J. Linguistics **60** (2024), 253–284.
doi:10.1017/S0022226722000457

Formalising phonological perception: The role of voicing assimilation in consonant cluster perception in Emilian dialects

EDOARDO CAVIRANI
KU Leuven
SILKE HAMANN
University of Amsterdam

(Received 25 November 2021; revised 26 October 2022)

Speech perception is influenced by language-specific phonological knowledge. While phonotactics has long been established to play a role, the study of how phonological alternations influence perception is still in its infancy. In this paper, we make a case for the latter by investigating the role of regressive voicing assimilation (RVA) in the perception of obstruent clusters in Emilian dialects of Italian. We provide empirical evidence from a phoneme-detection task, in which Emilian listeners reported to have heard [b] significantly more often in stimuli with a /p/ before a voiced obstruent (RVA context) than before a vowel (non-RVA context). Our experimental findings add to recent work on the influence of phonology on speech perception. In addition, we provide an explicit formalisation, which bolsters the need for a rigid distinction between phonetic, surface and underlying representation, and an explicit mapping between all three, both in the process of speech production and comprehension.

KEYWORDS: BiPhon, Emilian dialects, Phonetics-Phonology interface, Regressive voicing assimilation

1. INTRODUCTION

The influence of language-specific phonotactic restrictions on speech perception (Polivanov 1931; Swadesh 1934; among others) has been recently backed up by studies on so-called illusory vowels, where listeners perceive a vocalic segment even though there are no corresponding formants in the acoustic signal (Dehaene-Lambertz et al. 2000; Berent et al. 2007; Kabak & Idsardi 2007; Boersma & Hamann 2009; Monahan et al. 2009; Dupoux et al. 2011; Kilpatrick et al. 2021; Whang 2021). A well-known example comes from the study by Dupoux et al. (1999), where native Japanese listeners presented with French realisations of nonce words, such as /ebzo/, with an obstruent cluster that is phonotactically illicit in Japanese speech, reported to have heard a vowel breaking up the illicit cluster (e.g. for /ebzo/, they reported to have heard [ebuzo] in approximately 70% of the cases).

While most of the work on illusory vowels focuses on the interplay between acoustic properties and phonotactic restrictions (e.g. McClelland & Elman 1986; Daland et al. 2019), Durvasula & Kahng (2015, 2016) provide evidence from Korean speech that phonological alternations are also of relevance in speech perception, as they can account, for example, for the quality of the illusory vowel. Adding to the experimental field of possible phonological influences on speech perception, the present study investigates the influence of regressive voicing assimilation (RVA) on the perception of voicing in obstruent clusters.

The languages of interest are a set of Gallo-Italic varieties spoken in Emilia (northern Italy), more specifically, in Parma, Modena, Bologna and Ferrara. These varieties display unstressed vowel reduction, which applies both word-medially and word-finally to various degrees. The effect of unstressed vowel reduction ranges from reduction to complete deletion (Loporcaro 2011; Passino 2013). As shown by the Bolognese examples in (1) and (2),[1] complete deletion results in highly marked consonant clusters, which can trigger readjustment processes, such as prothesis in Example (1a), epenthesis in Example (1b) and deletion in Example (1c).

(1) (a) [a ˈlɛk] ‘I lick’ - [**a**lˈkɛːr] ‘to lick’
(b) [aˈliːgra] ‘happy.F’ - [aˈliːg**e**r] ‘happy.M’
(c) [landa] < *lampda < LAMPADA(M) ‘lamp’

RVA is one of the possible readjustment processes. Its effect is shown in Example (2), where pairs are given that exhibit assimilation of voice in Example (2a), assimilation of voicelessness in Example (2b) and the inactivity of sonorants in the assimilation process in Example (2c).

(2) (a) [(a) ˈ**p**ai̯**z**] ‘(I) weigh’ - [ˈ**bz**ɛːr] ‘to weigh’
[(a) ˈsai̯**g**] ‘(I) sew’ - [ˈ**zg**ɛːr] ‘to sew’
(b) [ˈ**bak**a] ‘mouth’ - [ˈ**pk**æŋ] ‘mouthful’
[ˈ**vɛt**ʃ] ‘old’ - [ˈ**ft**ʃats] ‘old geezer’
(c) [ˈ**p**ai̯**r**] ‘pear’ - [ˈ**pr**eŋ] ‘little pear’
[ˈ**p**ai̯**l**] ‘hair’ - [ˈ**pl**eŋ] ‘little hair’

The form pairs in Example (2) are morphologically related. This strongly supports the hypothesis that in the varieties under consideration, RVA is a synchronic process. Note that speakers are provided plenty of morphophonological evidence for the underlying voicing specification of the relevant segments. Besides the base-diminutive and PRS.1SG - INF (PRS = present, 1SG = 1 person singularind = indicative, PL = plural) pairs in Example (2), this is particularly clear in the case of verbal paradigms. For instance, in the IND.PRS paradigm of /p(ai̯)z-ˈɛːr/ ‘weigh-INF’, forms with the diphthong [ai̯] – [a ˈpai̯z] ‘I weigh’, [ət ˈpai̯z] ‘you.SG weigh’, [al ˈpai̯za] ‘s/he weighs’, [i ˈpai̯zeŋ] ‘they weigh’ – alternate with forms in which the stress is attracted by the inflectional suffixes /ˈɛŋ/ and /ˈɛ/ for 1PL and 2PL,

[1] The forms in Examples (1) and (2c) have been suggested by Daniele Vitali personal communication (p.c.). Those in Examples (2a) and (2b) have been produced by one of our participants (P2).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

respectively, and [aḭ] gets deleted, thereby triggering RVA – [a ˈbzɛŋ] 'we weigh', [a ˈbzɛ] 'you.PL weigh'. In all these cases, the speaker can easily recover the underlying voicing specification of the relevant consonant.

The presence of RVA in Emilian dialects has been reported by several scholars.[2] Rohlfs (1966: 341) claims that RVA 'can be frequently observed in Northern Italian dialects', and, in particular, in Romagnolo and Emilian varieties, where RVA applies 'as a consequence of the deletion of the intermediate vowel' in word-initial (*BOCC-ONE > Romagnolo [pkõ] 'mouthful'), word-medial (*bragh-ettina > Imolese [braktēna] 'underwear') and word-final (*tevedo > Imolese [teft] 'lukewarm') position, as well as across word-boundaries (Emilian [um brank at pegər] 'a herd of sheep', where the preposition [at] derives from /d/ by means of RVA and [a] prosthesis). Similarly, Vitali & Pioggia (2014: 22) claim that syncope feeds RVA in all Emilia-Romagna dialects, whereas Gaudenzi (1889: 58) describes RVA as 'exceedingly frequent' in Bolognese. RVA is reported to apply regularly also in Ferrarese (Baiolini & Guidetti 2005). Bertoni (1905: 43) documents the presence of RVA in the Modena variety, where it applies in an asymmetric fashion: while regressive assimilation of voicelessness for plosives is systematic (Old French *bouton* > [ptoṵn] 'button', *BECC-ARIU(M) > [pkær] 'butcher', *BOCC-ONE > [pkoṵn] 'mouthful'), the assimilation seems optional in the case of sibilants (VESICA > [vsiga]/[psiga] 'bladder') and in the case where the second consonant of the cluster is voiced (PEDALE > [pdæl] ~ [bdæl] 'pedal'). Some optionality with respect to RVA of [+voice] is also reported for Grizzanese by Loporcaro (1998: 162), who mentions [a t ˈvɛd] ~ [a d̥ ˈvɛd] ~ [a d ˈvɛd] 'I see you', where the object clitic /t/ is variably realised as [t], [d̥] or [d].

