

Response to revision

21/01/2019

1

Overall Evaluation

The paper presents an acoustic study of the “voicing effect” in Italian and Polish. The author seeks evidence to bear on the question of where the “voicing effect” comes from. The results, particularly those of release-to-release interval in Figure 5, appear to indicate that the so-called effect is in fact a case of compensation. Overall, this is an excellent paper and can be published with only MINOR REVISIONS. The methodology is solid and transparently described, save for one issue to be mentioned below. I have a few comments concerning the implications of the paper with respect to phonological aspects of laryngeal contrasts and prosodic structure. The author is invited to consider these issues, which may help in strengthening the theoretical background and discussion sections.

General comments

Page 6, line 171 regarding Polish as a voicing language: Polish VOTs for /ptk/ are longish in many cases, and for some speakers might be described as semi-aspirated. I’ve witnessed this many times myself. For one relatively recent published description, see e.g. Waniek-Klimczak (2011). I think the values in Keating’s PhD are also a bit longer than true short-lag.

Lines X mention the concern about the contested status of Polish as a true voicing language. Line X has been reworded so that now reads ‘On the other hand, they [Italian and Polish] are both traditionally classified as ‘true voicing’ languages, in which the phonologically voiced stops are typically articulated with vocal fold vibration’. The main point here is that while actual voicing during closure of lenis stops in aspirating languages is optional, voicing during the closure of phonologically voiced stops in both Italian and Polish is robust. Further theoretical reasons are also given on how the phonological distinction between voicing and aspirating languages is not clear-cut in face of empirical data, although a thorough discussion would be beyond the scope of the paper.

Page 7, lines 213-215. Since the recording locations were Italy and Manchester, I suppose the Polish participants were recorded in the UK? If so, how long have they been in the UK? The >6 month column is quite vague. From the table in the appendix, it appears as though all but two of the Polish speakers have lived abroad (i.e. have been out of the country for more than 6 months), and all of them have some knowledge of English. Since English has probably the most robust ‘voicing effect’ around, and it seems that the Polish participants are quite proficient in English, with regard to the voicing effect I think it is difficult to claim that the sample of Polish participants for this study is representative. There needs to be more information about the participants. How often do they use English/Polish? What is their education level? Do they use English in their professional lives? What about the language in which the recording sessions were conducted? Unless they were conducted in Polish (i.e. with a native Polish speaker giving instructions in Polish to the participants), I think it’s almost certainly the case that these speakers’ L1 Polish has been affected by English. I don’t think this disqualifies the paper. It’s just that the possibility of L2 influence on L1 needs to be acknowledged in the text, as well as in the discussion. For information on the effects of language mode with bilingual participants, see Grosjean (2004). If the recording sessions were indeed carried out entirely in Italian and Polish, Antoniou et al. (2010) is a study that may be used to argue that the data in the paper may indeed be representative of the L1s. For other work on L2-induced effects on L1 pronunciation, see Chang (2012, in press), Schwartz et al. (2015a, 2015b), Herd et al. (2015).

As mentioned in lines X [Procedure], the participants were spoken in their native language prior to the experiment, and instructions/equipment fitting was conducted in the native language as well. Lines X mention the concern of L2 influence in light of the results in Antoniou et al. (2010) and Schwartz 2015b, suggested by the reviewer.

Phonological comments

Page 10, line 289. Isn't it an interesting question why Italian has lenition but Polish doesn't? See comments below.

Page 10, lines 291-298. This is interesting, and completely goes against the predictions of Laryngeal Realism (stated explicitly I believe by Beckman et al. 2013), that intervocalic voicing should be absent from voice languages, since it would mean "adding" a feature [voice] in a weak prosodic position. Perhaps there is no feature [voice]? See Schwartz & Arndt (2018) for an account of the phonology of voicing languages in which there is no feature [voice], and some perceptual evidence in favor of that account.

