety resembled more closely the Slovak group, which was likewise associated with a higher amount of voicing in this context but which manifested additional differences in speaker accentedness. Finally, we can also examine the performance of individual learner groups within the given context. Crucially, the more accented MorCZ speakers differed significantly from the corresponding BohCZ speakers but not from the corresponding SK speakers. In contrast, there were no significant differences among the varieties in the less accented speakers. The results of post hoc multiple comparisons using Tukey contrasts are provided in Appendix B.

However, the patterns become more complex when the word-initial sonorant consonant is preceded by a lenis (phonologically voiced) consonant (transparent colours in Fig. 2b and Tukey post-hoc tests in Appendix B). On the one hand, an analogous effec – a shift towards significantly higher voicing percentages before sonorants – was associated with the MorCZ and SK speakers. On the other hand, it was also found in the less accented BohCZ group, which does not seem to follow the expectations. Even so, we suggest that it is not a clear argument for assimilation, as discussed by Skarnitzl and Šturm (2017). Whereas the more accented BohCZ speakers produced a low amount of voicing before sonorants, identical to the pre-voiceless contexts, the less accented BohCZ speakers might have been targeting the devoiced word-final lenis obstruent in native English by maintaining phonetic voicing. This would make them seemingly pattern with the "assimilating" groups, but it could in fact reflect their higher awareness that word-final voicing is not neutralized in English, and an attempt to aim for such a target. It is thus difficult to say what the finding about MorCZ speakers in the voiced-sonorant context means: is the higher amount of phonetic voicing due to L1 transfer, or to a higher pronunciation proficiency level, as suggested for the BohCZ less accented speakers?

Results from the contexts *before word-initial voiced consonants* (Fig. 2c) indicate, first, that there is a significant increase in each learner group in the amount of voicing compared to the contexts before voiceless consonants. Moreover, with the exception of the less accented SK speakers, all groups yielded a higher percentage of voicing before voiced sounds than before sonorants. This, of course, clearly reflects the L1 assimilation rules, although it needs to be explained why the percentages are not higher, approaching 100% (at least for the more accented speakers). Second, the voiced targets again yielded higher rates of assimilation in comparison to the voiceless targets. Finally, the more accented speakers of each variety were more prone to assimilation than the less accented speakers, although the effect size differed for individual varieties.

Interestingly, there is a recurring pattern in the data regarding the more accented speakers: Slovaks were associated with a higher amount of phonetic voicing than MorCZ speakers, and these in turn with more phonetic voicing than BohCZ speakers. The lan-

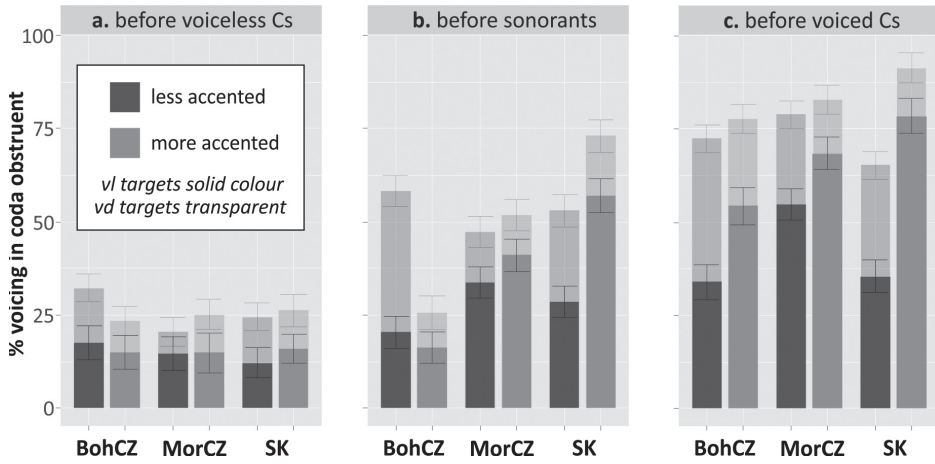

Figure 2. Mean amount of phonetic voicing (% of the target sound duration) in different assimilatory contexts and different learner groups (BohCZ = Bohemian Czech, MorCZ = Moravian Czech, SK = Slovak, vl = voiceless, vd = voiced). Effect plots from an LME analysis.

guage background provides a possible explanation: Bohemian Czech assimilates obstruents before other obstruents, Moravian Czech before obstruents and sonorants, and Slovak before obstruents, sonorants and also vowels. The speakers might be differentially primed from their native systems in their L2 English production, triggering different rates of transfer. Another pattern is apparent in Figure 2 if we examine the contexts which are arguably most conclusive about transfer, i.e., voiceless targets before voiced and sonorant consonants. The higher degree of assimilation in the less accented Moravian Czech speakers suggests they are less advanced L2 users than the corresponding "la" Slovaks (and Bohemian Czechs), despite being judged similarly in the perception test. Indeed, the accentedness effect sizes are greater for the latter group than for Moravian Czechs.

4. General discussion

The objective of the study was to investigate whether voicing assimilation strategies in Moravian Czech English tend to be more similar to Bohemian Czech English or to Slovak English, taking into account the speakers' pronunciation competence in English. We observed differences between less and more accented speakers, the effect being most pronounced in the Slovak speakers. One could argue that the degree of accentedness, although treated identically in the three L2 varieties (i.e., as a binary variable), was in fact not comparable, and that, for instance, the Slovaks showed a higher degree of phonetic voicing because of a generally higher degree of accentedness. However, the results of the perceptual tests (see Fig. 1 and Skarnitzl & Šturm, 2017) indicate that this was not the case, given that all members of the more vs. less accented groups received similar scores irrespective of their dialect. Nevertheless, it could still be the case that the tendency to

assimilate is not necessarily correlated with the evaluation of the speakers' foreign accent strength, as it was based on overall pronunciation skills and not on assimilatory behaviour. In other words, a speaker who behaves in terms of assimilation entirely according to the L2 rules might still be perceived as heavily-accented, and vice versa.

Our results showed a clear evidence of L1 interference in the English language production in all the three groups, especially with regard to the more accented speakers. On the one hand, all three L2 English groups of speakers produced word-final obstruents with substantial phonetic voicing before phonologically voiced consonants, whereas in the contexts before initial voiceless consonants, the target sounds were associated with a low degree of phonetic voicing. This finding can be interpreted as a transfer of L1 assimilation rules into an L2, which is manifested identically in the different varieties. On the other hand, the pre-sonorant contexts were associated with different assimilation strategies in the two varieties of Czech – the Moravian Czech speakers approximated the Slovak speakers, exhibiting assimilatory behaviour to a greater extent than did the Bohemian Czech speakers (but to a smaller extent than did the Slovaks). Moravian Czech English thus seems to be in this respect intermediate between Bohemian Czech and Slovak English. Moreover, we can conclude based on the effect sizes that Moravian Czech English is closer to Slovak English, a different source language that nevertheless has similar assimilation rules, than to Bohemian Czech English, the same source language but with different assimilation rules. There was even a statistically significant difference from the Bohemian Czech group if we consider the more accented speakers. In any case, the finding suggests that the two varieties of Czech exhibit L2 English assimilatory behaviour in a dissimilar way, and the direction of the effect is influenced by the assimilation rules of the speakers' own, variety-specific phonological system. These findings are in line with Lew's results for Polish dialects (Lew, 2002).

One aspect of our data is particularly interesting. In all the groups and regardless of the type of the word-initial consonant, word-final phonologically voiced sounds were associated with a generally higher degree of voicing than the corresponding voiceless counterparts (compare the transparent and solid colours in Fig. 2). Why is it so? The most probable explanation is that speakers are aware – not necessarily consciously – of the fact that, in native English, word-final /d z/ etc. do not turn into /t s/ (unlike in Czech or Slovak). This would be supported by the differential behaviour of the more and less accented speakers. The less accented speakers, who might be hypothesized to be more proficient and more aware of the L2 phonological system, seem to show greater differences between underlying voiced vs. voiceless target consonants compared to the more accented speakers (see Fig. 2). This would correspond to the findings of Cebrian (2000), where such positive transfer (i.e., /d/ realized with phonetic voicing before a voiced sound) was stronger than negative transfer (/t/ realized with phonetic voicing before a voiced sound).

Future research might further examine L2 voicing assimilation patterns by applying a more refined control of the assimilatory contexts, taking into account the voicing status of the preceding sounds as well. Apart from that, it would also be useful to measure vowel duration, which is an important cue to phonological voicing in native English (Chen, 1970; Jansen, 2004; Davidson, 2016). This approach might help differentiate between assimilation per se, when the phonological voicing status is changed, and obstruent devoicing, when the underlying category remains the same despite changes in phonetic

voicing. Moreover, it would then be clearer whether the presence of phonetic voicing in word-final obstruents in L2 English reflects the transfer of L1 voicing assimilation patterns, or a higher level of pronunciation proficiency in Bohemian, Moravian, and Slovak speakers of English. Finally, in addition to measuring the percentage of voicing in the target word-final obstruents, the voicing profile method (Möbius, 2004) could be applied to capture the dynamics of voicing, which may provide a more detailed insight into individual assimilatory contexts.

ACKNOWLEDGEMENTS

The work was supported by the European Regional Development Fund-Project 'Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World' (No. CZ.02.1.01/0.0/0.0/16_019/0000734). The second author was supported from the project 'Language and its research tools', solved at Charles University from Specific university research in 2018.

REFERENCES

Altenberg, E. & Vago, R. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning,* 33(4), 427–447.

Bárkányi, Z. & Beňuš, Š. (2015). Prosodic conditioning of pre-sonorant voicing. *Proceedings of the18th ICPhS*. Retrieved from https://www.internationalphoneticassociation.org/icphspproeedings/ICPhS2015/Papers/ICPHS0336.pdf

Bárkányi, Z. & Kiss, Z. G. (2015). Why do sonorants not voice in Hungarian? And why do they voice in Slovak? In: K. É. Kiss, B. Surányi & É. Dékány (Eds.), *Approaches to Hungarian: Volume 14: Papers from the 2013 Piliscsaba conference* (pp. 65–94). Amsterdam: John Benjamins Publishing Company.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics,* 32(2), 117–166.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer (version 6.0.36) [Computer software]. Retrieved from http://www.praat.org.

Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan–English interlanguage. *Studies in Second Language Acquisition,* 22(1), 1–26.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129–159.

Chládková, K. & Podlipský, V.J. (2011). Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. *Journal of the Acoustical Society of America*, 130, EL186–192.

Cruttenden, A. (2014). *Gimson's Pronunciation of English* (Eighth Edition). London: Routledge.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54(1), 35–50.

Farnetani, E. & Recasens, D. (2010). Coarticulation and connected speech processes. In: W. J. Hardcastle, J. Laver & F. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd edition) (pp. 316–352). Oxford: Wiley-Blackwell.

Fejlová, D. (2013). Pre-fortis shortening in fluent read speech: A comparison of Czech and native speakers of English. *AUC Philologica 1/2014, Phonetica Pragensia XIII*, 91–100.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15), 1–27.

