AERODYNAMIC VOICING CONSTRAINT IN STOPS: FROM WOLFGANG VON KEMPELEN TO JOHN OHALA

Jürgen Trouvain

Saarland University, Language Science and Technology, Saarbrücken, Germany
trouvain@lst.uni-saarland.de

Abstract: Voicing in stop consonants is restricted by aerodynamical constraints, as illustrated by John Ohala, among others, more than three decades ago. It seems to be unnoticed in modern phonetic research that Wolfgang von Kempelen already described these challenges two centuries earlier. With additional bellows in his speaking machine he offered a mechanical solution for the vocal tract enlargement in stop production to achieve a longer voiced phase. A look back can help to develop a better understanding of the main mechanism for contrasting stops in many languages.

1 Introduction

The aim of this paper is to enlighten the non-trivial mechanism of voicing in stop consonants. This mechanism is strongly constrained by aerodynamics, which has been described in modern phonetics, mainly by Ohala [20, 21] in the last quarter of the 20th century, but also in historical sources, in particular by von Kempelen [14] at the end of the 18th century.

There is no doubt that there is a huge body of research on stops in the phonetic sciences in the last century or so. However, it does not seem to be the case that the voicing of stops is fully understood. For instance, investigations on the well-known phonological phenomenon of final devoicing in German ("Auslautverhärtung") reveal particular patterns of partial devoicing, e.g. [15]. In addition, the degree of voicing and devoicing plays a central role for the acquisition of the voiced-voiceless contrasts in stops and fricatives in second language learning [6].

A look at aerodynamical constraints on stop voicing seems also promising when attempting to model voiced stops. In addition to transcription-based auditory phonetics and acoustics-based instrumental phonetics, the modelling of concrete sound events provides extra information on the details of speech production. This phonetic modelling can for instance be performed through software-based articulatory synthesis [3] or through hardware-based experiments, e.g. [4]. The historical descriptions by von Kempelen have the advantage that replicas of his speaking machine offer mechanical models for investigating the aerodynamic voicing constraints on stops.

2 Voiced-voiceless contrasts in stops

Most languages of the world have stop consonants in their sound inventories. Most sets of stops in those languages show a two-way distinction for different places of articulation, though there are three-way distinctions as in Korean or Armenian or even four-way distinctions as in Hindi. Among the two-way distinctions the predominant set is to have the contrasts /p-b/, /t-d/ and /k-ɡ/. The term voiced vs. voiceless is mainly used to characterise this phonological contrast. It must be noted that the distinction in voicing is often used as an umbrella term for linguistic categorisation but that on the phonetic side voicing is not, or only partially, used to maintain the contrast in some languages.

There are alternative ways to distinguish between the phonologically voiced and unvoiced stops. This includes differences in voice onset time (*VOT*) and *aspiration*, e.g. in Eng-

10.21437/HSCR.2021-7

lish and many other languages, see e.g. Adjarian [1, 8] for very early investigations and some overviews such as Lisker & Abramson [17] and Cho & Ladefoged [10]. For German this contrast is often called lenis-fortis instead of voiced-voiceless [7, 12]. In addition, there is a tendency among many languages, not only German, to *final devoicing*, as Keating et al. [13] show.

There are also cases where stops contrast only in *quantity* and where aspiration and voicing play no role. For instance, in many Swiss German varieties there are differences in the duration of unvoiced closure phases between phonologically voiced and unvoiced plosives [16]. Also Ohala [21] observed a general tendency across languages for longer duration in voiceless obstruents compared to voiced ones.

Another example of not using voicing of the stops is the effect of *pre-consonantal vowel lengthening* (also called pre-consonantal voicing effect). It refers to vowels preceding phonologically voiced obstruents being consistently longer than their phonologically voiceless counterparts, such as the differences in the English minimal pairs 'bad-bat' and 'phase-face' [11, 24].

Voiced obstruents seem to be less frequent than their voiceless counterparts in phoneme inventories of the world's languages [18, 27]. However, voiced types of obstruents are nonetheless relatively common with more than 60% in the data of the World Atlas of Language Structures [18, 19]. There is a remarkable number of languages that lack a /g/ in the consonantal inventory. Maddieson [18] concludes that "[b]ecause of the aerodynamic facts, /g/ can be seen to be a less favored plosive than /b/ or /d/."