Besides the few cases just mentioned, the literature thus describes RVA applying in the varieties of Bologna, Ferrara, Modena and Parma as a fairly robust generalisation. The accounts discussed above, however, mainly focus on the diachronic dimension and provide lists of forms showing RVA, rather than morphologically correlated pairs exhibiting RVA in action.

In the present study, we investigate whether speakers of Emilian varieties synchronically apply RVA in production, and then check whether RVA influences speech perception by testing the perception of C_1C_2 obstruent clusters in which C_1 is voiceless and C_2 voiced. In addition to this empirical contribution, which provides an experimental ground to observations regarding RVA reported in the literature, as well as new pieces of evidence for the role of phonology in speech perception, we also present a theoretical modelling of our findings. The latter represents a contribution to the debate concerning the phonetics-phonology interface and, more generally, the architecture of the grammar, as it challenges traditional production-oriented models, for which the application of the same

[2] In this paper, by Italian dialects, we refer to the Italo-Romance varieties descending directly from the Latin spoken in the Italian peninsula, more specifically, to their synchronic grammatical systems. For detailed discussions on the complex (sociolinguistic) status of such varieties, we refer the interested reader a.o. to Maiden & Parry (1997) and Loporcaro (2013).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

phonological process both in production and perception poses problems. The role played by phonological knowledge in perception is not easy to model in traditional rule-based generative theories, which restrict their formalisation to the production process, that is, the mapping from underlying to surface form. In such models, one could think about perception as a process of rule inversion (Leben & Robinson 1977), which, though, has been shown to come with several problems (Churma 1981).

Optimality Theory (Prince & Smolensky 1993; henceforth: OT), with its evaluation of the best output given a certain input, lends itself to the formalisation of any decision mechanism, hence, also for the formalisation of the perception process. Nevertheless, most OT models are restricted to formalising the phonological production, where phonotactic restrictions apply to the output, whereas perception has only indirect influence via constraints referring to extra-grammatical information on perceptibility (as represented, e.g. in the p-map by Steriade 2001).

We remedy these shortcomings by providing a formal account of how phonological restrictions and auditive cues interact in RVA production and perception, using the BIDIRECTIONAL PHONETICS AND PHONOLOGY optimality theoretic model (henceforth: BiPhon; Boersma 2007, 2011; Boersma & Hamann 2009), where one and the same set of phonotactic constraints triggering phonological processes hold both in production and in perception.

This article is structured as follows. Section 2 covers the experimental part, describing the production data illustrating that RVA is a productive process in Emilian dialects (Section 2.1), and a segment detection task testing the influence of RVA on speech perception (Section 2.2). Section 3 provides a formal account of our experimental findings in BiPhon. Section 4 discusses our results in the context of recent studies on speech perception resorting to Bayesian reverse inference, and Section 5 concludes.

2. EXPERIMENTAL EVIDENCE OF REGRESSIVE VOICING ASSIMILATION

The following data provide experimental evidence of RVA in the Emilian dialects spoken in Parma, Modena, Bologna and Ferrara, and tests the production and perception of the labial plosives /b/ and /p/. Our restriction to labial plosives has purely practical reasons, as a systematic testing of all places of articulation would have resulted in a very long experiment that would have exceeded the attention span of the participants.

All data have been collected in a set of fieldwork sessions performed in 2017, with 13 participants. Apart from P4, all speakers were male. The relevant details are given in Table 1, where Age refers to the participants' age in 2017.

All participants have lived in the respective regions since their birth, and are native speakers of the respective dialects, which they use on a daily basis.[3] They are

[3] Participant P10 was born in Asmara, the capital of Eritrea. His parents were Italian immigrants from Parma and moved back to their hometown right after his birth.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

Participant	Provenience	Age	Birthplace	Elicitation task	Perception task
P1	Bologna	84	Bologna	✓	✓
P2	Bologna	70	Budrio	✓	✓
P3	Bologna	78	Bologna	✓	✓
P4	Ferrara	69	Ferrara	✓	✓
P5	Ferrara	76	Ferrara	✓	✓
P6	Ferrara	85	Ferrara	✓	✓
P7	Modena	72	Campogalliano	✓	✓
P8	Parma	77	Parma	✓	✓
P9	Parma	74	Parma	✓	✓
P10	Parma	74	Asmara	✓	✓
P11	Parma	61	Parma	✓	✓
P12	Parma	74	Noceto	—	✓
P13	Parma	70	Parma	—	✓

Table 1
Participant information.

all speakers of (regional) Italian too, which they learnt at school and use in more formal contexts. Mean age of our speakers was 74 years. We employed older speakers, as they are more competent in their dialect. Dialect competence was assessed based on peer-declaration. The choice of having only older speakers was determined by the language shift towards standard Italian that has been going on in the last decades, which makes it difficult to find proficient dialect speakers among the youth (especially in northern Italy; for precise quantitative data and discussion, see Manzini & Savoia 2005: 29–34 and Loporcaro 2013: 180f.). None of the participants reported any hearing problems. Participants P12 and P13 did not participate in the elicitation task; however, all 13 took part in the perception experiment.

The participants were first interviewed and recorded, and then performed the perception experiment. The whole session lasted about 25 minutes. The sessions took place in a quiet room at the participant's home and were performed by means of the Praat computer software package (Boersma & Weenink 2017), installed on a MacBook Air (OS X El Captain, version 10.11.6). The recordings were made with the built-in microphone positioned in front of them, with a sampling rate of 48 kHz.

2.1. *Elicitation task*

For this small-scale task, we elicited the morphologically correlated forms given in Example (3) by means of a series of questions that forced the participants to produce the dialectal forms without the interviewer producing the corresponding standard Italian forms. For instance, the form BOCCA 'mouth' was elicited by asking the participant 'how do you call this in your dialect', while indicating the mouth. Such questions were followed by a further question that would prompt the participant to repeat the relevant form in a post-vocalic context, for example, 'so this is…?'

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

(expected answer: LA/UNA BOCCA). Most speakers produced the words once, some rendered a repetition. Many of the forms occurred in utterance-medial or -final position, preceded by an article, a clitic subject pronoun or a preposition, all ending with a vowel.