Gordon, M. K. (2016). *Phonological Typology*. Oxford: Oxford University Press.
Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363.
Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. PhD Thesis. University of Groningen, The Netherlands.
Jansen, W. (2007). Phonological 'voicing', phonetic voicing, and assimilation in English. *Language Sciences*, 29(2), 270–293.
Kanioková, Z. (2011). Voicing assimilation in English spoken by Czech and Slovak learners. BA thesis.
Lew, R. (2002). Differences in the scope of obstruent voicing assimilation in learners' English as a consequence of regional varieties in Polish. In: E. Waniek-Klimczak & P. James Melia (Eds.), *Accents and Speech in Teaching English Phonetics and Phonology: EFL Perspective* (pp. 243–264). Frankfurt am Main: Peter Lang.
Lindblom, B. (1983). Economy of speech gestures. In: P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 217–246). Berlin: Springer.
Machač, P. & Skarnitzl, R. (2009). *Principles of Phonetic Segmentation*. Prague: Epocha.
Mády, K., & Bárkányi, Z. (2015). Voicing assimilation at accentual phrase boundaries in Hungarian. *Proceedings of the 18th ICPhS*. Retrieved from https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2015/Papers/ICPHS0796.pdf.
Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, 38, 5–26.
Palková, Z. (1994). *Fonetika a fonologie češtiny* [Phonetics and phonology of Czech]. Prague: Karolinum.
Pauliny, E. (1979). *Slovenská fonológia* [Slovak phonology]. Bratislava: Slovenské pedagogické nakladateľstvo.
R Core Team (2019). R: A language and environment for statistical computing (version 3.5.3) [Computer software]. R Foundation for Statistical Computing, Vienna. Retrieved from http://www.R-project.org.
Rubach, J. (1984). Rule typology and phonological interference. In: S. Eliasson (Ed.), *Theoretical issues in contrastive phonology: Studies in descriptive linguistics* (pp. 37–50). Heidelberg: Julius Groos.
Simon, E. (2010). Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences,* 32(1), 63–86.
Skarnitzl, R. & Poesová, K. (2008). Typology of voicing changes in Czech English. In: A. Grmelová, L. Dušková, M. Farrell & R. Pípalová (Eds.), *Plurality and Diversity in English Studies – Proceedings from the Third Prague Conference on Linguistics and Literary Studies* (pp. 8–17). Prague: Faculty of Education, Charles University in Prague.
Skarnitzl, R. & Šturm, P. (2014). Assimilation of voicing in Czech speakers of English: The effect of the degree of accentedness. *Research in Language,* 12, 199–208.
Skarnitzl, R. & Šturm, P. (2016). Pre-fortis shortening in Czech English: A production and reaction-time study. *Research in Language,* 14, 1–14.
Skarnitzl, R. & Šturm, P. (2017). Voicing assimilation in Czech and Slovak speakers of English: Interactions of segmental context, language and strength of foreign accent. *Language and Speech*, 60(3), 427–453.
Šimáčková, Š., Podlipský, V., & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association,* 42(2), 225–232.
Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* New York: Springer.
Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08.* Retrieved from http://www.ling.upenn.edu/~jiahong/publications/c09.pdf.

APPENDIX A

Regression coefficients of fixed effects in the LME model (vl = voiceless, vd = voiced, son = sonorant, BohCZ = Bohemian Czech, MorCZ = Moravian Czech, SK = Slovak, la = less accented, ma = more accented). The intercept corresponds to fricatives in grammatical words in the vl-vl context spoken by less accented BohCZ speakers when no prosodic boundary was present.

Fixed effect	Estimate	SE	t-value
Intercept	17.31	5.16	3.36
ASSIMILATORY CONTEXT (vl-vd)	16.35	4.90	3.33
ASSIMILATORY CONTEXT (vl-son)	2.74	4.64	0.59
ASSIMILATORY CONTEXT (vd-vl)	14.63	4.21	3.48
ASSIMILATORY CONTEXT (vd-vd)	54.82	4.22	12.99
ASSIMILATORY CONTEXT (vd-son)	40.72	4.49	9.08
LEARNER GROUP (BohCZ_ma)	−2.71	6.23	−0.44
LEARNER GROUP (MorCZ_la)	−3.05	6.16	−0.50
LEARNER GROUP (MorCZ_ma)	−2.74	6.81	−0.40
LEARNER GROUP (SK_la)	−5.38	5.84	−0.92
LEARNER GROUP (SK_ma)	−1.65	5.78	−0.29
PROSODIC BREAK (yes)	−28.02	5.14	−5.45
LEXICAL STATUS (lexical)	−4.37	2.86	−1.53
TARGET MANNER (stop)	17.91	3.91	4.58
CONTEXT (vl-vd) : GROUP (BohCZ_ma)	23.15	7.13	3.25
CONTEXT (vl-son) : GROUP (BohCZ_ma)	−1.32	6.35	−0.21
CONTEXT (vd-vl) : GROUP (BohCZ_ma)	−6.07	5.72	−1.06
CONTEXT (vd-vd) : GROUP (BohCZ_ma)	8.07	5.70	1.42
CONTEXT (vd-son) : GROUP (BohCZ_ma)	−29.97	6.37	−4.70
CONTEXT (vl-vd) : GROUP (MorCZ_la)	23.89	6.50	3.68
CONTEXT (vl-son) : GROUP (MorCZ_la)	16.56	6.28	2.64
CONTEXT (vd-vl) : GROUP (MorCZ_la)	−8.75	5.63	−1.56
CONTEXT (vd-vd) : GROUP (MorCZ_la)	9.51	5.66	1.68
CONTEXT (vd-son) : GROUP (MorCZ_la)	−7.93	6.07	−1.31
CONTEXT (vl-vd) : GROUP (MorCZ_ma)	37.3	7.17	5.20
CONTEXT (vl-son) : GROUP (MorCZ_ma)	23.49	7.04	3.34
CONTEXT (vd-vl) : GROUP (MorCZ_ma)	−4.33	6.39	−0.68
CONTEXT (vd-vd) : GROUP (MorCZ_ma)	13.28	6.34	2.09
CONTEXT (vd-son) : GROUP (MorCZ_ma)	−3.88	6.75	−0.57
CONTEXT (vl-vd) : GROUP (SK_la)	6.92	6.41	1.08
CONTEXT (vl-son) : GROUP (SK_la)	13.53	6.02	2.25

CONTEXT (vd-vl) : GROUP (SK_la)	−2.38	5.35	−0.45
CONTEXT (vd-vd) : GROUP (SK_la)	−1.69	5.29	−0.32
CONTEXT (vd-son) : GROUP (SK_la)	0.05	5.97	0.01
CONTEXT (vl-vd) : GROUP (SK_ma)	46.14	6.40	7.21
CONTEXT (vl-son) : GROUP (SK_ma)	38.43	6.08	6.33
CONTEXT (vd-vl) : GROUP (SK_ma)	−4.40	5.58	−0.79
CONTEXT (vd-vd) : GROUP (SK_ma)	20.67	5.36	3.86
CONTEXT (vd-son) : GROUP (SK_ma)	16.42	5.90	2.78
PROS. BREAK (yes) : LEX. STATUS (lexical)	14.17	5.39	2.63
LEX. STATUS (lexical) : TAR. MANNER (stop)	−8.75	4.18	−2.10

APPENDIX B

Tukey multiple comparisons of means (only selected contrasts of interest). Positive estimates indicate that the first member in the comparison was associated with higher proportions of voicing in the word-final obstruent (BohCZ = Bohemian Czech, MorCZ = Moravian Czech, SK = Slovak, la = less accented, ma = more accented, vl = voiceless, vd = voiced, son = sonorant).

Table B.1: Effect of CONTEXT.

Learner group	Pairwise comparison	Estimate	SE	*z*	adjusted *p*
BohCZ_la	vl-son – vl-vl	2.74	4.64	0.59	1.00
BohCZ_ma	vl-son – vl-vl	1.42	4.43	0.32	1.00
MorCZ_la	vl-son – vl-vl	19.30	4.49	4.30	<0.01
MorCZ_ma	vl-son – vl-vl	26.23	5.46	4.80	<0.001
SK_la	vl-son – vl-vl	16.27	3.94	4.14	<0.05
SK_ma	vl-son – vl-vl	41.17	4.22	9.76	<0.001
BohCZ_la	vd-son – vd-vl	26.09	3.55	7.34	<0.001
BohCZ_ma	vd-son – vd-vl	2.19	3.90	0.56	1.00
MorCZ_la	vd-son – vd-vl	26.92	3.61	7.46	<0.001
MorCZ_ma	vd-son – vd-vl	26.55	3.77	7.05	<0.001
SK_la	vd-son – vd-vl	28.52	3.98	7.17	<0.001
SK_ma	vd-son – vd-vl	46.91	4.49	10.46	<0.001

Table B.2: Effect of LEARNER GROUP.

Context	Pairwise comparison	Estimate	SE	*z*	adjusted *p*
vl-son	MorCZ_la – BohCZ_la	13.51	5.83	2.32	0.90
vl-son	MorCZ_la – SK_la	5.36	5.74	0.93	1.00
vl-son	BohCZ_la – SK_la	−8.15	5.90	−1.38	1.00
vl-son	MorCZ_ma – BohCZ_ma	24.78	5.75	4.31	<0.01
vl-son	MorCZ_ma – SK_ma	−16.04	6.03	−2.67	0.68
vl-son	BohCZ_ma – SK_ma	−40.81	5.85	−6.98	<0.001
vd-son	MorCZ_la – BohCZ_la	−10.98	5.60	1.96	0.99
vd-son	MorCZ_la – SK_la	−5.65	5.83	0.97	1.00
vd-son	BohCZ_la – SK_la	5.33	5.82	0.92	1.00
vd-son	MorCZ_ma – BohCZ_ma	26.06	5.85	4.45	<0.01
vd-son	MorCZ_ma – SK_ma	−21.38	5.91	−3.62	0.08
vd-son	BohCZ_ma – SK_ma	−47.45	6.08	−7.81	<0.001

Table B.3: Effect of ACCENTEDNESS.

Context	Variety	Pairwise comparison	Estimate	SE	*z*	adjusted *p*
vl-son	BohCZ	ma – la	−4.04	5.83	−0.69	1.00
vl-son	MorCZ	ma – la	7.23	5.78	1.25	1.00
vl-son	SK	ma – la	28.63	5.94	4.82	<0.001
vd-son	BohCZ	ma – la	−32.68	5.84	−5.59	<0.001
vd-son	MorCZ	ma – la	4.36	5.62	0.78	1.00
vd-son	SK	ma – la	20.09	6.07	3.31	0.20

RESUMÉ

Osvojování znělostního kontrastu a jeho fonetická realizace představuje u nerodilých mluvčích angličtiny nelehký úkol, zejména, jedná-li se o konsonantické shluky v souvislé řeči. Tato studie se zaměřuje na asimilaci znělosti anglických finálních obstruentů u mluvčích pocházejících ze tří různých jazykových oblastí – Čech, Moravy a Slovenska. Jelikož moravští mluvčí uplatňují asimilaci znělosti podobně jako slovenští mluvčí, a naopak rozdílně než mluvčí obecné češtiny, lze předpokládat, že osvojování cizího jazyka bude ovlivněno nejen jazykově specifickými charakteristikami, ale rovněž odlišnostmi mezi varietami téhož jazyka. Autoři studie sledují míru fonetické znělosti (kvantifikovanou jako podíl trvání znělé části cílového konsonantu) při produkci čteného anglického textu. Cílový segment se nacházel v pozici před třemi typy iniciálních konsonantů: před neznělými a znělými obstruenty a před sonorami. Výsledky analýz potvrzují, že se mluvčí z Čech a Moravy řídí v kontextech před sonorami odlišnými strategiemi, což odpovídá asimilačním pravidlům z rodné variety jazyka. Anglický čtený projev mluvčích z Moravy se tak blíží z hlediska asimilací spíše mluvčím ze Slovenska než mluvčím z Čech.