3 Aerodynamic constraint on voicing in stops in human speech

3.1 Air pressure and phonation

Typically voicing ceases after 15 ms [20]. The closure phases of fully voiced plosives are usually considerably longer than that and it is assumed that the vocal tract is enlarged in order to maintain a trans-glottal flow and thus delay the cessation of voicing. The data in Ohala and Riordan [20] showed a duration of about 82 ms as the limit for voicing in labials and, as expected, much shorter (52 ms) in velars, if purely *passive* expansion of the vocal tract walls is considered. The authors estimate that each cm^3 gives an estimated additional 10 ms of voicing.

One of the prerequisites for phonation is a sufficiently large difference of the transglottal air pressure to maintain airflow. After the oral closure for a stop, supra-glottal air pressure increases and leads to a reduction of the transglottal air pressure drop. When the difference is not large enough anymore, devoicing applies.

Catford [9] estimates the threshold value of the pressure differential at about 2-3 cm H_2O with adducted vocal folds in a modal phonation setting. He further estimates the volumes contained between the glottis and the oral closure for different places of articulation as 30-50 cm^3 for velar stops, 70-100 cm^3 for alveolar stops, and 120-160 cm^3 for bilabial stops. For a [g] there is less volume than for [d] and [b] and it can be expected that the devoicing happens earlier, as the place of closure is closer to the glottis. This is the explanation behind the frequent absence of /g/ mentioned above.

3.2 Vocal tract enlargement

This *passive* enlargement of the vocal tract happens through the compliance of tissue [20, 21, 22, 23, 26]. An *active* expansion of the vocal tract can be achieved for example by lowering the larynx or by lowering the tongue body or by elevating the already closed soft palate a bit more, and expanding the pharyngeal walls [20, 21, 22, 23, 26].

4 Aerodynamic constraints in stops in Kempelen's speaking machine

4.1 Voicing in stops in human speech

The aerodynamic constraints in voiced stops were already observed and described by Wolfgang von Kempelen [14] two centuries earlier. In his famous book, chapter 4 is devoted to the speech production mechanisms of "the sounds or letters of the European languages". For the description of [b] he mentioned that the voiced phase has a limited duration [14, p. 242]:

"Man kann auch bey geschlossenem Munde, und Nase eine Stimme hören lassen, jedoch nur eine kurze Zeit, und nicht sehr laut."

"When the mouth and nose are closed, one can hear the voice, but only for a short time and not very loudly."

What follows is the explanation of the difference of the transglottal air pressure and the drop of this pressure difference needed for phonation after a short while [14, p. 243]:

"Wenn nun die Stimme ansprechen soll, so wird die in der Lunge enthaltene Luft zusammengedrückt, das Stimmhäutchen öffnet sich ein klein wenig, und gestattet ihr einen nur ganz engen Durchgang. In der in dem Munde vorräthigen, viel weniger zusammengedrückten Luft findet sie noch so viel Raum, daß sie in dieselbe mit einem Laut hineinströmen kann, indem die diese immer mehr und mehr zusammendrückt. So bald die Luft in dem Munde eben so sehr zusammengedrückt ist, als die in der Lunge, so ist zwischen beyden das Gleichgewicht hergestellt, der Strohm der Luft hört auf, und mit ihm die Stimme. Und dieses ist die Ursache, warum die Stimme nur eine kleine Weile, etwan 1 Sekunde lang anhalten kann."

"If now the voice wishes to speak, the air contained in the lungs is compressed, the vocal cords open a small amount, and allow it only a narrow passage. In the much less compressed air that is found in the mouth, it finds so much space that it can flow in, producing a sound, as it compresses the latter more and more. As soon as the air in the mouth is as compressed as that in the lungs, an equilibrium is established between the two, the stream of air ceases, and with it the voice. And this is the reason why the voice can only last for a short while, about one second."