Pages 20-21. It seems to me that the results of these studies suggest that Italian and Polish behave somewhat differently with regard to the voicing effect. The mean differences are greater in Italian than they are in Polish (16 ms vs. 7.5 ms as stated in 4.1). Do you think it is indeed the case that Polish and Italian differ systematically? If the difference between the languages is in fact systemic, I think phonology needs to have something to say about it. For discussion of this issue, particularly with regard to Polish, see Schwartz (2016). Essentially the claim in that paper is that Polish does not really have trochees (see Newlin-Lukowicz 2012 for phonetic evidence that is compatible with this claim). For Schwartz, true trochees show a recursive structure, which puts the second C in a CVCV in an inherently weaker position. Perhaps language-specific differences between Polish and Italian are due to prosodic differences in the structure of CVCV sequences? This might explain the earlier comment about why Italian shows lenition but Polish doesn't?

While there is an apparent systematic difference in the magnitude of the voicing effect between Italian and Polish, the behaviour of the individual speakers is quite idiosyncratic and does not pattern with their native language (as revealed by the by-speaker random coefficients extracted from the mixed-effects models). Line X now offer an overview of the individual differences found in relation to the voicing effect and how these individual variation does not seem to support clear across-language patterns.

References

- Antoniou, M., Best, C., Tyler, M. & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38, 640–653.
- Beckman, J., Essen, M. & Ringen, C. (2013). Evidence for laryngeal features: Aspirating vs. true-voice languages. *Journal of Linguistics*, 49(2), 259–284.
- Chang, Charles (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40. 249-268.
- Chang, Charles (in press). Phonetic drift. In M. S. Schmid, B. Köpke, M. Cherciov, E. de Leeuw, T. Karayayla, M. Keijzer & T. Mehotcheva (Eds.), *The Oxford Handbook of Language Attrition*, Chapter 17. Oxford, UK: Oxford University Press.
- Grosjean, François (2004). Studying bilinguals: Methodological and conceptual issues. In Tej Bhatia & William Ritchie (eds.), *The Handbook of Bilingualism* (pp. 32-63). Oxford: Blackwell Publishing.
- Herd, W., Walden, R., Knight, W. & Alexander, S. (2015). Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics* 23. Acoustical Society of America.
- Newlin-Lukowicz, L. 2012. Polish stress – looking for evidence of a bidirectional system. *Phonology* 29 (2), 271-329.
- Schwartz G, Balas A, Rojczyk A. 2015a. Language mode vs. L2 interference: evidence from L1 Polish. In: *The Scottish Consortium for ICPHS 2015 (ed.) Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow, online. URL: <http://www.icphs2015.info/pdfs/Papers/ICPHS0550.pdf>
- Schwartz G, Balas A, Rojczyk A. 2015b. Phonological Factors Affecting L1 Phonetic Realization of Proficient Polish Users of English. *Research in Language* 13(2): 181-198.
- Schwartz, G. 2016. On the evolution of prosodic boundaries – parameter settings for Polish and English. *Lingua* 171, 37-74.

Schwartz, G. & Arndt, D. 2018. Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences* 69, 98-112.

Waniek-Klimczak, E. (2011). Aspiration in Polish: A sound change in progress?. In Pawlak, M. & J. Bielak (eds), *New Perspectives in Language, Discourse and Translation Studies*. Berlin: Springer.

2

The paper presents acoustic data on vowel duration before voiced and voiceless stops in two languages, Italian and Polish. The paper confirms the long-described voicing effect in these languages and claims that release-to-release interval remains constant regardless of the voicing effect. This latter result is argued to support the compensatory lengthening explanation of durational differences of vowels in positions before voiced and voiceless stops.

The paper is clearly written and well structured. The statistical analysis is presented transparently and I have no major reservations (for minor comments, see below). The acoustic analysis and experimental design are standard. A few shortcomings in experimental design: the procedure includes a high number of repetitions of the same word (5-6 repetitions), which is not desired, but a general practice, and a relatively small inventory of nonce words.

Results are relevant for the discussion of the voicing effect for the following reasons. Polish has been reported not to feature the voicing effect. The paper provides additional data arguing that voicing effect exists in Polish as well. The data suggest that the release-to-release interval remains constant regardless of voicing of the following stop. According to the author, this supports the compensatory lengthening approach to the voicing effect.