Pavel Šturm and Lea Tylečková
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: pavel.sturm@ff.cuni.cz

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 145–158
PHILOLOGICA 2/ PHONETICA PRAGENSIA

THE SIZE OF PROSODIC PHRASES IN NATIVE AND FOREIGN-ACCENTED READ-OUT MONOLOGUES

JAN VOLÍN

ABSTRACT

The objective of this study is to provide quantitative data concerning size of prosodic phrases in foreign-accented Czech. The speech production of Anglophone users of the Czech language is contrasted with that of Czech professional and non-professional speakers. Each of the three groups of speakers of Czech is represented by 12 speakers. The fourth group of speakers (also 12 subjects) are English professional news readers. They provide data pertaining to the mother tongue of the target group. As expected, the prosodic phrases produced by non-native speakers are shorter and our data provide basis for their modelling that can be used in perceptual testing. One of the interesting outcomes of the study is the revelation that although Czech professional speakers make longer phrases than English professionals if counted in syllables (10.78 against 7.76 syllable per phrase), if counted in words, the difference disappears (4.56 against 4.54 words per phrase). This suggests that semantic constraints on prosodic phrase length are stronger than purely structural ones.

Key words: prosodic phrase, prosodic boundary, foreign accent, clear speech, Czech, English

1. Introduction

Discussions of prosodic phrasing customarily start with the cases of contrastive representational meaning. Linguists are primarily concerned with, and laymen understandably interested in pairs of identical sequences of words which can be uttered so that they mean different things. If the sequence *Definitely not Archie* materializes as one phrase, then it may sound as a strong objection against *Archie*. If, however, there is a clear prosodic boundary after *not*, then Archie is offered as an alternative to something that was decidedly rejected: *Definitely not! Archie*. Moreover, we could speculate about shallow prosodic boundary in case of addressing Archie with a vocative tag: *Definitely not, Archie*. (This third case would presume that vocatives are typically separated from the message itself, which is a rather risky premise outside the declamational style.) Similarly, in a sentence like *I thought that you invited Kate and Amy Martin* it is unclear, whether *Amy Martin* is someone who was supposed to be invited or whether *Amy* was supposed to invite *Martin*.

https://doi.org/10.14712/24646830.2019.23

In the recent decades, many studies have been devoted to investigating the prosodic cues that allow for disambiguation of analogous structures (see, e.g., Lehiste, 1973; Nespor & Vogel, 1983; Price, Ostendorf, Shattuck-Huffnagel & Fong, 1991; Pynte & Prieur, 1996; Carlson, Clifton & Frazier, 2001, etc.). It might be argued that in natural (meaning non-laboratory) settings, the disambiguation is simply provided by the context. In the example above, the interacting individuals should know, whether Amy's surname is Martin or whether she is acquainted with someone called Martin. Nevertheless, Schafer and her colleagues observed in their semi-spontaneous material that speakers signal syntactic differences with prosody even when the context fully disambiguates the structure (Schafer, Speer, Warren & White, 2000). This might suggest that competent speakers of a language use prosodic phrasing habitually to prevent misunderstanding or to serve a purpose other than disambiguation.

The question is whether the cases of potential ambiguity pose a real threat to speech communication. One might wonder how often during an ordinary working day such an ambiguous sentence is produced. We can quite probably testify that within our past few weeks' experience we have uttered thousands of propositions, yet we do not remember one that would expand the list of the above. Unclear meaning is more probably caused by incomplete information or differences in the context evoked in the minds of the provider and the receiver of the message. Does this render intonational phrasing irrelevant? Not in the least. Any time we talk to someone, this someone has to recover the meaning we are trying to convey and there is always some processing cost involved on the part of the listener. Proper phrasing can decrease the cost, the lack of thereof otherwise.

A number of experiments in the 1960s and 1970s demonstrated that speech is processed faster and its content remembered better if it is presented with clear phrasal prosody (e.g., Leonard, 1974; Martin, 1968; O'Connell, Turner & Onuska, 1968; Zurif & Mendelsohn, 1972). Many of the studies were probably inspired by the article by Epstein (1961) who showed that groups of non-words were more successfully recalled by respondents if they were presented with sentence morphology, i.e., if the non-words simulated conventional grammar. Yet, as it became clear soon afterwards, the effect of morphology in spoken stimuli only held if the non-words were presented with phrasal prosody. If presented with list prosody (i.e., as isolated items), the effect disappeared. Similarly, a list of isolated numerals is more difficult to remember than the same numerals prosodically grouped (e.g., Reeves, Schmauder & Morris, 2000).

In contrast, disrupted prosodic structure has been demonstrated to lead to longer reaction times in word, syllable or phoneme monitoring experiments (Meltzer, Martin, Mills, Imhoff & Zohar, 1976; Martin, 1979; Buxton, 1983; Tyler & Warren, 1987) or to compromise listeners' ability to retrieve the intended interpretation of an ambiguous utterance (Ferreira, Anes & Horine, 1996). Similarly, faulty turn construction causes awkward exchange of conversational floor, while proper boundary cues lead to successful transition of turns in conversations (Auer, 1996).

Another important question in the area of our present research involvement is that of syntax – prosody relationship. Due to the relatively long tradition of linguistic description, references to syntax are fairly easy to make and usually plausible enough to accept. We should always remember, however, that educational focus can cause bias. It is not necessarily true that what is taught at schools is somehow more real than what schools currently

ignore or neglect. Just because syntactic rules and units are part of elementary school syllabi, while prosodic structure is not, does not mean that prosody of real speech is in some sort of inferior position to syntax. The traditional belief that prosodic boundaries reflect syntactic structure (e.g., Selkirk, 1984; Price et al., 1991, but also, though not explicitly Kentner & Féry, 2013) is quite difficult to uphold outside the domain of laboratory speech.

Auer argued more than two decades ago that syntax and prosody do not serve one another. Rather, they complement each other to serve the communicative meaning and to manage the recipient's behaviour (Auer, 1996). Although this enlightened proposition is not as yet specific enough for precise phrase boundary predictions, there have been many attempts since to build boundary placement models for various speech materials (e.g., Cooper & Paccia Cooper, 1980; Gee & Grosjean, 1983; Taylor & Black, 1998; Parlikar & Black, 2011). Breen and colleagues discuss two fundamental options in this area of research: meaning-based approach and balance-based approach (Breen, Watson & Gibson, 2011). In their own experiments they also managed to obtain some practicable solutions, but they admit that precise modelling still requires more research. Our current study should contribute to that.

Foreign-accented speech adds one important aspect to the exploration of prosodic boundaries and that is the cognitive load. By this we mean lower-level (i.e., not intellectual) processing demands on the neurophysiology of the speaker's brain. A learner of a foreign language is constrained in the efforts to create proper prosodic phrasing, arguably by substantial detrimental processing factors (e.g., tedious search for words, uncertainty about morphosyntactic rules, or neurophysiological planning of articulatory gestures in phonotactically unfamiliar sequences). In foreign-accented speech, the prosodic boundaries can be involuntary or unplanned – the speaker just has to break the speech continuum when he struggles with the actual cognitive constraints. The results of this labour have to be mapped too for at least two reasons. First, the prosodic phrasing in foreign-accented speech must be eventually tested perceptually in rigorously planned experiments if we want to identify individual factors that impact on the listener. Second, various attempts to understand speech mechanisms have led to the appreciation of the fact that the devil is in detail. This study is motivated by ambition to provide clear, contextually grounded detail that will find its use in future research.

2. Method

The sample of professional native speakers was represented by news readers from respectable national radio stations. It was the BBC for English and Czech Radio (Český rozhlas) for the Czech language. Twelve established news readers (6 men + 6 women) were recorded for each language directly from a broadcast of news bulletins. (The professional experience of individual speakers was ascertained on the web pages of the respective radio stations.)

News reading exemplifies the so-called 'clear speech' – the speaking style used outside ordinary conversational settings, usually under special acoustic or social conditions. The use of clear speech in news reading is understandably essential due to the lack of visual cues for the listeners, the limited amount of shared context between the speaker and the

listeners, and due to relative semantic and syntactic complexity of the texts. Our tentative presupposition is that clear speech manifests prosodic structures more explicitly than common conversational speech thanks to greater production efforts exerted during its production (see also Dellwo et al., this volume).

The news bulletins were quite similar in form for both languages. They comprised 7 to 8 paragraphs (news items) with initial, final and occasionally medial greetings or contact phrases. The mean number of words in the English bulletins was 505, in the Czech bulletins it was 517.

The sample of professional news readers was complemented with twelve Czech speaking non-professionals: university students of 19–23 years of age (8 women + 4 men) who were asked to read out the text of one of the news bulletins in a recording studio. The students were given sufficient time to familiarize themselves with the text. They were well acquainted with both the recording studio and the experimenter who was present. Hence, we expect little impact of nervousness or performance anxiety. These non-professional readers were also advised to make longer pauses between the consecutive paragraphs to avoid performance stress.

Finally, the foreign accent bearers were Anglophone speakers (6 women + 6 men) living and working or studying in the Czech Republic for at least a year with proficiency in the Czech Language of at least B2 of CEFRL (Common European Framework of Reference for Languages). The length of residence was established in an interview after the recording, but since it did not correlate even remotely with the language proficiency, we do not report it.

In the graphs and tables below, the Czech and English professional speakers will be referred to as CzP and EnP respectively, the Czech non-professionals will be CzN, and the speakers of English-accented Czech will be represented by ECN.

Individual recordings were processed in Praat software package (Boersma & Weenink, 2019). The text was aligned with the sounds, and position of individual phones and words was estimated with Prague Labeller (Pollák, Volín & Skarnitzl, 2007) followed by manual corrections of boundaries. Syllabic peaks were established in a special tier with a Praat script.

Prosodic boundaries (or breaks) were located through auditory inspection. Two levels of division were sought, both compatible with ToBI break indices conventions (Price et al., 1991; Beckman & Ayers Elam, 1997) and with other similar recommendations (e.g., Xu, 2011). In this study the break index 4 (BI4) will be referred to as *major phrase boundary.* (Major prosodic phrase is called intonation phrase in some texts.) Such prosodic boundary is indisputable as it is signalled by multiple cues, especially by a very clear F0 pattern, decrease in tempo (i.e., lengthening of the phrase-final syllable or two), occasionally accompanied with a declination reset, change in phonation and amplitude, or specific juncture phenomena and pauses.

Minor phrase boundary (minor prosodic phrase is called intermediate phrase in some texts) is equivalent to what the ToBI transcription system labels as break index 3 (BI3). Such boundaries lack either the phrase-final lengthening or clear F0 pattern, or they may display weakened version of both. The BI3s also leave quite unambiguous feeling of discontinuity, but they require immediate restoration of the flow of speech. Thus, for instance, it would be unnatural to place a silent pause after them.