He indicates an entire second for this aerodynamic manoeuvre of voicing in the (bilabial) stop, but such a time interval should be regarded as a rather short duration of auditory observation rather than a precise measurement that would be possible much later. Interestingly, he also offers the vocal tract enlargement in its passive and in its active form as an explanation of the underlying speech production mechanism [14, p. 244]:

"Nebst der Zusammendruckbarkeit der Luft bedienet sich die Natur noch eines anderen Kunstgriffes um der aus der Lunge kommenden Luft Raum zu verschaffen. Die Wände des ober dem Stimmhäutchen befindlichen Behältnißes, nämlich die fleischigen und daher nachgebenden Theile des inneren Halses erweitern sich, oder werden vielmehr von der anschwellenden Luft aufgebläht. […] Wenn diese Theile einmal auf das höchste aufgetrieben sind, so muß die Stimme aufhören. Bläßt man über dieses noch die beyden Backen auf, so kann man die Stimme noch eine Weile länger aushalten."

"Besides the compressibility of the air, nature provides another artifice to make space for the air coming from the lungs. The walls of the chamber above the vocal cords, namely the fleshy and thus compliant portion of the inner throat expand, or rather are inflated by the air that is forced in. […] When these parts are expands as far as they will go, the voice must cease. If one in addition expands the two cheeks, one can maintain the voice a bit longer."

4.2 Vocal tract enlargement in the speaking machine

Von Kempelen was also interested in modelling the observed effect of vocal tract enlargement. In his speaking machine [14] the vocal tract enlargement can be obtained by means of the 'plosive bellows'. These bellows are located directly beneath the 'nasal cavity', and the two cavities are linked with a small tube (Figures 1 and 2).

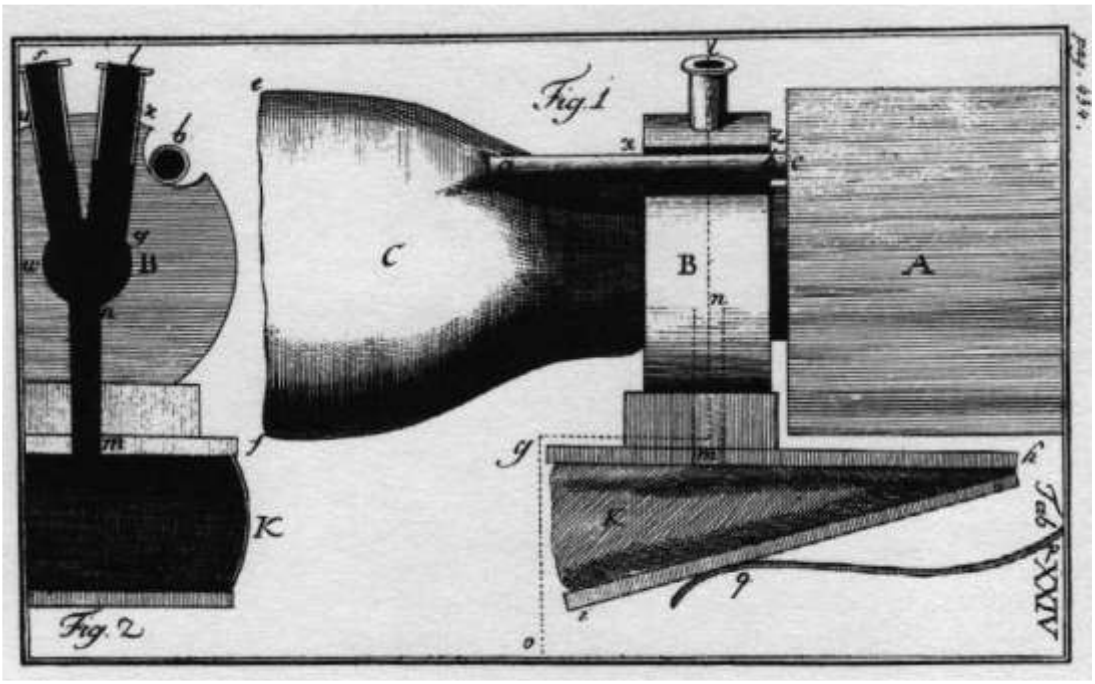

Figure 1. Plosive bellows "K" seen in cross-section in the front view (left-hand side) and in the side view of the inner life of the speaking machine (taken from the original engraving in Kempelen's book from 1791 [14]). The 'nasal cavity' ("B") is linked with a small tube ("n") to the plosive bellows located beneath.