The main objection against the paper is that it fails to interpret its result in relation to a vast body of literature that exist on the voicing effect. It is true that the constant rate of the release-to-release interval is predicted under the compensatory lengthening approach. However, it is likely that more than one factor influences vocalic durational differences before voiced and voiceless stops. It is possible that the durational differences result from a sum of different mechanisms (each of which contributes a weighted portion).

In sum, a well-written paper with results that yield some new information for a longstanding discussion. At the current stage, I cannot recommend publication, but with a better discussion on different factors that can influence durational differences, the paper is publishable in a journal like *Glossa*.

Major comment

As already mentioned above, the paper should clearly state that the constant value of the release-to-release interval can be due to different mechanisms operating on the vowel duration differences and does not per se point to the compensatory lengthening explanation. If the author has additional evidence in favor of why compensatory lengthening is likely the main contributing factor, this should be stated as well.

If I understand correctly, the reviewer is saying that the temporal stability of the release to release interval could be a consequence of the fact that vowels are lengthened when followed by voiced stops, so that the temporal distance between two consecutive releases in this context matches that in the context of a voiceless stop. However, there are reasons to argue that the temporal stability is a logical antecedent, as now discussed in lines X. It is anyway recognised that further work is needed to disambiguate between these two interpretations by looking at general properties of gestural phasing that might encompass temporal patterns other than the voicing effect.

Furthermore, the author should discuss the hypothesis that laryngeal features influence vowel durations in greater detail. The few works that measure vowel duration before stops other than voiced vs. voiceless (Durvasula and Luo 2014, Beguš 2017) find evidence for the effect of laryngeal features. Because the present paper measures vowel duration before voiced and voiceless stops which also have different closure durations, the laryngeal effects and the closure effects are of course conflated. This should be addressed in the paper.

This point has been discussed in more details in lines X. The results on laryngeal effects reported in Beguš 2017 are discussed and compared with those from this study. Lines X in the conclusion section stresses that a compensatory account is likely not the only mechanism underlying the observed durational patterns, and that multiple causes can conspire to produce the difference in vowel/closure durations.

Minor comments

Abstract

“factors (like perceptual biases) could”

To be fair to all proposals, I think at least the laryngeal effect hypothesis should be mentioned here.

The abstract now has “(like perceptual biases and laryngeal effects)”.

186-198 Does the author have any guesses why the results are so different for Polish? The review of the literature is sufficient, but it would be very beneficial to the reader if the author would evaluate methodology of the studies, especially since Polish has been mentioned as a case of non-existence of the voicing effect. Is experimental design equally balanced in all studies? If not, are there patterns of experimental design in studies that don't report the voicing effect?

Lines X discuss the possible reasons behind the diverse results. While no clear pattern can be seen between methodology and presence/absence of a significant voicing effect, the small reported estimates could indicate inability to detect small differences due to low statistical power.

Statistical analysis.

The analysis is generally sound and clear. Unfortunately, the manuscript does not mention whether assumptions of the linear model are checked. Sometimes modeling durations can produce patterns in the residual plot, so it would be good to inform the reader that the residual plots look without patterns.

The residual and QQ plots of the models do not show patterns in the residuals (see X). Some of the residual plots show different blobs, this is related to idiosyncratic differences in mean durations between participants and it is not an issue (it is taken care of by the random effects). Line X added in the manuscript: “Inspection of residual plots and QQ plots of the models described below indicated absence of patterns in the residuals.”

A short justification of why Bayes factors vs. the more standard AIC criterion is used.

This has been now added as footnote X in the manuscript.

Release to release

Why release to release? The author mentions in the Discussion section that release-to-release measurement is somewhat arbitrary. A broader discussion on why particularly this interval would be useful. Has the author tested any other measurements? Do the result change if, say, VOT is excluded from the interval?

3

Short justification of recommendation (maximum of 10 sentences): I recommend that the paper be accepted for publication, pending only minor revisions. The paper presents an interesting empirical study which sheds light on the so-called voicing effect and thus tackles more general questions of gestural coordination during speech production. The design of the study is sound, the analyses sophisticated relying on the latest trends in speech sciences, and the results and interpretations convincing.