(3)	STANDARD ITALIAN FORMS	EXPECTED REALISATIONS	MEANING
(a)	bocca - boccone	[ˈbakˑa] - [ˈpkæŋ]	'mouth' - 'mouthful'
(b)	becco - beccheria/ beccaio	[ˈbɛːk] - [pkaˈria]/ [ˈpkɛːr]	'buck' - 'butcher'
(c)	becco - beccata/ beccare	[ˈbɛːk] - [ˈpkɛːda]/ [ˈpkɛːr]	'I peck' - 'peck'/'to peck'
(d)	peso - pesare	[ˈpa͡iz] - [ˈbzɛːr]	'I weigh' - 'to weigh'
(e)	piede - pedale	[ˈpa] - [ˈbdɛːl]	'foot' - 'pedal'
(f)	piede - pedana	[ˈpa] - [ˈbdɛːna]	'foot' - 'platform'

Of relevance are the second forms of each pair in the expected realisations, as they display adjacent segments contrasting in voicing and should, therefore, undergo RVA. In particular, in the clusters in Examples (3a, b, c), we expect /b/ to be realised as voiceless due to following /k/, whereas in Examples (3d, e, f), we expect /p/ to surface as voiced due to following /z/ or /d/ (note that the quality of the stressed vowel on the left side of the Expected realisation column can vary from dialect to dialect; this has no consequence for RVA).

In this elicitation task, we focus on forms with an initial CC cluster because those are the ones that result from the very productive morphological process of suffixation, which, crucially, triggers RVA: due to the stress shift triggered by suffixation, the vowel of the base gets unstressed and dropped and the two relevant Cs result adjacent to each other, feeding RVA.

The recordings of the participants were acoustically analysed in Praat to check whether RVA was applied. Though plosive voicing can be conveyed by several acoustic means, we restricted our analysis to the presence or absence of a voice bar during stop closure. The literature has shown that Italian voiced stops are characterised by the presence of a voice bar throughout the whole duration of the closure, and voiceless stops by its complete absence during closure, and that this presence/absence is the most important perceptual cue (Pape & Jesus 2015: 225; Vagges et al. 1978). Lacking any evidence supporting the opposite, we assume that this holds for the varieties under consideration, too. Other potential cues to voicing reported for Italian (by, e.g. Esposito 2002) are duration of the preceding vowel, duration of release and frequency of f0 in the following vowel. The structure of our data, though, did not allow us to rely on these cues. The first cue could only be checked for if the relevant form was preceded by a vowel-final form (e.g. an article or a clitic). As this is not the case for all the forms in Example (3), we could not rely on this cue throughout the whole study, and we decided not to consider it. The other two cues cannot be relied on either because of the cluster-initial position of the stop, so we left them out.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

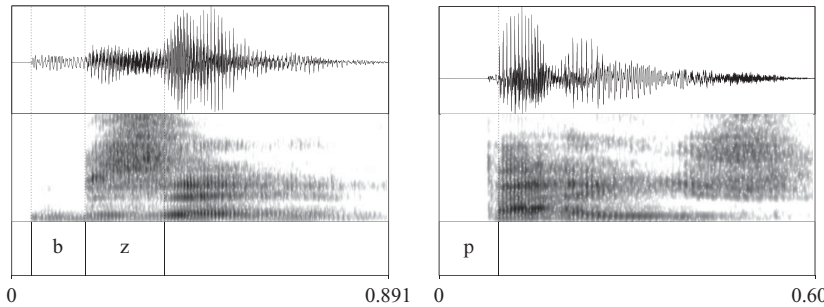

Figure 1
Illustration of regressive voice assimilation in the word [ˈbzɛːr] 'to weigh' on the left, compared to the voiceless realisation of the corresponding initial plosive in [ˈpaiz] 'I weigh' on the right (speaker P1).

RVA of voicing is illustrated with the spectrogram on the left of Figure 1, where an underlyingly voiceless plosive /p/ is produced by P1 as fully voiced in the word [bˈzɛːr], as in Example (3d): the plosive displays a very clear voice bar of considerable duration (102 ms), whereas in the non-RVA context in [ˈpa̰iz] on the right, the voice bar is completely absent.

When our speakers applied RVA to underlyingly voiced stops, it was always categorical, that is, there was no partial devoicing, as shown by the total absence of a voice bar for the respective bilabial plosive. For the cases of voiced stops directly preceded by a vowel in the preceding word, application of RVA was also categorical and nongradual, as could be ascertained by the presence of a voice bar throughout the complete closure. For the cases of voiced stops not directly preceded by a vowel, the beginning of the closure phase could not be determined, and, hence, it could not be inferred whether the complete closure was voiced. We, therefore, also measured the duration of the voice bar of all underlyingly voiceless plosives that underwent RVA and compared it to the voice bar duration in the underlyingly voiced plosives in non-RVA context, namely, the first words in the pairs in Examples (3a)–(3c). The results of these measurements, given in Appendix A, show that the voice bar duration of [b] from underlying /pD/ is similar and often even longer than that of [b] from underlying /b/ in non-RVA context. We interpret these results as a categorical application of RVA in pD words. However, our participants did not always apply RVA in the context where it could be applied. This is summarised in Table 2, where the application of RVA is split by participant and token.

As can be seen in Table 2, four speakers applied RVA in every applicable context, four in 80% of the cases, two in 75% and one in 60%. On average, the speakers applied RVA to the labial plosive in 86% of the cases.

There were 7 of the 11 speakers that produced a vowel between the two relevant consonants for Example (3e) ([pəˈdɛːl; see the top of Figure 2 below) and therefore could not apply RVA. The same result happened for Examples (3a) and (3f) produced by two speakers ([bəˈkæŋ] and [pəˈdɛːnɐ], respectively), indicating that

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

Participant	(3a) [pkæŋ]	(3b) [pkɛːr]	(3c) [pkɛːda]	(3d) [bzɛːr]	(3e) [bdɛːl]	(3f) [bdɛːna]	RVA(%)
P1	*vowel*	yes	yes	yes	yes	yes	100
P2	yes	—	no	yes	yes	yes	80
P3	yes	yes	yes	yes / no	yes	yes / no	80
P4	yes	yes	yes	yes	*vowel*	yes	100
P5	*vowel*	yes	yes	yes	*vowel*	*vowel*	100
P6	yes	yes	yes	yes	*vowel*	no	80
P7	yes	yes	—	yes	*vowel*	*vowel*	100
P8	no	yes	*vowel*	yes	*vowel*	yes	75
P9	yes	yes	yes	yes	no	—	80
P10	no	yes	—	yes	*vowel*	—	67
P11	no	yes	yes	yes	*vowel*	—	75

RVA: regressive voicing assimilation; *vowel*: a vowel occurred between the two relevant consonants; —: the speaker did not produce the word or background noise did not allow a decision on voicing. The last column summarises how often a speaker applied RVA (in percent of total possibilities to apply RVA).

Table 2
Results of the elicitation task: application of RVA (yes, no) split by participant and token.

these speakers might not have been familiar with the dialectal forms of the respective words, possibly due to the low frequency of these forms.

One speaker – P3 – produced two forms each for Examples (3d) and (3f), one with RVA applied and another without; this is illustrated at the bottom of Figure 2 with spectrograms of the word as in Example (3f), with RVA (on the left) and without RVA (on the right). The speaker did not comment on the two different pronunciations, but given their low frequency, it is reasonable to assume an influence of the standard Italian forms, which display no syncope and therefore no RVA.