These second, much smaller bellows also have the effect of lengthening the voiced phase of the plosives. There is evidence of this effect for various replicas. However, replicas without these bellows are better at generating voiceless plosives with an aspiration phase before voicing begins. An aspiration effect is not possible in replicas with these bellows, where a 'Papa' sounds more like a 'Baba' (at least to speakers of West Germanic languages). The possibility of switching the second bellows on and off would clearly offer a good solution to control the 'aerodynamic voicing constraint' [21].

It must be noted, however, that it was von Kempelen's belief that he was invoking exactly the opposite effect with the installation of the additional bellows [14, p. 437f.]:

"Um die Explosion bey den stummen Mitlautern zu verstärken, habe ich noch einen anderen, nicht minder wichtigen Zusatz gemacht. Ich habe nämlich an das Stück *B* unten einen kleinen Blasebalg *g h i* angebracht […] Wenn nun Mund und Nase geschlossen ist, und der große Blasebalg gedrückt wird, so bläht die gepreßte Luft diesen kleinen Blasebalg mit auf. Wird sodann die Hand, die den Mund verschlossen hält, jäh von da weggerückt, so stoßt der kleine Blasebalg, der durch die an der Windlade befestigte Drahtfeder *p q* zusammen gedrückt wird, die Luft rasch von sich, wodurch die Explosion zum Munde hinaus nothwendig stärker wird. Und so haben wir nun den Buchstaben P in seiner Vollkommenheit."

"In order to strengthen the explosion with the voiceless consonants, I made yet another no less important addition. I installed underneath piece *B* a small bellows *g h i* […] If the mouth and nose are closed, and the big bellows is pressed, the compressed air inflates this small bellows. If then the hand, which holds the mouth closed, is taken away, the small bellows, which is pressed together by the compression spring *p q* that is fastened to the wind chest, presses the air quickly out, by which means the explosion at the mouth is necessarily stronger. And thus we have the letter P in its full form."

Although the reasoning for generating an optimal [p] – instead of [b] – with the additional mechanical detail was not correct, it is remarkable that von Kempelen's observations and explanations on the aerodynamic voicing constraint on stops and the vocal tract enlargement were correct – a long time before modern and instrumental phonetics. In addition, his engineering approach of integrating small bellows for the stop production is admittedly sensational.