Does the paper present an empirical discovery potentially of interest to most of this journal's readers? Please elaborate your answer.: The paper examines a phenomenon which has been addressed by a number of researchers since the 1960s but which was still not explained unequivocally. I believe the submitted paper makes a valid contribution to this discussion by first identifying an interval of speech which is not affected by the proposed temporal compensation mechanism and then analyzing its extent between vowel duration and stop closure duration. As such, the study will be interesting for the journal's readers. The results are

convincing, the charts with the results are very illustrative and well designed, and the fact that the author analyzes two languages, Polish and Italian, which have been reported to differ in the extent of the voicing effect, makes the results even more interesting.

Is the empirical content of the paper sound (e.g. fieldwork includes proper controls and comparisons, experiments well designed, etc.)? Please elaborate your answer.: The empirical content of the paper is sound. The study is based on controlled speech material, which was further constrained by the fact that ultrasound (UTI) data were recorded alongside the speech signal within a larger project. However, that is not detrimental to the reported study. The artificial disyllabic sequences allow the author to more reliably answer the research question and compare tendencies across the two languages. The amount of data (ca. 1,200 tokens analyzed in total) is adequate; it is only a pity that the speaker sample is not more balanced (11 speakers of Italian vs. 6 speakers of Polish). The acoustic recordings have been manually segmented according to established criteria, which is crucial when dealing with fine differences in duration of speech sounds. Statistical analyses rely on the latest trends in speech sciences.

Does the paper make a broader proposal about an aspect of linguistic theory potentially of interest to most of this journal's readers? Please elaborate your answer.: As mentioned in the reply to question 3, temporal relationships within vowel-consonant sequences depending on the voicing of the consonant have been studied for a long time in a number of languages, but reasons for this "voicing effect" are still not absolutely clear. While being rather a specific question, it (and the submitted study) is closely related to more general questions of the coordination of articulatory gestures during speech production. I therefore believe that the topic will be interesting for the journal's readers.

Is the argumentation linking the paper's broader conclusions to its empirical or theoretical premises sound? In answering this question, please elaborate your answer without regard to your personal judgments concerning the plausibility of these premises — see question 7.: The author's argumentation is very well presented and it makes clear links to the state-of-the-art in researching the question. While proposing a well-reasoned articulation-based account of compensation between the duration of a vowel and that of the following stop consonant, the author acknowledges the possibility of other (not incompatible) explanations, predominantly those based perceptually.

Comment on the paper's premises or the conceptual framework that it assumes, if you believe that issues in this area are relevant to the overall evaluation of the paper.: N/A

Any other comments relevant to the evaluation of the paper as a whole.: N/A

What are your suggestions for improving the paper? (optional if your publication recommendation is "accept" or "reject", strongly recommended otherwise). If your publication recommendation was "revisions required" or "resubmit for review", your recommendations may be taken by the editor as requirements for future acceptance, unless you explicitly state otherwise, so please try to distinguish your high-priority requirements for revision from weaker suggestions.: I have no recommendations concerning the scientific aspect of the submitted paper: this is a very well-written and well-presented study. The paper needs proofreading, and some minor language aspects are mentioned in the following point.

Specific line-by-line comments on details of the paper. Begin each comment with a page number, example number or paragraph/line from top or bottom, as appropriate.:

p. 5, line 150 – plausible on the light of → plausible in the light of

p. 7, line 223 – Hawlett-Packard → Hewlett-Packard

p. 8, continuation of footnote 6 – for sake of → for the sake of

p. 9, line 267 – which yields to a grand total → which yields a grant total

p. 9, footnote 7 – I am not sure this footnote is necessary or relevant for the paper

p. 13, line 370 – Vowel tend to be → Vowels tend to be

p. 20, line 473 – voiced then when → voiced than when

These typos have been corrected, plus a few other instances of ‘then’ for ‘than’. Footnote 7 has been removed.