The first plosives in both realisations at the bottom of Figure 2 have clear release bursts with noise of considerable duration (40 ms and 34 ms), indicated by dotted lines and the word 'burst' on top of the figures, probably due to very careful pronunciation. A less careful pronunciation can be seen in the realisations in Figure 3. The /p/ burst on the right is stronger/noisier than that of the /b/ on the left, as expected for a voiceless release (see, e.g. Repp 1979 for English and van Dommelen 1983 for French; both studies also show the relevance of burst amplitude as perceptual cue to voicing), though periodicity (due to voicing) starts in the later part of this burst. Vowel-like formants, like those of the inter-plosive vowel at the top of Figure 2, indicated by 'vowel' on top of the figure, are absent from the spectrograms of the bursts of the initial plosives at the bottom of Figure 2, and there are no vowel-like complex periodic patterns in the corresponding oscillograms either. We therefore interpret these burst noises as not containing any excrescent vowel (see, e.g. Miatto et al. 2019 for a similar definition with the additional criterion that excrescent vowels need to be at least three glottal cycles long).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

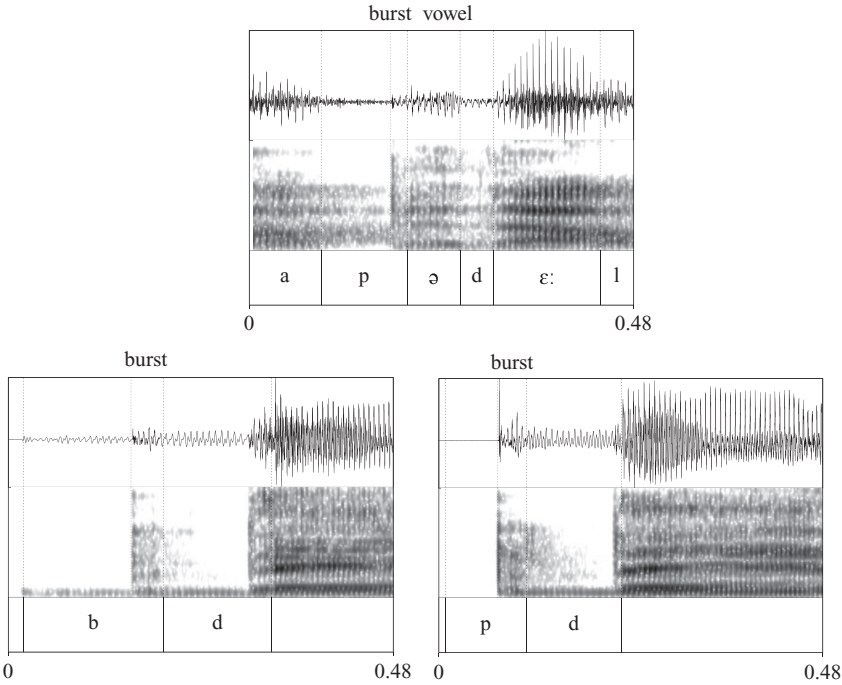

Figure 2
Top image: Oscillogram and spectrogram of [a pəˈdɛːl] 'I pedal', displaying a vowel between the two plosives (speaker P6). Bottom images: Oscillograms and spectrograms of the relevant parts of [ˈbdɛːna] (with RVA) left and [ˈpdɛːna] (without RVA) right (both speaker P3).

2.2. *Segment-detection task*

In a perception experiment, we tested whether our participants detect a /p/ followed by a voiced obstruent. For instance, given a nonce word, such as [apda], we tested whether our participants perceive the /p/ as such, or whether they apply RVA in perception and perceive the obstruent as assimilated, namely, as /b/. Since this experiment required a considerable amount of concentration from the participants, we restricted ourselves to testing regressive assimilation of voice, as in Example (2a), and did not include the assimilation of voicelessness, as in Example (2b). We employed a forced-choice segment-detection task (Zimmerer & Reetz 2014), where participants had to press either 'b' or 'no b' after every stimulus word they heard. The overall duration of the experiment was around 20 minutes. Participants could cope well with the experiment, and it was not too demanding, as an exploratory statistical analysis of the correctness of answers over time showed.[4]

[4] The correctness of answers to all stimuli excluding those to the pD words (because these words did not have a default correct answer, see explanation in Section 2.2.1) was tested in a generalised

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

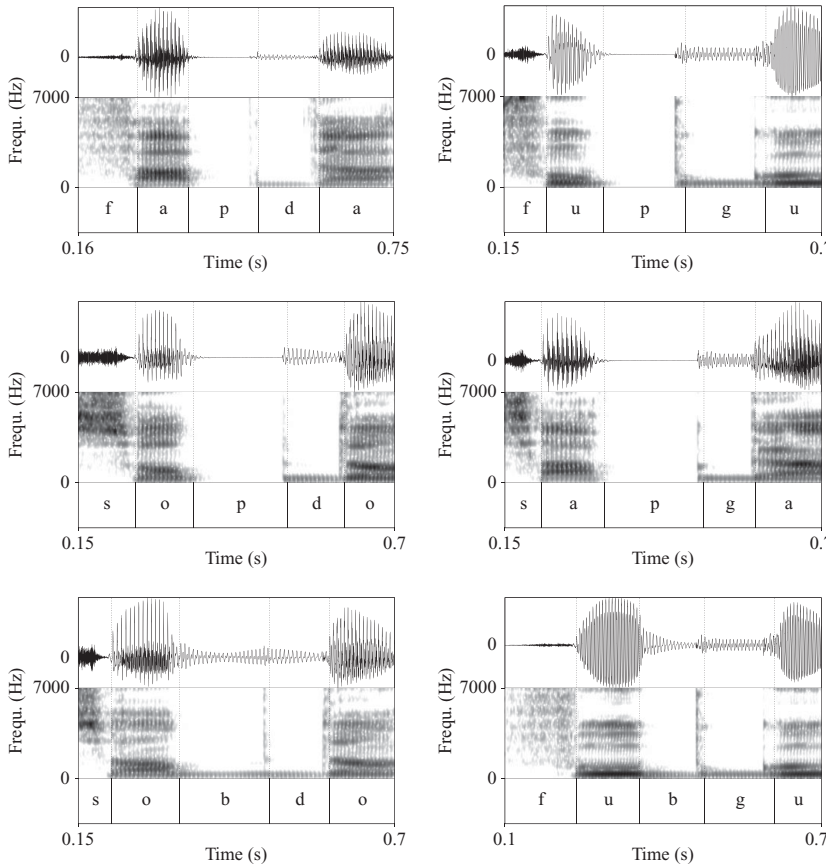

Figure 3
Oscillograms and spectrograms of the pD stimulus items /fapda/, /fupgu/, /sopdo/ and /sapga/ (first two rows) and of the bD stimulus items /sobdo/ and /fubgu/ (bottom row).

2.2.1. *Stimuli and procedure*

All stimuli were bisyllabic nonce words of the form CVC(C)V, with two identical vowels of the set /a e i o u/ and stress on the first vowel. There were 16 test items that had a medial cluster with /p/ followed by a voiced obstruent of the set /d g z/ (henceforth: D), referred to in the following as pD words (cf. Example (4a) for examples). We decided, for this post-vocalic occurrence of the relevant pD cluster, to ensure that participants could use the end of the preceding vowel as an indication of the beginning of the voiceless closure phase. This post-vocalic RVA environment can

linear mixed effects model with stimulus number (as a measure of time) as predictor (cf. - Section 2.2.2 for further details on the models we employed). The outcome was not statistically significant ($p = 0.308$), ruling out an accuracy degradation effect.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

also be found in natural speech in finite verbal forms, which are often preceded by vowel-final clitics, and in nominal forms preceded by, for example, vowel-final articles.