There were four groups of speakers (altogether 26 female and 22 male subjects) and the following four research questions were asked.

Research Question 1 – What is the mean length of a prosodic phrase in syllables
a) in Czech professional presentation of spoken texts?
b) in English professional presentation of spoken texts?
c) in Czech non-professional renderings of spoken texts?
d) in English-accented renderings of Czech spoken texts?

Research Question 2 – What is the mean length of a prosodic phrase in words?
with a), b), c) and d) subspecifications as above

Research Question 3 – What is the proportional representation of major and minor prosodic breaks in the spoken texts?
with a), b), c) and d) subspecifications as above

Research Question 4 – Is there any correlation between the articulation rate and prosodic phrase length?

To extract information about numbers of syllables, words, prosodic phrases and articulation rates, the scripting facility of the Praat software was used. Where appropriate, testing of statistical significance of differences was related to conventional $\alpha = 0.05$.

3. Results

Figure 1 presents the graphic answer to the research question 1a. Czech professional speakers produced phrases of 10.78 syllables on average. The longest phrases were made by the female speaker CzP02 – 13.2 syllables per phrase. Incidentally, the shortest phrases were also produced by a female speaker. CzP03 only used 8.7 syllables per phrase on average. The outcome then suggests that the male speakers form a more homogenous group, but since the size of the subgroup is only 6 individuals, this fact should not be overemphasized.

Figure 2 provides a set of results that is analogical to the previous one, but describes the phrase production in the groups of English professional news readers. Thanks to the

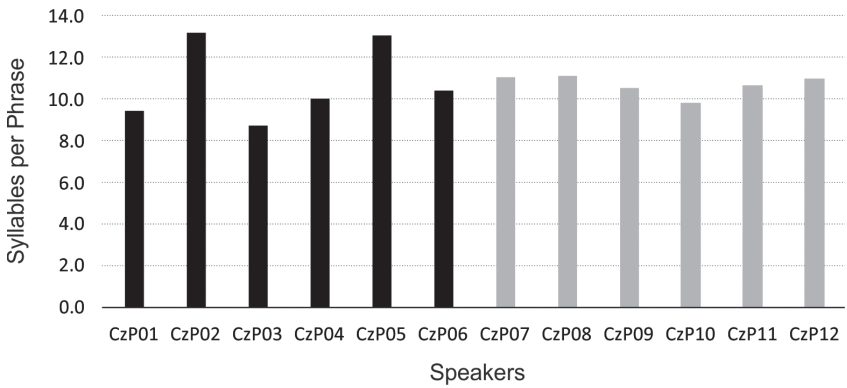

Figure 1. Mean lengths of major prosodic phrases in syllables by individual Czech professional speakers. Black columns = women, grey columns = men.

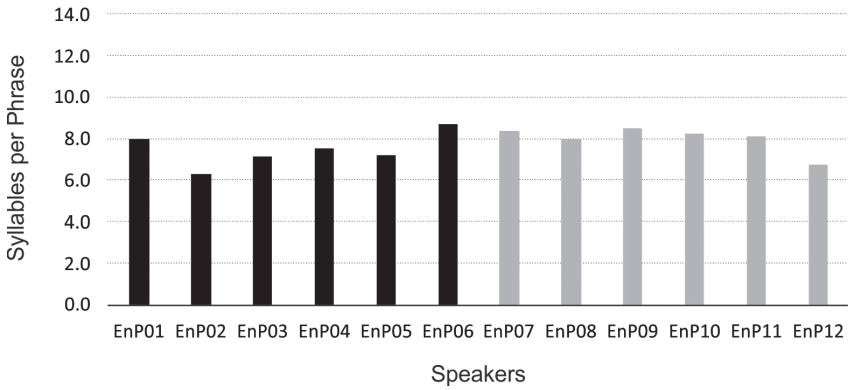

Figure 2. Mean lengths of major prosodic phrases in syllables by individual English professional speakers. Black columns = women, grey columns = men.

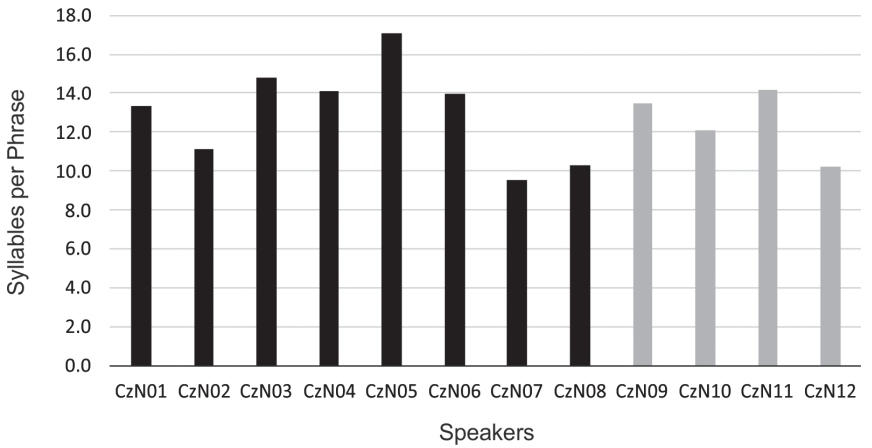

Figure 3. Mean lengths of major prosodic phrases in syllables by individual Czech non-professional speakers. Black columns = women, grey columns = men.

identical scaling of both graphs it is immediately noticeable that the English professionals produced shorter phrases – the mean length across the sample was only 7.76 per phrase, i.e., by three syllables fewer than in the Czech news reading. Interestingly, the highest and the lowest values were again produced by women: EnP06 produced phrases of 8.8 syllables on average, while EnP02 used merely 6.3 syllables. Similarly to the situation in the Czech professional sample, the values provided by men are again more balanced (with the same caveat).

Quite surprisingly, the axis scaling for the phrase production of the Czech non-professional speakers had to be changed (Figure 3). While the Czech professional news readers made prosodic phrases of 10.78 syllables (see above, Fig.1), the non-professional speakers reached the mean length of 12.89 syllables. Nine of the twelve non-professional speakers produced values above the mean of the Czech professionals. Figure 3 also exposes the lon-

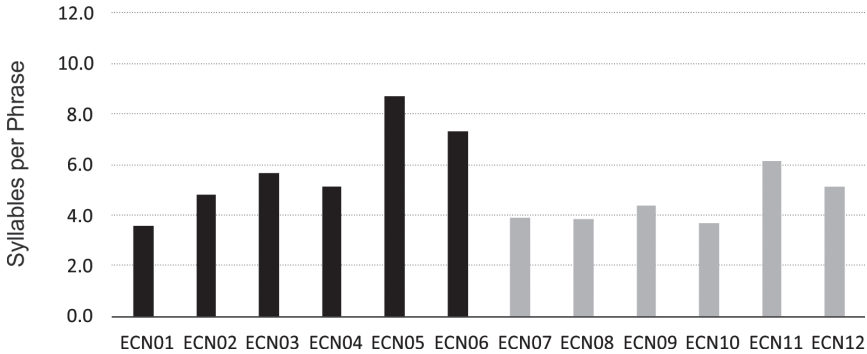

Figure 4. Mean lengths of major prosodic phrases in syllables by individual English speakers of Czech. Black columns = women, grey columns = men.

gest phrases in the sample: the speaker CzN05 created phrases with mean length of 17.1 syllables. That is almost 4 syllables more than the maximum in the Czech professional group. The shortest phrases were also produced by a woman: CzN07 delivered mean length of 9.5 syllables. (It has to be noted, though, that this sample is unbalanced gender-wise.)

The graph in Figure 4 had to be rescaled as well, but this time quite predictably. Based on everyday experience, foreign-accented speech can be anticipated to be more fragmented. This was, indeed, the case: the mean length of a prosodic phrase was only 5.22 syllables. Four of the twelve speakers even produced values under 4 syllables per phrase, three, on the other hand, exceeded 6 syllables per phrase. Yet again, it seems that the male speakers form a more homogenous group (and yet again we warn against over-generalizations from small samples).

Table 1. Mean length and variation of prosodic phrases in Czech professional news reading (CzP), English professional news reading (EnP), Czech non-professional news reading (CzN), and Czech spoken by Anglophone foreigners (ECN). Values of the arithmetic means and standard deviations are in syllables per phrase, coefficients of variation are percentages.

	CzP	***EnP***	***CzN***	***ECN***
mean	10.78	7.76	12.89	5.22
std. dev.	1.31	0.75	2.23	1.57
C_{var} *(%)*	12.15	9.66	17.30	30.07

Table 1 summarizes the results for the Research Question 1. It shows that in terms of syllables per phrase the longest units were produced by Czech non-professionals. These were followed first by Czech, and then by English professional speakers. As expected, the foreign accented Czech consisted of the shortest phrases. Although this study is designed to provide descriptive data and not to test hypotheses, one-way ANOVA was computed to ascertain the significance of the differences. The outcome was highly significant: $F(3, 44) = 55.83$; $p < 0.001$. (The general α was set to 0.05 – see Method.) Post-hoc Tukey HSD test returned high significance of all the differences between the four conditions.

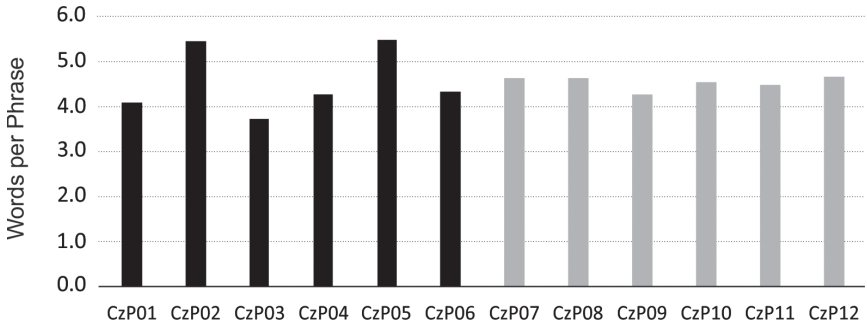

Figure 5. Mean lengths of major prosodic phrases in words by individual Czech professional speakers. Black columns = women, grey columns = men.

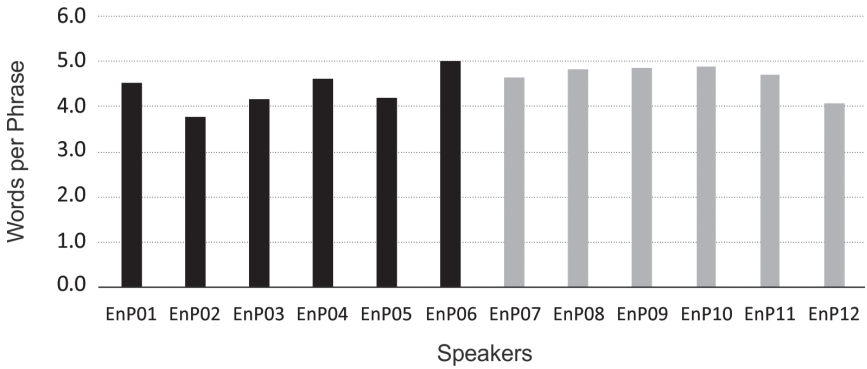

Figure 6. Mean lengths of major prosodic phrases in words by individual English professional speakers. Black columns = women, grey columns = men.