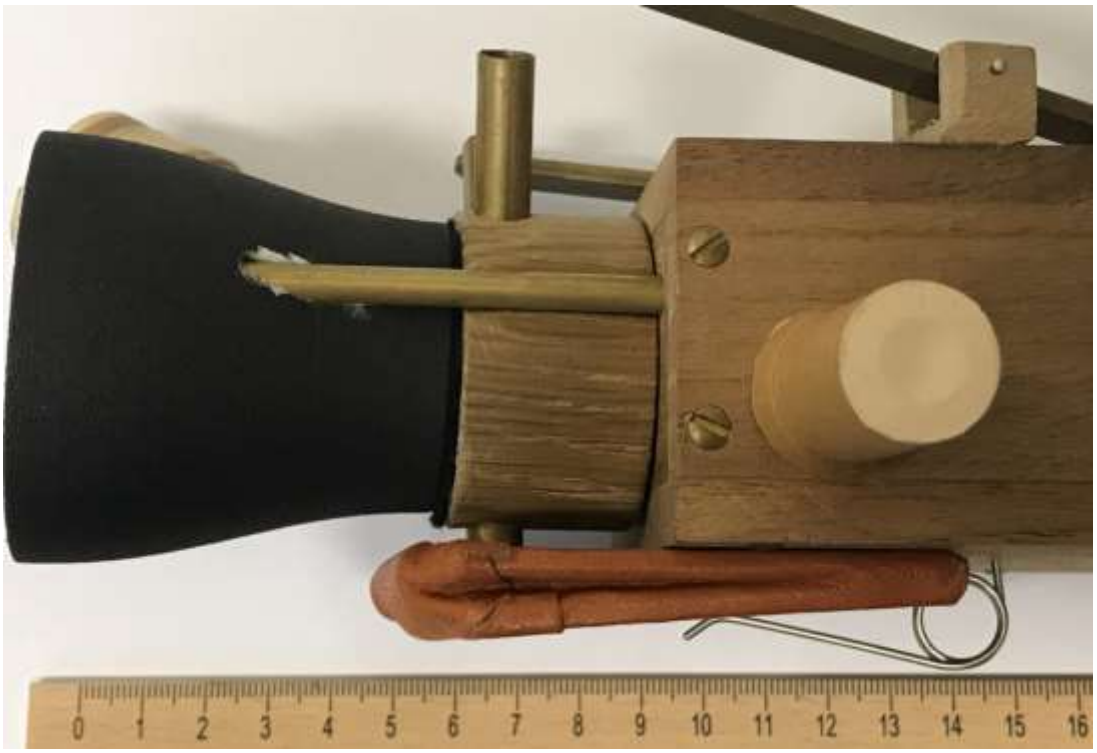

Figure 2. Side view of the inner life of the speaking machine of the Saarbrücken replica (measures in centimetres). Photography corresponds to the right-hand side illustration of Figure 1. Plosive bellows at the bottom are not inflated.

In order to obtain a distinction between [b] and [p] with the 'speaking machine', the player has to control the air pressure (by her/his elbow). The air pressure is increased for [p] compared to [b] immediately before and during the closure of both apertures. However, the initiation of 'subglottal' pressure must not be too strong, otherwise the onset of the 'vocal fold' vibration fails.

5 Considerations for future studies

The aerodynamic voicing constraint in stop consonants has been described by John Ohala, among others, more than three decades ago. It seems to be unnoticed in modern phonetic research that Wolfgang von Kempelen already described these challenges two centuries earlier. But still today it is unclear how exactly the enlargement of the vocal tract is executed for phonetically voiced stops.

A better understanding of the mechanism of voicing in stops can also help to explain further phenomena linked to this mechanism. One example refers to differences in how languages and language varieties use voicing in stops, maybe along with variation in VOT, closure duration and pre-consonantal vowel lengthening. Another example is variation in microprosodic details in, before, and after stops, that helps to improve articulatory synthesis [5]. A further example would apply to human non-verbal vocalisations such as crying but also screaming, with, so far, hardly any knowledge about vocal tract configurations, adjustments of vocal fold tension and sub-glottal pressure.

With the integration of the 'plosive bellows' in his speaking machine, von Kempelen seems to have been far ahead of his time. It will remain the task of today's research to continue to develop models, that are able to temporally overcome the aerodynamic voicing constraint, an often-ignored phenomenon that was frequently pointed out by Ohala. Those models are important for demonstration and research, as was shown by Arai with "modern" material, e.g. in [2], or with replicas, e.g, [25].

Acknowledgements

The author would like to thank the editors, two anonymous reviewers and colleagues from Saarbrücken for their suggestions that substantially helped to improve the quality of this paper.

References

[1] ADJARIAN, H. Les explosives de l'ancien arménien étudiées dans les dialectes modernes. *La Parole ou Revue internationale de Rhinologie, Otologie, Laryngologie et Phonétique expérimentale* (Paris), pp. 119–127, 1899.

[2] ARAI, T.: Education in acoustics and speech science using vocal-tract models. *The Journal of the Acoustical Society of America* 131(3), pp. 2444–2454, 2012.

[3] BIRKHOLZ, P.: *VocalTractLab* [software] Available online at: http://www.vocaltractlab.de, 2017.

[4] BIRKHOLZ, P., STONE, S & KÜRBIS, S.: Comparison of different methods for the voiced excitation of physical vocal tract models. *Proc. Elektronische Sprachsignalverarbeitung (ESSV)*, Dresden, pp. 84–94, 2019.