A further 16 items were identical to the first set but had a medial cluster with /b/ followed by a voiced obstruent, referred to as bD words (cf. Example (4b)). All these items had a fricative or affricate as onset consonant.

(4)	Example stimuli	Type of stimuli
(a)	/ˈfopdo/, /ˈsupgu/, /ˈtʃipzi/	pD word
(b)	/ˈfobdo/, /ˈsubgu/, /ˈtʃibzi/	bD word
(c)	/ˈpaka/, /ˈtʃepe/, /ˈsupu/	p word
(d)	/ˈbuvu/, /ˈzobo/, /ˈvaba/	b word
(e)	/ˈsuku/, /ˈtʃidli/, /ˈgolo/	filler

Furthermore, we included 48 items with /p/ or /b/ in nonassimilating position (cf. Examples (4c) and (4d)), in either initial or medial position and 122 fillers without /b/ or /p/ (cf. Example (4e)). This amounted to a total of 202 stimuli.

For the initial training, we employed an additional list of 16 words of the same CVC(C)V structure as test stimuli. Of those, 6 had a target /k/ in either initial or medial position and 10 contained no /k/. None of these training words involved a context where voicing assimilation could apply.

Each stimulus was read several times by a phonetically trained native speaker of Italian, recorded in a soundproof booth at a 44.1 kHz sampling rate. It was not difficult for the speaker to produce such stimuli, as standard Italian allows /pD/ sequences both across word boundaries (e.g. *sto*[p d]*ietro* ‘stop after’) and within words (in borrowings, e.g. [futˈbollə] ‘football’, as shown by Huszthy 2016). From the recordings, we selected one token for each stimulus, controlling the test items for the total absence of epenthesis and partial voicing (i.e. we selected pD words whose p part was completely voiceless and bD words whose b part was completely voiced). The stimuli were then normalised to a mean intensity of 60 dB. In Figure 3, we give six examples of stimulus items with two plosives, which illustrate that neither vowel-like formants after the release of the first stop were present nor partial voicing during the closure of the first stop. Furthermore, the stimulus items have no or very short burst releases (especially obvious when compared to the careful pronunciation of the words in the elicitation task in Figure 2, bottom). We follow Henderson & Repp (1982) in categorising such bursts as inaudibly released: ‘visible release burst in records of the signal, but not readily detectable by ear’ (p. 79). See also the overview in Wright (2004) on the difficulty to perceive very short bursts.

Participants had to read an instruction text, which was translated into the specific dialects to ensure that they activated the participants’ dialect (see, e.g. Grosjean 2001; Yazawa et al. 2020 on the importance of language mode in perception studies). In order to minimise a priming effect from standard Italian, we adopted the spelling convention that is considered ‘standard’ by most of the associations preserving and promoting the relevant dialects (Vitali & Pioggia 2014; Vitali 2020).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

The translation of the instructions was made by Daniele Vitali. No participant showed disagreement with the translation. The instruction explained that they would hear words via headphones. In the introduction phase, they had to indicate as quickly as possible for each word whether it contained a [k], by clicking <f> on the keyboard, or not, by clicking <j>. We chose these keys because, in a qwerty keyboard, they are symmetrically placed at the center of the keyboard and can be easily reached with the left and right index fingers, respectively. This should allow for minimising the reaction time. After this introduction, the participants had time to ask the instructor questions. Another instruction text in their dialect then explained that they now had to detect the presence or absence of [b] in each word, by using the same keys. The 202 stimuli of the experiment were presented in randomised order, with a self-timed break after every 51 stimuli. All stimuli were presented via headphones with an ExperimentMFC script in Praat, which collected both response category and reaction time for every stimulus.

To answer our research question – whether RVA influences the perception of voiceless [p] before voiced obstruents – we planned to compare 'b'-responses for pD words to those of p words: had RVA no effect on perception, then the responses to these two categories should be very similar. If, however, RVA did influence perception, then there should be considerably more 'b'-responses to pD words than to p words.

2.2.2. *Analysis and results*

We analysed the responses to all items, as in Examples (5a–d) (80 x 13 participants = 1,040). There were 25 of them that had to be excluded because they were faster than 500 ms or slower than 5 s. Many of the excluded responses had a negative reaction time, indicating that participants pressed an answer button before they had heard the stimulus. We decided for a rather long reaction time window of 5 s, because our participants were elderly and were not used to performing psycholinguistic experiments. An overview of the results is given in Figure 4.

To test the validity of our perception experiment and whether our participants paid attention during the experiment and were able to perform it, we checked their performance on the p words and b words. Participants responded with 'b' to b words in 85% of the cases, and to p words in 4% of the cases. Based on these two stimulus types, we calculated mean accuracy rates per participants (where 'b'-responses to b words and 'no b'-responses to p words were considered correct), as given in Table 3.

Accuracy rates for p words ranged between 76% and 100%, with most participants reaching ceiling level, and those for b words between 65% and 100% (only one participant with ceiling performance). This shows that our participants paid attention, were able to perform the test and to perceive the stimuli correctly, and that they did not suffer from any hearing impairment. The accuracy is nevertheless lower than what is usual in perception experiments, likely due to two factors. Firstly, the testing did not take place in the lab but in a quiet room at the participants' home (see,

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

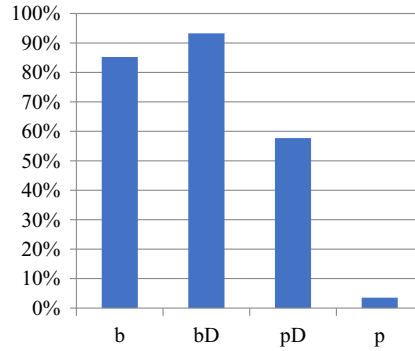

Figure 4
Percentage of 'b'-responses to the categories: b = initial or medial /b/; bD = assimilated cluster; pD = nonassimilated cluster; p = initial or medial /p/.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
p word	1.00	1.00	1.00	0.79	1.00	1.00	0.95	1.00	1.00	1.00	1.00	0.76	1.00
b word	0.65	0.96	0.83	0.92	0.92	0.82	0.90	0.78	0.83	0.88	0.96	1.00	0.92

Table 3
Mean accuracy rates to b words and p words per participant.

e.g. Phatak et al. 2008 on the influence of noise on the perception of voicing), and secondly, our participants were elderly (see, e.g. Strouse et al. 1998 who found that elderly with normal hearing performed poorer in perception experiments).

The comparison of 'b'-responses for pD words to those of p words, which allows us to answer our research question, resulted in considerably more 'b'-responses to pD words than to p words, as can be seen in the two rightmost columns of Figure 4: while mean percentage of 'b'-responses to p words is a mere 4%, it is 58% to pD words. The percentage of 58 indicates that the participants perceived these stimuli not consistently but sometimes as containing a [b] and sometimes a [p]. As shown by classical studies on categorisation, performances at 50% indicate that participants are not sure to which category the stimuli belong (Liberman et al. 1957).