As to the variance in the data, Czech non-professionals produced the largest standard deviation, but after normalization by mean (i.e., computation of the coefficient of variation) the foreign-accented speech turned out to be most variable.

Figures 5 and 6 present mean lengths of prosodic phrases measured in words as produced by Czech and English professional news readers respectively. They are pertinent to Research Question 2 above (see final part of Method). Interestingly, despite the substantial difference between values expressed in syllables per phrase (Fig. 1 and 2), there is practically no difference in lengths expressed in words per phrase. The Czech grand mean is 4.56 and the English one is 4.54 words per phrase. The original difference of 3 syllables per phrase translates into negligible 0.02 words per phrase. This implies that semantic constraints are very similar for both languages, provided the word is the natural semantic building block. Structural constraints obviously differ. The explanation that offers itself first rests in the fact that Czech words are longer due to the rich inflectional system. In other words, there are much fewer monosyllables in Czech texts. There might also be the syllable phonotactics involved: Czech syllables avoid codas to much greater extent than the English ones. (In terms of syllable onsets, the complexity is comparable.)

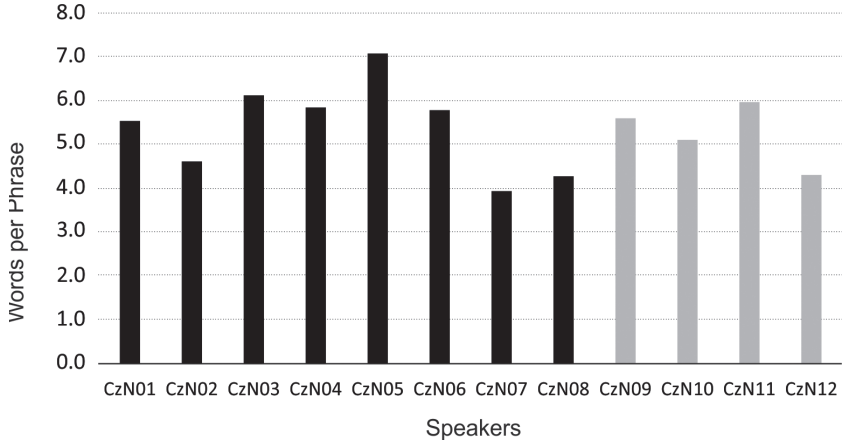

Figure 7. Mean lengths of major prosodic phrases in words by individual Czech non-professional speakers. Black columns = women, grey columns = men.

Figure 7 displays the mean values for the non-professional Czech speakers. The grand mean for this group exceeds both professional groups: it is 5.35 words per phrase. The non-professionals make their phrases by almost one word longer than Czech and English skilled news readers. There is, however, substantial variation within the non-professional group: the female speaker CzN05 makes phrases almost twice as long as the speaker CzN07.

Finally, but crucially for the primary motivation of this study, we have measured the lengths of prosodic phrases produced by Anglophone speakers of Czech. The results are displayed in Figure 8. The grand mean across the whole group is 2.19 words per phrase. That is less than half of the mean length produced by both Czech and English professional speakers (see Fig. 5 and 6). If we disregard speakers ECN05 and ECN06, the grand mean drops to 1.95 words per phrase. This signals quite a substantial number of phrases consisting of one word only, which contributes to the disfluent character of the foreigners' speech production. Indeed, out of 3132 prosodic phrases produced by ECN speakers there were 1722 (55%) containing just one word, of which 176 (about 10%) were monosyllables.

Table 2 summarizes the results for the Research Question 2. One-way ANOVA returned a highly significant effect: $F(3, 44) = 53.15$; $p < 0.001$, and post-hoc Tukey HSD test found no difference between English and Czech professionals, but both these groups were significantly different from Czech non-professionals and foreigners speaking Czech. Variation in the data is analogical to that of lengths in syllables per phrase (see C_{var} in Table 1 above).

Table 2. Mean length and variation of prosodic phrases in Czech and English professional samples (CzP and EnP respectively), Czech non-professional group (CzN), and Czech spoken by Anglophone foreigners (ECN). Values of the arithmetic means and standard deviations are in words per phrase, coefficients of variation are percentages.

	CzP	***EnP***	***CzN***	***ECN***
mean	4.56	4.54	5.35	2.19
std. dev.	0.51	0.39	0.92	0.66
C_{var} *(%)*	11.13	8.53	17.17	30.19

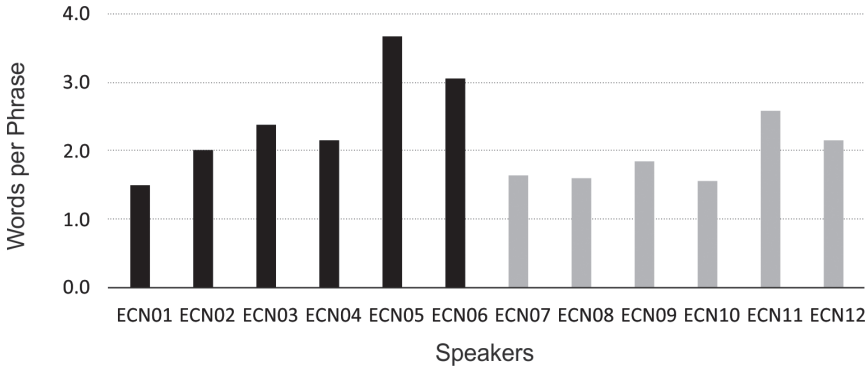

Figure 8. Mean lengths of major prosodic phrases in words by individual English speakers of Czech. Black columns = women, grey columns = men.

The Research Question 3 asked about the relative proportions in the occurrence of minor and major prosodic breaks. Table 3 provides the answer to that. As explained above, the metric chosen for the ratio is the percentage of major prosodic boundaries from the entire set of boundaries. (Since only break indices BI3 and BI4 were included, 75% of major breaks, for instance, would mean 25% of minor breaks). Table 3 indicates that the relative incidence of major breaks was very similar across the four groups of speakers, specifically about 78%, which means that about 22% of the boundaries found were minor phrase boundaries. One-way ANOVA found the actual differences clearly insignificant ($p \approx 0.608$).

Table 3. Mean occurrences of major prosodic boundaries expressed as a percentage of the whole set of boundaries (i.e., major and minor – see Method).

	CzP	***EnP***	***CzN***	***ECN***
major phrase (%)	76.6	80.1	78.4	78.0
std. dev. (%)	6.5	7.3	6.1	5.4
C_{var}*(%)*	8.5	9.1	7.7	6.9

The fourth and final Research Question asked about a relationship between the mean length of the phrase and articulation rate. The Pearson correlation was found as very high: $r = 0.89$ and highly statistically significant ($p < 0.001$). Figure 9 depicts the situation with 48 data points, i.e., each speaker is represented by one data point.

The high correlation coefficient is clearly influenced by non-homogeneity of the whole assembly, especially by the behaviour of the group of foreigners speaking Czech (in the lower left part of the graph). After their exclusion, the correlation drops to $r = 0.51$, but stays highly significant nonetheless ($p < 0.001$). Faster talkers then produce longer prosodic phrases. This comes as no surprise, but it should be remembered that the primary objective of this study was not to discover new trends but, instead, to provide reliable quantitative data about Czech, English and English-accented Czech.

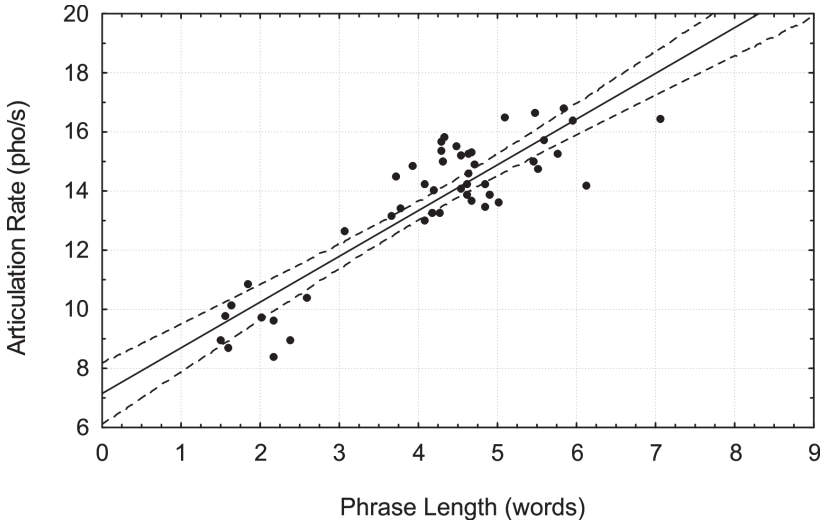

Figure 9. Scatterplot of 48 data points (1 for each speaker) capturing the relationship between articulation rate (in phones per second) and mean length of a prosodic phrase (in words).

4. Discussion

As expected, foreign-accented speech was found considerably more broken than native speech. L1 professionals, who are often considered 'model speakers', made about 22 major prosodic boundaries in every 100 words in our sample, whereas L2 speakers made more than 46 of those. Jun's observation that all languages use prosodic grouping even if different languages use them in very different ways (Jun, 2005), can be expanded then: even various groups of the same language users may build phrases in different ways.

Our material showed that a major prosodic phrase in foreign-accented speech very often consisted of one word only and cases when this was a monosyllabic grammatical word were not exceptional. The resulting impression of such fragmentations is typically that of struggle. Our results provide some practical guidance for future perception experiments, which should address the impact of abundant phrase boundaries on the listener. Ultimately, the impact of one's speech is what matters in everyday interactions in multilingual environment (*cf*. Lev-Ari & Keysar, 2010).

Somewhat unexpectedly, our sample of non-professionals produced significantly longer phrases than skilled speakers. Even though this trend is in the opposite direction than that of foreign-accented speech, it is quite probably not advantageous either. Larger continua of speech can pose extra demand on cerebral processing and listeners may find them as tiresome as the texts that do not 'hold together'. However, this can only be hypothesized if we accept the premise that professionals master the language better on all levels. The hypothesis still deserves empirical testing in the future perception experiments that should focus both on comprehension and on memory retention under different phrasing conditions.

One of the interesting details revealed in this study is that although foreigners make many more boundaries in spoken texts, the proportion of major and minor prosodic breaks is virtually the same as in other groups of speakers (see Table 3). This finding should be tested in other speech styles and genres. A speculative interpretation of this fact could state that the minor prosodic boundary is just an auxiliary agency while the major break is the principal way of prosodic grouping. Minor phrases are then used when an occurring semantic unit is too large and needs some sort of internal structure together with the necessity of preserving unity. Another possible explanation could be based on the ambiguity of the minor prosodic break. Analogically to the metrical structure of English, where despite the existence of secondary stresses and full unstressed syllables speakers evidently prefer primary stresses and reduced unstressed syllables in continuous speech, there might be a preference for a clear boundary or no boundary at all in prosodic 'chunking'. Be that as it may, the speech style of read-out news led to more than three quarters of all prosodic phrase boundaries being major.