[5] BIRKHOLZ, P. & ZHANG, X.: Accounting for microprosody in modelling intonation. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8099–8103, 2020.

[6] BONNEAU, A.: Realization of French voiced fricatives by German learners as a function of speaker level and prosodic boundaries. *Proc. 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, 5 pp., 2015.

[7] BRAUN, A.: *Zum Merkmal "Fortis/Lenis". Phonologische Betrachtungen und instrumentalphonetische Untersuchungen an einem mittelhessischen Dialekt*. Stuttgart: Steiner, 1998.

[8] BRAUN, A.: An early case of "VOT". *Proc. Interspeech* 2013, Lyon, pp. 119–122, 2013.

[9] CATFORD, I.: *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press, 1977.

[10] CHO, T. & LADEFOGED, P.: Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27, pp. 207–229, 1999.

[11] HOUSE, A. S. & FAIRBANKS, G.: The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, pp. 105–113, 1953.

[12] JESSEN, M.: *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam: Benjamins, 1998.

[13] KEATING, P., LINKER, W. & HUFFMANN, M.: Patterns of allophone distribution for voiced

and voiceless stops. *Journal of Phonetics* 11, pp. 277–290, 1983.
[14] VON KEMPELEN, W. [1791]: *Mechanismus der menschlichen Sprache / The Mechanism of Human Speech*. Kommentierte Transliteration & Übertragung ins Englische / Commented Transliteration & Translation into English by FABIAN BRACKHANE, RICHARD SPROAT & JÜRGEN TROUVAIN (eds) Part 1 (Studientexte zur Sprachkommunikation Band 87) and Part 2 (Studientexte zur Sprachkommunikation Band 88). TUDpress: Dresden, 2017.
[15] KLEBER, F., JOHN, T. & HARRINGTON, J.: The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38, pp. 185–196, 2019.
[16] KRAEHENMANN, A. & LAHIRI, A.: Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *Journal of the Acoustical Society of America* 123, pp. 4446–4455, 2008.
[17] LISKER, L. & ABRAMSON, A.S.: A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, pp. 384–422, 1964.
[18] MADDIESON, I.: Voicing in plosives and fricatives. In: M.S. DRYER & M. HASPELMATH (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. (Available online at http://wals.info/chapter/4)
[19] MADDIESON, I.: Voicing and gaps in plosive systems. In: M.S. DRYER & M. HASPELMATH (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. (Available online at http://wals.info/chapter/5)
[20] OHALA, J.J. & RIORDAN, C.J: Passive vocal tract enlargement during voiced stops. In: J.J. WOLF & D. H. KLATT (eds.), *Speech communication papers*. New York: Acoustical Society of America, pp. 89–92, 1979.
[21] OHALA, J.J.: The origin of sound patterns in vocal tract constraints. In: P.F. MACNEILAGE (ed.), *The Production of Speech*. New York: Springer, pp. 189–216, 1983.
[22] OHALA, J.J.: Accommodation to the aerodynamic voicing constraint and its phonological relevance. *Proc.* 17th *International Congress of Phonetic Sciences*, Hong Kong, pp. 64–67, 2011.
[23] SVIRSKY, M. A., STEVENS, K. N., MATTHIES, M. L., MANZELLA, J., PERKELL, J. S., & WILHELMS-TRICARICO, R.: Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America* 102, pp. 562–571, 1997.
[24] TANNER, J., SONDEREGGER, M., STUART-SMITH, J. & FRUEHWALD, J.: Toward "English" phonetics: Variability in the pre-consonantal voicing effect across English dialects and speakers. *Frontiers in Artificial Intelligence* 3, Article 38, 15 pp., 2020.
[25] TROUVAIN, J. & BRACKHANE, F.: Wolfgang von Kempelen's speaking machine as an instrument for demonstration and research. *Proc.* 17th *International Congress of Phonetic Sciences*, Hong Kong, pp. 164–167, 2011.
[26] WESTBURY, J. H.: Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73, pp. 1322–1336, 1983.
[27] ŻYGIS, M., FUCHS, S. & KOENIG, L.L.: Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology* 3(2), pp. 299–336, 2012.