We tested the significance of this difference with a generalised linear mixed effects model (logistic regression) in R (glmer from the package lme4; Bates et al. 2015) with the binary response 'b' or 'no b' as dependent variable, item (pD word and p word) as within-subjects factor, a random intercept per word and per participant and a random slope per participant for item. Our participants gave significantly more 'b'-responses to pD words than to p words ($p = 0.00587$; confidence interval [C.I.] of odds ratio: $75..1.0{\cdot}10^{8}$). We conclude from this that Emilian speakers are influenced in their perception of pD words by the phonological process of RVA. The between-participant standard deviation (SD) in the model is

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

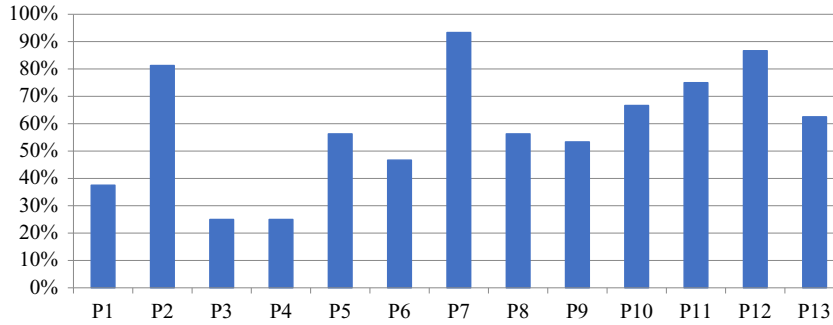

Figure 5
Percentage of 'b'-responses to pD words split by participants.

reported as 2.565 (log-odds), which we interpret as significant inter-speaker variation.[5] Figure 5 shows the percentage of 'b'-responses to pD words split by speakers, illustrating this high individual variation in the responses, ranging from 25% (for P3 and P4) to 93% (for P7), with a mean of 58%.

The mean reaction time (RT) to p words was 1.188 s, with a SD of 0.356 s; pD words was 1.469 s, with a SD of 0.524 s. We tested this difference in RT with a linear mixed effects model in R. For this, we normalised the RT values by first ranking them and then applying an inverse cumulative normal distribution to the ranked values.[6] Again, we used item (pD word and p word) as within-subjects factor, a random intercept per word and per participant and a random slope per participant for item. Our participants had a significantly longer RT to pD words than to p words ($p = 0.0000201$). This is as expected for stimuli with conflicting information.

2.3. *Discussion of experimental results*

In Section 2.1, we saw that the speakers of the Emilian varieties from Parma, Modena, Bologna and Ferrara all applied RVA, in a high percentage of cases (86%). The production data thus show that RVA is a synchronically active process, though not obligatory for all speakers in all cases.

The segment detection experiment in Section 2.2 shows that RVA also influences the perception process but, again, not systematically in all cases: the participants reported to have perceived a 'b' in pD words in 58% of the cases, and this was significantly more often than they reported for p words (4%). Participants

[5] A Monte Carlo simulation (with 10 million replications), under the null hypothesis that all participants have the same /b/ probability of 0.6, shows that the chance of finding a between-participant standard deviation greater than the one observed (namely, 3.31) is $p = 0.00037$. We therefore have strong evidence of individual variation.

[6] In R, we used the formula: qnorm ((rank (reactionTimes) – 0.5) / length (reactionTimes)).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

considered /p/ in RVA context sometimes as voiced, thus showing an influence of RVA, and sometimes as voiceless, showing the impact of the present auditory cues, in this case, the silent closure phase. The fact that RVA did not fully determine the outcome of their perception suggests that phonological knowledge cannot override all perceptual cues, and that speech perception is an integration of auditory cues and phonological restrictions and processes. The conflict between these two types of information is reflected in the variation observed in the listeners' answers. For the same reason, perception experiments on so-called illusory vowels show similar 'non-categorical' results: In their second experiment (Dupoux et al. 1999), Japanese listeners reported an illusory [ɯ] in 59% of the tokens, and in an identification task (Durvasula et al. 2018), Mandarin listeners reported an illusory [i] in 29% of the tokens.

We also found individual variation with respect to the alignment of the results of the production and the perception experiment, as shown in Table 4.

While for participants 10 and 11, the percentages of producing and perceiving a /b/ in RVA context are identical, for all other speakers, the percentage of perceiving /b/ is lower than producing it, with an extreme difference in participant 4 with 100% versus 25%.

As we show in the following section, the stochastic implementation of BiPhon allows for the formal modelling of the observed individual variation, whereas its three-level architecture allows to account for the misalignment of the production and perception results. As discussed below, this would not be possible in more traditional approaches assuming a two-level grammar architecture.

3. A FORMAL ACCOUNT

In this section, after we present a formalisation of Emilian RVA in production (Section 3.1), we illustrate how to formalise the integration of auditory and phonological information accounting for speech perception (Section 3.2). In the final subsection (Section 3.3), we show how this model can account for the observed variation.

Before formalising RVA in the two processing directions, a word on our choice of voicing feature is in order. As RVA in Emilian dialects is triggered both by voiced and voiceless obstruents but not by sonorants, we employ a binary feature

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
Production of /b/ (%)	100	80	80	100	100	80	100	75	80	67	75	–	–
Perception of /b/ (%)	38	81	25	25	56	47	93	56	53	67	75	87	63

Table 4
Percentage of production of /b/ in RVA context and perception of /b/ in RVA context (in pD words) per participant.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

[±voice], where [–voice] is as active as [+voice] (Rubach 1997, 2008; Wetzels & Mascaró 2001), and the inactivity of sonorants is due to them lacking any voicing specification. In doing this, we depart from approaches proposing the privative feature [voice] (Lombardi 1995a, 1999), as privative [voice] leads to several theoretical and empirical problems (Kim 2002). For instance: (i) it does not allow to formalise the three-way contrast [+voice] VERSUS [0voice] VERSUS [–voice] required in some languages (Inkelas & Orgun 1995; Krämer 2000; Wetzels & Mascaró 2001); (ii) it does not allow to account for the phonetic and phonological differences between [–voice] and [0voice] (Dixit 1987; Hsu 1998) and (iii) it requires the introduction of ad hoc stipulations, such as final exceptionality (Lombardi 1995b) to account for languages that have RVA of [–voice] but not [+voice] (Wetzels & Mascaró 2001).

For the modelling of RVA, we employ BiPhon (Boersma 2007, 2011; Boersma & Hamann 2009), whose architecture is given in Figure 6. BiPhon can account for both speech production and comprehension. Production consists in the mapping of underlying to surface form (phonological production) and the mapping from surface to phonetic form (phonetic implementation), analogous to the modularity assumed in psycholinguistic models of speech production (e.g. Levelt 1989).[7] Comprehension consists of the mapping from phonetic to surface form (speech perception) and the mapping from surface to underlying form (word recognition), analogous to psycholinguistic models of speech comprehension (e.g. McQueen & Cutler 1997).