The study also highlighted a thought-provoking relationship between two ways of measuring the size of prosodic phrases. Even though the results for two typologically disparate languages, Czech and English, differed in numbers of syllables per phrase, they were virtually identical in numbers of words per phrase. If we consider the syllable a basic structural unit, and the word a primary semantic unit, then our finding contributes to the debates on the relationship or interplay between the form and the meaning. It seems that our cognitive mechanisms are less constrained in terms of formal structures but more fastidiously tuned to certain 'amount of the meaning' (*cf*. Caplan & Waters, 1999 or, for instance, Hirotani, Frazier & Rayner, 2006). Naturally, our simple study does not allow for any far-reaching conclusions in this area, but rather invites experimenters with other language backgrounds to take read-out monologues (ideally news bulletins for direct comparison) and to try to replicate our measurements.

Future research should also investigate the acoustic and syntactic nature of the phrase boundaries. Even though informal observations suggest that the phonetic means of prosodic boundary markings are analogous in Czech, English and English-accented Czech, a detailed acoustic analysis might uncover interesting differences. The syntactic disparities, on the other hand, are quite obvious: foreign-accented Czech exhibits, for instance, some unusual breaks between the adjective and the modified noun, between the preposition and the following noun, or even between the first name and the surname of a person. The frequency of occurrence and other circumstances of such and similar cases should be known before further perceptual testing.

ACKNOWLEDGEMENT

This study was supported by MUP Project No. 68-01 „Political sciences, culture, media and language" funded by the Institutional Support for Long-Term Strategic Development of Research Organisations in 2019.

REFERENCES

Auer, P. (1996). On the prosody and syntax in turn-continuations. In: E. Couper-Kuhlen & M. Selting (Eds.) *Prosody in Conversation*, pp. 57–100. Cambridge: Cambridge University Press.

Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI Labelling, version 3*. The Ohio State University Research Foundation, Ohio State University.

Boersma, P. & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program], Version 6.0.47, retrieved 8 February 2019, http://www.praat.org.

Breen, M., Watson, D. G. & Gibson, E. (2011). Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes* 26(10), pp. 1532–1562.

Buxton, H. (1983). Temporal predictability in the perception of English speech. In: A. Cutler & D. R. Ladd (Eds.) *Prosody: Models and measurements* (pp. 111–121). Heidelberg: Springer-Verlag.

Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22(1), 77–94.

Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45(1), 58–81.

Cooper, W. E. & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Dellwo, V., Pellegrino, E., He, L. & Kathiresan, T. (2019). The dynamics of indexical information in speech: Can recognizability be controlled by the speaker? *Acta Universitatis Carolinae – Philologica XX, Phonetica Pragensia XV*, pp. 57–75.

Ferreira. F, Anes, M. D. & Horine, M. D. (1996). Exploring the use of prosody during language comprehension using the auditory moving window technique. *Journal of Psycholinguistic Research* 25(2), 273–290.

Gee, J. P. & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15(4), 411–458.

Hirotani, M., Frazier, L., Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language* 54(3), pp. 425–443.

Jun, S. A. (2005). Prosodic typology. In: S. A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 430–458, Oxford University Press.

Kentner, G. & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, 30(2), 1–35.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107–122.

Leonard, L. B. (1974). The role of intonation in the recall of various linguistic stimuli. *Language and Speech*, 16(4), 327–335.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096.

Martin, J. G. (1968). Temporal word spacing and the perception of ordinary, anomalous, and scrambled strings. *Journal of Verbal Learning and Verbal Behaviour*, 7(1), 154–157.

Martin, J. G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America*, 65(5), 1286–1297.

Meltzer, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L. & Zohar, D. (1976). Reaction time to temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, 2(2), 277–290.

Nespor, M., & Vogel, I. (1983). Prosodic structure above the word. In: A. Cutler & D. R.

Ladd (Eds.), *Prosody: Models and measurements*, pp. 123–140. Heidelberg: Springer-Verlag.

O'Connell, D. C, Turner, E. A. & Onuska, L. A. (1968). Intonation, grammatical structure, and contextual association in free recall. *Journal of Verbal Learning and Verbal Behaviour*, 7(1), 110–116.

Parlikar, A., & Black, A.W. (2011). A grammar based approach to style specific phrase prediction. In: *Proceedings of INTERSPEECH 2011*, Florence, Italy: ISCA, pp. 2149–2152.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation* [Ph.D. dissertation]. MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington].

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In: *Proceedings of XIIth Speech and Computer – SPECOM 2007*, pp. 537–541.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America*, 90(6), 2956–2970.
Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence processing. *Language and Cognitive Processes*, 11(1–2), 165–192.
Reeves, C., Schmauder, A. & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology – Learning, Memory, and Cognition*, 26(6), pp. 1638–1654.
Selkirk, E. (1984). *Phonology and Syntax*. Cambridge, Mass.: MIT Press.
Schafer, A. J., Speer, S. R., Warren, P. & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29(2), 169–182.
Taylor, P. & Black, A. W. (1998). Assigning phrase breaks from part-of-speech sequences, *Computer Speech and Language* 12(2), 99–117.
Tyler, L. K. & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language* 26(6), 638–657.
Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences* 1(1), 85–115.
Zurif, E. B. & Mendelsohn, M. (1972). Hemispheric specialization for the perception of speech sounds: The influence of intonation and structure. *Perception & Psychophysics* 11(5), 329–332.

RESUMÉ

Primárním cílem této studie je poskytnout kvantitativní data týkající se délky promluvového úseku (prozodické fráze) v češtině s cizineckým přízvukem. Řečová produkce anglofonních mluvčích, kteří si osvojují češtinu jako cizí jazyk (L2) je porovnávána s češtinou profesionálních i neprofesionálních mluvčích, ale také s angličtinou jakožto rodným jazykem cílové skupiny. Každý ze čtyř vzorků mluvčích je reprezentován dvanácti mluvčími (tj. celkově $n = 48$). Materiálem jsou čtené monology, konkrétně rozhlasová zpravodajství, která reprezentují mluvu k neznámému publiku a jsou charakteristická požadavkem zřetelnosti.

Podle očekávání byly promluvové úseky v řeči s cizineckým přízvukem kratší a naše data poskytují konkrétní základ pro modelování tohoto rysu v percepčních testech. Tak např. 55 % promluvových úseků vyprodukovaných cizinci bylo jednoslovných a 10 % z těchto jednoslovných bylo dokonce jednoslabičných.

Jedním ze zajímavých výstupů studie je také zjištění, že čeští hlasatelé (profesionální mluvčí) produkují delší promluvové úseky než hlasatelé angličtí co do počtu slabik (v průměru 10.8 slabik na úsek v češtině proti 7.8 slabikám na úsek v angličtině), ale tento rozdíl se vytratí, pokud se délka promluvového úseku vyjádří v počtu slov na úsek (4.56 a 4.54 slova na úsek). Tento výsledek naznačuje, že v daném mluvním stylu je délka promluvového úseku (prozodické fráze) nejspíše vymezena sémanticky. Slovo je totiž jednotkou sémantickou, zatímco slabika jednotkou strukturní. Strukturní rozdíl tří slabik mezi češtinou a angličtinou odpovídá sémantickému rozdílu 0,02 slova.

Studie také přináší doklad o relativně stabilním poměru mělčích a hlubších prozodických předělů u všech čtyř sledovaných skupin mluvčích.

Jan Volín
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: jan.volin@ff.cuni.cz

2019 ACTA UNIVERSITATIS CAROLINAE PAG. 159–167
PHILOLOGICA 2/ PHONETICA PRAGENSIA

THE ELECTROPALATOGRAPHIC STUDY OF THE COARTICULATORY EFFECT OF VOWELS ON CORONAL STOPS IN PERSIAN

MARAL ASIAEE, MANDANA NOURBAKHSH,
SAEED RAHANDAZ

ABSTRACT

Using electropalatographic (EPG) data, we study the coarticulatory effect of intervocalic contexts on the Persian coronal stops [t] and [d]. The EPG patterns demonstrate that [d] is produced in a more anterior place than [t], proving the former to be a dentialveolar consonant and the latter to be an alveolar one. The coarticulation index (CI) is calculated for each consonant flanked by the same vowels. The results obtained show that there is no significant difference between [t] and [d], in terms of coarticulation; however, based on the data we have, we can say that [t] is more resistant to coarticulatory effect than [d]. This result is in agreement with previous investigations which propose that laminals show stronger coarticulation resistance than apicals.

Key words: coarticulatory effect, coarticulation index, electropalatography, coronal stops, Persian

1. Introduction

Phoneticians and phonologists use place of articulation as one of the most important parameters when they describe segments. The phonetic representation of a phoneme is not always the same and it may vary according to the adjacent segments or its position in the syllable. This mostly occurs because of the overlapping gestures of the two neighboring segments. The change that occurs in articulation and acoustic signal of any segment due to its adjacent segments is called coarticulation; however, not all segments allow the same degree of coarticulation. Some are more resistant. Bladon and Al-Bamerni (1976) – the first scholars who originally proposed Coarticulatory Resistance (CR) – postulated that a numerical CR value can be designated to segments and their extrinsic allophones and this CR value can be used in the speech production mechanism to plan the coarticulatory directionality and magnitude for articulators.

Using their ears, linguists can detect some coarticulations which have traditionally been called allophones; but, with quantitative instrumental investigation, they can detect more detailed and meticulous variations (Kühnert and Nolan, 1999: 7). EPG is one of the

https://doi.org/10.14712/24646830.2019.24

instruments that can be used to assess whether the place of articulation of a phoneme has changed due to the coarticulatory effect.

EPG is a technique which depicts the contact between the tongue and the hard palate. The technique was first used by Grützner (1879) and since then it has gained popularity amongst phoneticians and speech therapists. In EPG, an artificial palate with 62 electrodes embedded on its lingual surface specially made for each individual is put against the hard palate to record the tongue/ palate contact. These 62 electrodes are arranged in 8 rows, each of which has 8 electrodes except the first row. The schematic arrangement of electrodes with their corresponding places of articulation is shown in Figure 1.

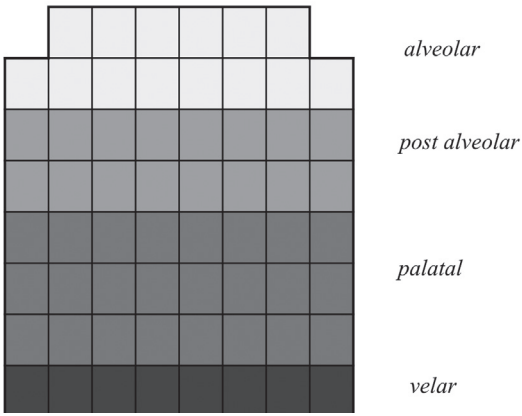

Figure 1. Schematic electrode arrangements with their corresponding places of articulation.

Numerous studies have investigated the coarticulatory effect using EPG. Some of the most extensive and comprehensive studies in this field were done by Recasens (1983, 1984a, 1984b, 1987, 1989). He compared Spanish and Catalan lingual consonants with different places and manners of articulation by means of EPG. He concluded that there was an inverse relationship between the coarticulatory effect and the degree of tongue dorsum elevation required for the consonants; that is, a progressive decrease in the degree of the tongue dorsum contact with the palate causes a progressive increase in the coarticulatory effect (Farnetani, 1989). His findings also indicated that those segments produced with an elevated tongue dorsum, were the ones most resistant to coarticulation and any contextual variation, but they were at the same time, the segments that caused the neighboring segments to vary the most.