In BiPhon-OT, phonological production (Figure 6, top right) is an interaction of FAITHFULNESS and STRUCTURAL constraints (as in traditional OT, see McCarthy & Prince 1995), and perception (Figure 6, bottom left) is an interaction of CUE and STRUCTURAL constraints. The same STRUCTURAL constraints thus apply to the surface form in both processing directions but interact with different sets of constraints depending on the direction, allowing for a divergence between perception and production, as we have observed in our data.

3.1. *Phonological production*

In this section, the application of RVA in production is formalised. As shown in Sections 1 and 2.1, Emilian varieties display a synchronic process of unstressed

[7] BiPhon includes more levels of representations than shown in Figure 6. The phonetic representation, for example, can be further split into auditory and articulatory representations. As the latter does not play a role in perception, we give a single phonetic representation, corresponding to the auditory form. As for the 'underlying' and 'surface' forms, they refer to phonological structures, which are ontologically different from the phonetic structures the surface forms are mapped onto. This view complies with the modular tenets assumed by a great deal of generative grammar literature and conceives of phonology and phonetics as two distinct modules, where phonology deals with abstract and categorical representations, and phonetics with concrete and continuous objects. The mapping between the two different levels is taken care of by a set of CUE constraints formalising the phonology-phonetics interface.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

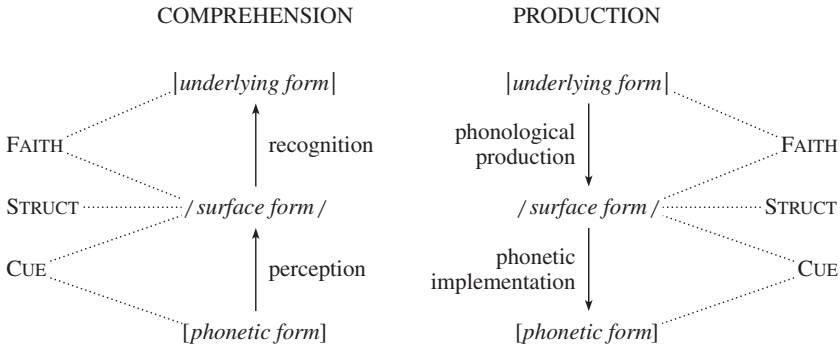

Figure 6
A single three-level model for production and comprehension.

vowel deletion, which feeds RVA. We formalise unstressed vowel deletion as triggered by the STRUCTURAL constraint *V_{weak} (for different incarnations of the reduction-triggering constraint, see, e.g. Crosswhite 2001; Gouskova 2003; Coetzee 2006; de Lacy 2006; McCarthy 2008; Iosad 2012; Cavirani 2015). For the formalisation of voicing assimilation, we resort to the STRUCTURAL constraint AGREE (Lombardi 1999: 272). The latter defines the phonotactic well-formedness of consonant clusters sharing the same voicing specification, and triggers assimilation. The definitions of these constraints are given in Example (5):

(5) (a) *V_{weak} Assign a violation mark if a place-bearing vowel is in a metrically weak, i.e. unstressed, position.
(b) AGREE Assign a violation mark if an obstruent cluster does not agree in voicing.

As for the assimilation direction, following Rubach (2008), we argue that the regressive directionality results from the interaction of a general FAITHFULNESS constraint IDENT[voice] with the more specific IDENT[voice]$_{\_V}$, which formalises a preference for maintaining the underlying voicing specification of segments before vowels. These constraints are defined in Example (6):

(6) (a) IDENT [voice] Assign a violation mark if the feature value for [±voice] on an input segment is not preserved on the correspondent output.
(b) IDENT [voice] $_{\_V}$ Assign a violation mark if the feature value for [±voice] on an input segment is not preserved on the correspondent output segment in prevocalic position.

Further support for an analysis resorting to the constraints in Example (6) is provided by the fact that Emilian varieties show word-final devoicing, as illustrated with the examples from Bolognese in Example (7) (Vitali 2020):

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

(7) [fraːp] ‘blacksmith’ ~ [fraˈbatː] ‘bad blacksmith’
[nuːt] ‘naked$_{M.SG}$‘ ~ [ˈnuːda] ‘naked$_{F.SG}$’
[noːf] ‘new$_{M.SG}$‘ ~ [ˈnoːva] ‘new$_{F.SG}$’

Together with the assimilated patterns described in the previous sections, Example (7) suggests that a [+voiced] segment can only occur before a vowel, a sonorant or another [+voiced] segment.

The working of our four constraints is illustrated in Example (8) with the production of [ˈbzɛːr] in Example (3d).[8] The ranking of AGREE between the two IDENT constraints is motivated by the observed variation (see Section 3.3 below).

(8) Phonological production of [+voice] RVA[9]

|paiz+ɛːr|	*V$_{weak}$	IDENT[voice]$_{\_V}$	AGREE	IDENT[voice]
/paiˈzɛːr/	*!			
/ˈpzɛːr/			*!	
☞ /ˈbzɛːr/				*
/ˈpsɛːr/		*!		*

The structural constraint AGREE and the binary feature [±voice] ensures that RVA also applies in cases where the first obstruent is voiced and the second voiceless. This is shown in Example (9) with the production of [ˈpkɛːr], as in Example (3b).

(9) Phonological production of [–voice] RVA

|bɛːk+ɛːr|	*V$_{weak}$	IDENT[voice]$_{\_V}$	AGREE	IDENT[voice]
/bɛːˈkɛːr/	*!			
/bkˈɛːr/			*!	
☞ /pkˈɛːr/				*
/bgˈɛːr/		*!		*

3.2. *Phonology in speech perception*

The process of speech perception is modelled in BiPhon as a mapping from an auditory onto a surface phonological form (Figure 6, lower left). Compared to production, the STRUCTURAL constraint AGREE still evaluates the surface phonological form, but

[8] The stress shift caused by the suffixation process is not accounted for in the following tableaux, as it goes beyond the scope of this paper.

[9] In contrast to standard OT tableaux, the following tableaux employ ‘| |’ to delimit underlying forms, ‘/ /’ for surface forms and ‘[ ]’ for phonetic forms (following Boersma 2007).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

it now interacts with CUE constraints. In the following formalisation, we focus on the interplay of several cues and a language-specific STRUCTURAL constraint. With this, we provide a simplified formalisation of speech perception, ignoring other kinds of knowledge that might play a role in it. Furthermore, our description is restricted to cues of voicing in plosives because we only employed plosives in our perception experiment. A complete description of all cues to obstruent voicing would go beyond the scope of this paper.

The most reliable cue to voicelessness in /p/, and in plosives in general, is the silence during the closure, transcribed as [_] in the auditory form. If the voiceless plosive is released, a strong labial release burst [^{p}] is another cue to its (place of articulation and) voicelessness (recall the strong burst in Figure 2, bottom right). The auditory cues to voiced plosives are the presence of vocal fold vibration during closure, transcribed as [〰], and a weak (because voiced) labial release burst [^{b}].

How listeners employ the silence and vocal murmur in the closure to correctly perceive the voicing specification of plosives is captured with two CUE constraints given in Example (10).