Data from other languages also confirm Recasens' findings. Butcher and Weiher (1976) and Farnetani et al. (1985) studied German and Italian respectively and found [i] to be the most resistant vowel to the coarticulatory effect amongst other vowels; however, it was the most aggressive (or dominant) vowel in influencing the adjacent segments.

Farnetani (1989) studied coarticulation of VCV sequences for lingual consonants. She proposed a coarticulation index (CI) which was defined palatographically. Coarticulation index (CI) shows the contextual effects on a segment; that is, it shows variation both in the position of the contact and the amount of the contact (Farnetani, 1989: 113). She studied the effect of place of articulation and voicing on laterals and dentoalveolar stops in Italian. She also examined the coarticulatory effect on consonants flanked by stressed and unstressed

vowels. The data obtained indicated that the coarticulation of tongue body varies inversely with the degree of tongue dorsum elevation. The CI decreased from alveolars to palatals. Within the alveolar category, CI decreased from laterals to voiced and voiceless stops.

Zharkova (2008) used EPG and ultrasound to investigate the lingual coarticulation in vowel- consonant sequences. She used an EPG measure and an ultrasound measure to compute the difference between /p, f, t, s, l, r, k/ in /a, i/ contexts in Scottish English. Results showed that labial consonants and /r/ were the most affected segments.

Chen, Chang & Iskarous (2015) studied the speech of seven Taiwan Mandarin speakers. Their data consisted of CV syllables. Their results indicated that the high front vowel, [i], was more resistant to the coarticulatory effect than [a, u].

In this survey, we use EPG to compute the coarticulation index of Persian coronal stops and study the effect of different vowels on them. Persian belongs to the Indo-Iranian branch of Indo-European languages. It has six vowels /ɪ, e, a, ɑ, o, u/ and 23 consonants including eight plosives /p – b, t – d, c – ɟ, ɢ, ʔ/, eight fricatives /f – v, s – z , ʃ – ʒ , x, h/, two affricates /t͡ʃ, d͡ʒ/ and five sonorants /m, n, j, l, r/. Previous impressionistic studies on Persian consider the place of articulation of coronal stops, namely [t] and [d], to be either dental, alveolar or dentialveolar (Mahootian, 1997; Nourbakhsh, 2009; Modarresi Ghavami, 2013 and 2018: 95; Bijankhan, 2018: 112). However, using EPG, Asiaee, Nourbakhsh and Skarnitzl (2018) showed that there was an asymmetry between the places of articulation of these speech sounds; they found [d] to be a dentialveolar stop, whereas [t] was an alveolar stop. Moddaresi Ghavami (2018) mentions the effect of the places of articulation of vowels on the production of their neighboring dorsal consonant – [c, ɟ]; however, there have been no studies on the effect of vowels on coronal stops in Persian. This study aims to tackle this subject.

2. Data and Method

The speech was recorded at the Phonetic laboratory at Alzahra University, Tehran, Iran. The speaker was asked to wear the palate 30 minutes prior to the actual recording session to minimize unwanted possible effects of wearing an artificial palate. Using the EPG system by Rose Medical, we got the linguopalatal contact at the sampling rate of 100Hz, we also recorded the audio at the same time with the sampling rate of 16000 Hz. We asked a female Persian speaker to read 12 disyllabic Persian words and pseudowords, three times. The words consist of coronal stops -[t] and [d]- located in intervocalic contexts where the two flanking vowels were the same. Persian has three front and three back vowels. Figure 2 shows the vowel space of Persian.

After the data was recorded, we used icSpeech Professional, a software which was provided by Rose Medical along with the EPG system to determine the onset and offset of the coronal stops in the intervocalic contexts.

To study the effect of vocalic coarticulation on the tongue-palate contact pattern, we needed a dataset in which coronal stops were flanked by the same vowels. The data that we collected fulfilled this need. To do so, coarticulation index was computed for [t] and [d] in each intervocalic context. Coarticulation index is calculated as the mean absolute difference between the percentages of contacted electrodes in all rows for each context.

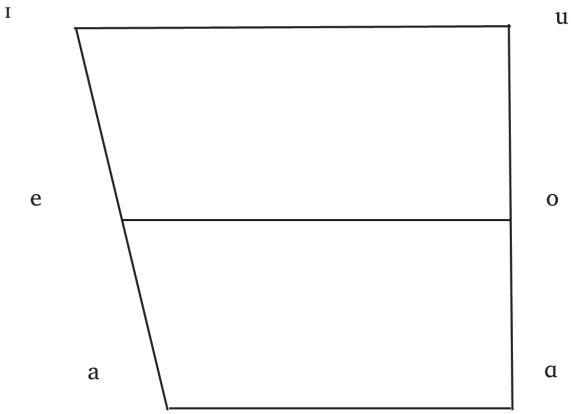

Figure 2. Persian vowel space (form Bijankhan, 2013: 136).

3. Results and Discussion

Figure 3 presents the EPG pattern for [t] and [d] as a mean across all repetitions and all intervocalic contexts.

The first row was consistently contacted in [d], which is expected from a dentialveolar segment. The EPG template and the CA index for the place of articulation of [t] and [d] have shown that the former speech sound is an alveolar stop and the latter a dentialveolar one (for a full review see Asiaee et al., 2018).

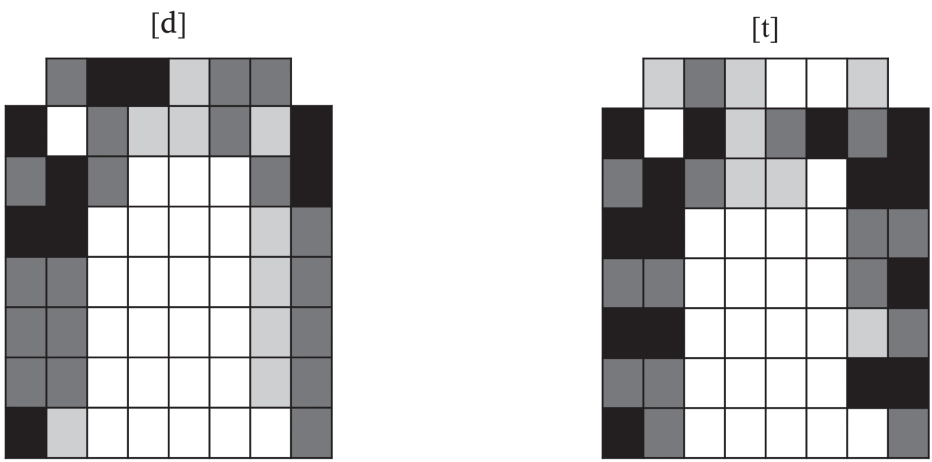

Figure 3. Mean of contacted electrodes across all three repetitions and all intervocalic contexts (black squares correspond to more than 90% electrodes being contacted, dark grey squares to 50–90 %, light grey squares to 20 – 50% contacted electrodes).

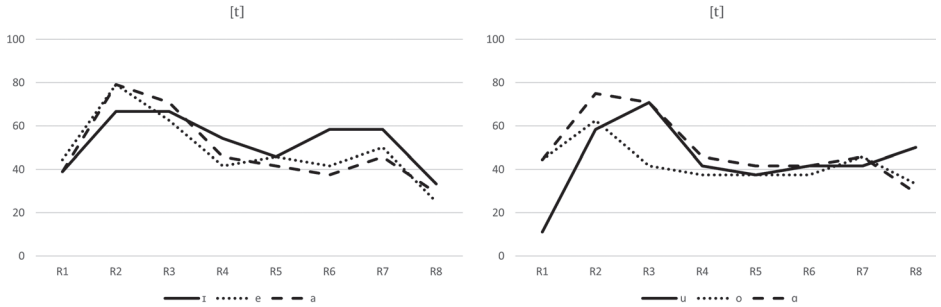

Figure 4. Percentages of contacted electrodes when [t] is flanked by Persian front vowels (left); Persian back vowels (right).

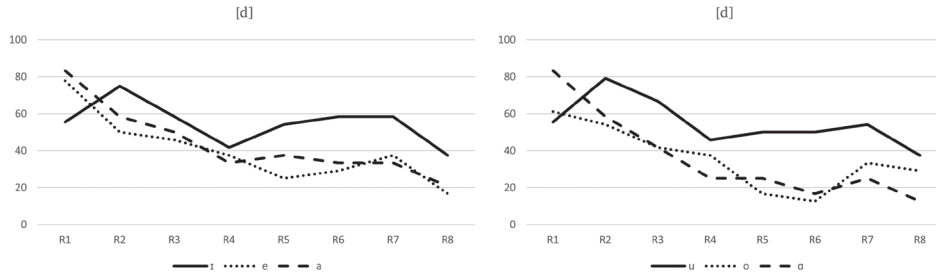

Figure 5. Percentages of contacted electrodes when [d] is flanked by Persian front vowels (left); Persian back vowels (right).

Figures 4 and 5 illustrate the mean percentage of contacted electrodes for [t] and [d] in different intervocalic contexts. Figure 4 shows the percentage of the contacted electrodes in each row where [t] is flanked by front and back vowels. Since [t] had the same percentage of contacted electrodes in both [ɪ] and [a] contexts the absolute difference between these two intervocalic contexts was zero. In the second row of [t], the percentage of contacted electrodes was stronger than the first row in all three intervocalic contexts. As we moved backward, the number of contacted electrodes decreased gradually except for the [ɪ] context- in which there was an increase in rows 5 to 7.

One possible reason for this pattern is the fact that [ɪ] is a high vowel and as Recasens (1991: 179) mentions there is a decrease in the linguopalatal contacts in a progression from higher to lower front vowels all over the palate surface. In the context of back vowels, [t] was produced in a posterior place. The number of contacted electrodes was the smallest in the context of [u] and it was more prominently drawn backward. It can be explained by the fact that according to Farnetani (1989) and Recasens (1987, 1989), the more elevated the tongue dorsum is, the more resistant the segment is to coarticulation, and the more aggressive that segment is in affecting the neighboring sounds. Hence, being a [+high] vowel, [u] affects the neighboring sounds more than a [-high] vowel. The absolute difference for [t] in the contexts of front and back vowels were 1.88, 1.62, 1.75, 1, 0.66, 1.16, 0.95 and 1.41 respectively from row one to row eight.

The coarticulation index for this speech sound which is calculated as the mean absolute difference was 1.30.

Figure 5 illustrates the percentage of the contacted electrodes in each row where [d] is flanked by Persian front and back vowels. [d] was produced in a slightly more anterior position than [t]. The percentage of contacted electrodes were significantly higher in the first row for [d]. With the exception of the first row, the percentages of contacted electrodes in all rows was higher in the [ɪ] context than the two other front vowels.

The high back vowel, [u], had almost the same pattern as [ɪ]. [o] and [ɑ] which are the mid and low back vowels respectively, affected [d] to be produced in a more posterior place, in comparison to [d] in the context of mid and low front vowels. The absolute differences for row one to row eight were 2.38, 2.08, 1.79, 1.29, 2.75. 3.33, 2.33 and 1.95. The coarticulation index for [d] was hence 2.24.

Since there was a noticeable difference between the coarticulation index values of [t] and [d], we examined the CI values of [t] and [d] in each intervocalic context separately

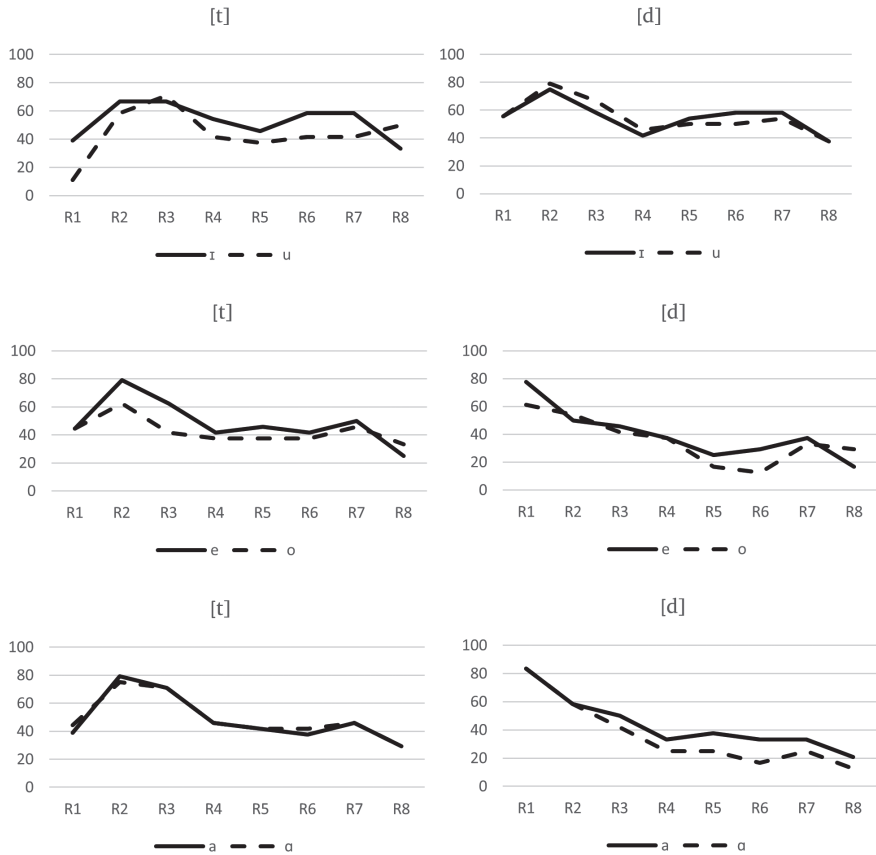

Figure 6. Percentages of contacted electrodes for [t] and [d] in each [-back]/[+back] pair of Persian vowels.

and conducted an independent sample t-test to compare the CI values for them. The Results obtained showed that there was no significant difference in the CI values for [t] (mean = 0.449722, SD = 0.0598) and [d] (mean = 0.418650, SD = 0.1145); t (10) = 0.423.

For a better understanding of how the [+back] feature has affected coronal stops in the intervocalic context, the charts for each [-back]/[+back] pair of Persian vowels are displayed in Figure 6 separately.

As it is obvious, in all the intervocalic contexts, the number of the contacted electrodes in [t] in the first row was less than the number of the contacted electrodes in [d] in the same row. When [t] is flanked by [-back]/[+back] high vowels, the [+back] vowel drew it backward to a more posterior place. The same pattern can be seen for the mid [-back]/[+back] vowels as well. However, for the low [-back]/[+back] vowels there was a subtle or no difference between the places of articulation of [t]. In these contexts, the absolute difference between R3-R5 and R7-R8 was zero. Contrary to [t], when [d] is flanked by [-back]/[+back] high vowels, not much noticeable difference can be recognized. In mid and low vowel contexts, [+back] vowels drew the place of articulation of [d] to a more posterior position than that of [t].

Recasens (1999) reported that voiced dental or alveolar stops are more affected by coarticulation than their voiceless counterparts. Since, in Persian, the distinction between [t] and [d] in intervocalic contexts is not realized by voicing, but by aspiration (Nourbakhsh, 2009), the difference between the degree of resistance in [t] and [d] cannot be explained by voicing. This variation might be due to the fact that [t] is a laminal sound, while [d] is an apical one. Bladon and Nolan (1977) studied the alveolar consonants in RP English. They found that laminal speech sounds were more resistant to coarticulation than the apical ones. They argue the reason is that it is the tip rather than the blade that has the more distance from the dorsum when it is active in vocalic gestures. Skarnitzl (2013) reported that [t] being a voiceless laminal sound had stronger coarticulation resistance than [d] as an apical voiced sound.

4. Conclusion

The objective of this study was to determine if intervocalic contexts have an effect on the coarticulation of coronal stops in Persian. We first used electropalatography (EPG) to delineate the precise places of articulation of [t] and [d]. We then analyzed our data by computing the coarticulation index proposed by Farnetani (1989). The overall CI value in all intervocalic contexts was 1.30 for [t] and 2.24 for [d]. Finally, an independent sample t-test was conducted to compare the CI values in [t] and [d]. The results manifest that there is no significant difference between the degree of the coarticulatory effect in [t] and [d]. However, by looking at the data and the CI values of [t] and [d] in each vocalic context separately, we can say that [t] shows a slightly stronger coarticulation resistance than [d], Of course the result is tentative since we only recorded and analyzed the speech of one speaker. This might be explained by the fact that [t] is a laminal stop, while [d] is an apical one.

REFERENCES

Asiaee, M., Nourbakhsh, M. & Skarnitzl, R. (2018). Coronal stops in Persian: An electropalatographic study. In: M. Nourbakhsh, H. Asadi & M. Asiaee (Eds.), *Proceedings of First International Conference on Laboratory Phonetics and Phonology*, 37–44. Tehran: Neveeseh Parsi Publications.

Bijankhan, M. (2013). *Phonetic System of the Persian Language*. Tehran: Samt.

Bijankhan, M. (2018). Phonology. In: A. Sedighi and P. Shabani- Jadidi (Eds.), *The Oxford Handbook of Persian Linguistics*, 111–141. Oxford: Oxford University Press.

Bladon, R. A. W. & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, 4, 137–150.

Bladon, R. A. W. & Nolan, F. (1977). A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, 5, 185–193.

Butcher, A. & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, 15, 111–126.

Chen, W., Chang, Y., & Iskarous, K. (2015). Vowel coarticulation: landmark statistics measure vowel aggression. *Journal of the Acoustical Society of America*, 138, 1221–1232.

Farnetani, E. (1989). V-C-V Lingual Coarticulation and its Spatiotemporal Domain. In: W. J., Hardcastle & A. Marchal (Eds.), *Proceedings of the NATO Advanced Study Institute on Speech Production and Speech Modelling*, 93–130. Dordrecht: Kluwer Academic Publishers in cooperation with NATO Scientific Affairs Division.

Farnetani, E., Vagges, K. & Magno-Caldognetto, E. (1985). Coarticulation in Italian /VtV/ Sequences: A Palatographic Study. *Phonetica*, 42(2–3), 78–99.

Grützner, P. (1879). Physiology of voice and speech. In *Hermman's Handbook of Physiology (1 & 2).* (p.165204). Leipzig.

Kühnert, B. & Nolan, F. (1999). The origin of coarticulation. In W. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 7–30). Cambridge: Cambridge University Press.

Mahootian, Sh. & Gebhart, L. (1997). *Persian (Descriptive Grammars)*. London: Taylor & Francis Routledge.

Modarresi Ghavami, G. (2013). *Phonetics: Scientific Study of Speech*. Tehran: Samt.

Modarresi Ghavami, G. (2018). Phonetics. In: A. Sedighi and P. Shabani- Jadidi (Eds.), *The Oxford Handbook of Persian Linguistics*, 91–110. Oxford: Oxford University Press.

Nourbakhsh, M. (2009). Distinctive role of voice onset time (VOT) in standard Persian oral stops. PhD dissertation. University of Tehran.

Recasens, D. (1983). Timing and coarticulation for alveolo-palatals and sequences of alveolar [j] in Catalan. *Haskins Laboratories Status Rep. Speech Res*, 74/75, 97–112.

Recasens, D. (1984a). V-to-C coarticulation in Catalan VCV sequences: an articulatory and acoustical study. *Journal of Phonetics*, 12, 61–73.

Recasens, D. (1984b). Vowel-to-Vowel coarticulation in Catalan VCV sequences. *Journal of acoustical Society of America*, 76, 1624–1635.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, 15, 299–312.

Recasens, D. (1989). Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8(4), 293–307.

Recasens, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics*, 19(2), 177–192.

Recasens, D. (1999). Lingual coarticulation. In W. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 80–104). Cambridge: Cambridge University Press.

Recasens, D. & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of Acoustical Society of America*, 125, 2288–2298.

Skarnitzl, R. (2013). Asymmetry in the Czech alveolar stops: An EPG study. *AUC Philologica 1/2014, Phonetica Pragensia XIII*, 101–112.

Zharkova, N. (2008). An EPG and Ultrasound Study of Lingual Coarticulation in Vowel-Consonant Sequences. In: R. Sock, S. Fuchs & Y. Laprie (Eds.). *Proceedings of the 8th International Seminar on Speech Production*, 241–244. Strasbourg: INRIA.

RESUMÉ

Studie analyzuje artikulaci koronálních exploziv [t] a [d] v perštině pomocí elektropalatografie (EPG). Zaměřuje se především na koartikulační vliv sousedních vokálů na jejich realizaci. Vzorce EPG naznačují, že [d] je realizováno s anteriornějším postavením jazyka jako dentialveolární hláska, zatímco [t] je alveolární. Pro každý konsonant je stanoven koartikulační index (CI). Výsledky nenaznačují výraznější rozdíl mezi oběma hláskami z hlediska koartikulace, ale [t] vykazuje vyšší míru koartikulační rezistence vůči okolnímu vokalickému kontextu než [d]. To je v souladu s výzkumy, které poukazují na silnější koartikulační rezistenci u laminálních hlásek než u hlásek apikálních.

Maral Asiaee and Mandana Nourbakhsh
Department of Linguistics
Alzahra University, Tehran, Iran
E-mail: m.asiaee@alzahra.ac.ir

Saeed Rahandaz
Department of Linguistics
Bu-Ali Sina University, Hamedan, Iran

ACTA UNIVERSITATIS CAROLINAE
PHILOLOGICA 2/2019

Editor: doc. Mgr. Radek Skarnitzl, Ph.D.
Cover and layout by Kateřina Řezáčová
Published by Charles University
Karolinum Press, Ovocný trh 560/5, 116 36 Praha 1
www.karolinum.cz
Prague 2019
Typeset by Karolinum Press
Printed by Karolinum Press
ISSN 0567-8269 (Print)
ISSN 2464-6830 (Online)
MK ČR E 19831

Distributed by Faculty of Arts, Charles University,
2 Jan Palach Sq., 116 36 Prague 1, Czech Republic
(books@ff.cuni.cz)