(10) (a) *[_] /+voice/ Assign a violation mark if the presence of a silent closure in the auditory signal is mapped onto a voiced plosive in the surface form.

(b) *[〰] /–voice/ Assign a violation mark if the presence of a voiced closure in the auditory signal is mapped onto a voiceless plosive in the surface form.

The use of release bursts is captured in a similar way with the constraints in Example (11).

(11) (a) *[^{p}] /+voice/ Assign a violation mark if the presence of a strong, voiceless release burst in the auditory signal is mapped onto a voiced plosive in the surface form.

(b) *[^{b}] /–voice/ Assign a violation mark if the presence of a weak, voiced release burst in the auditory signal is mapped onto a voiceless plosive in the surface form.

The workings of these constraints and the irrelevance of AGREE in nonassimilating contexts is illustrated in Examples (12) and (13), formalising the perception of intervocalic /p/ and /b/, respectively. The use of the symbol [a] in the auditory form is shorthand for specific formant values and should not be confused with a symbolic phonological representation, whereas [ˀ] stands for vowel transitions into a labial plosive. We restrict our illustration to nonce words, as this allows us to exclude the influence of lexical knowledge on speech perception (Ganong 1980; for a formalisation of such a Ganong-effect in BiPhon, see Boersma 2011).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

(12) Perception of a voiceless bilabial plosive in a nonassimilating context[10]

[aˈ_ᵖa]	*[‿‿] /–voice/	*[_] /+voice/	*[ᵇ] /–voice/	*[ᵖ] /+voice/	AGREE
☞ /a.pa/					
/a.ba/		*(!)		*(!)	

(13) Perception of a voiced bilabial plosive in a nonassimilating context

[aˈ‿‿ᵇa]	*[‿‿] /–voice/	*[_] /+voice/	*[ᵇ] /–voice/	*[ᵖ] /+voice/	AGREE
/a.pa/	*(!)		*(!)		
☞ /a.ba/					

A complete perception grammar would also contain constraints like *[‿‿] /+voice/ and *[ᵇ] /+voice/ that avoid that the cues are being mapped onto their corresponding phonemes, but since those constraints would very often be violated by the forms occurring in the language (also by the winning candidates in Examples (10) and (11)), they would be very low ranked. We did not include them in the tableaux for lack of space.

The constraints used in Examples (12) and (13), and in Examples (14) and (15) below, are not ranked with respect to each other (yet). This is because, up to now, we have neither theoretical arguments nor sufficient evidence from perception experiments that could inform us about a possible ranking (but see Section 3.3).

An obstruent cluster that does not agree in voicing causes a conflict between auditory cues and AGREE, as formalised in Example (14). The auditory input in this tableau does not occur natively in Emilian but reflects the pD words we presented to the participants in our segment detection experiment (Section 2.2). As we have shown and explained in Section 2.2.1, the first plosive in a cluster of two plosives as given here is usually not released (hence, we do not include a burst for it in our modelling), and the second plosive has no vowel transitions into the closure. As shown by the transcription of the burst release and the respective CUE constraints, the second consonant is a coronal plosive.

(14) Perception of a voiceless plosive in an assimilating context

[aˈ_‿‿ᵈa]	*[‿‿] /–voice/	*[_] /+voice/	*[ᵈ] /–voice/	*[ᵗ] /+voice/	AGREE
☞ /ap.da/					*
☞ /ab.da/		*			
/ap.ta/	*(!)		*(!)		

The evaluation results in two winning candidates, the first not assimilated, the second with RVA, mirroring the two possible answers we received in our

[10] A candidate marked twice with '(!)' indicates that one cannot determine which of the two violations is fatal, given the indeterminacy of the ranking between the two constraints involved (Mester & Padgett 1994).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

perception experiment. The third candidate shows progressive voice assimilation, thus does not violate AGREE. This candidate does not win because it violates two CUE constraints (it ignores both the weak burst and the presence of voicing murmur in the closure), while the second candidate with regressive assimilation violates only one (the silence during closure). Note that the STRUCTURAL constraint AGREE is satisfied in perception in a very different way from what we saw in production: here, CUE constraints determine the best output, while in production (Examples (8) and (9)), the best output was selected by the FAITHFULNESS constraint IDENT[voice]$_{V}$.

The tableau in Example (15) shows that, differently from what happens in the tableau in Example (14), in the perception of clusters agreeing in voicing, there is only one winner:

(15) Perception of a voiced plosive in a fully voiced cluster

[a͡ˀ_____d͡a]	*[_____] /–voice/	*[_] /+voice/	*[^{d}] /–voice/	*[^{t}] /+voice/	AGREE
/ap.da/	*(!)				*(!)
☞ /ab.da/					
/ap.ta/	*(!)		*(!)		

3.3. *Variation in the perception and production output*

In Example (14), with a nonassimilated pD word as input, the nonranking of the constraints predicts that both winning forms, /ap.da/ and /ab.da/, should be reported equally often. This does not reflect the speaker-specific results of the segment-detection task in Section 2.2, where participants varied in their 'b'-responses to pD words from 25% (P3 and P4) to 93% (P7). Nor does the nonranking in Example (13) for the perception of voiced bilabial plosive in nonassimilating context, and its winning candidate, /a.ba/, reflect the varied performance of our participants, ranging between 65% and 100% 'b'-responses.

Several reasons can be given for this deviation from the results predicted by the model we proposed up to now. Firstly, there might be extra-grammatical factors at play, such as the fact that the relevant cues might not be fully available in all positions. This might hold for voicing during closure in phrase-initial position: in our segment detection task, half of the b (and p) words had the contrast phrase-initially, where the voice bar is often shorter than in medial position. This could lead to an incomplete input to the perception tableau and could partly explain the observed asymmetry between voiced and voiceless input in the accuracy rates (Table 3). This possibility is, however, not supported by the results: our participants had a similar number of correct answers to initial b words (136) as to medial b words (130).

The asymmetry could also be explained by a grammar-internal factor, namely, a general difference in cue strength between voiced and voiceless plosives: voicing during closure can be easily mistaken as noise, and vice versa, low background noise can be mistaken as voicing. As a result, the perception of voicing during closure might not be as reliable and strong a cue as silence during closure, which, if

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

present, is a reliable indication that the perceived segment is /–voice/. This possible difference in cue strength would predict a difference in the ranking of the corresponding CUE constraints (*[_] /+voice/ >> *[‿] /–voice/) and, hence, a different treatment of voiced versus voiceless input. A second grammar-internal factor to be considered is that listeners might differ in the importance they give to CUE versus STRUCTURAL constraints. This last factor seems to be responsible for the large inter- and intraspeaker variation we observed in pD words (see, e.g. van Oostendorp 1997; Boersma & Hayes 2001; Coetzee 2016 for proposals dealing with variation in terms of constraint ranking or weighting). In the following, we formalise this constraint weightings variation in terms of listener-specific rankings and Stochastic Optimality Theory (Boersma 1997; Boersma & Hayes 2001).[11]

Participants P3 and P4 had 25% of 'b'-responses to pD words, showing that they paid more attention to the acoustic cues of the voiceless plosive than to the restriction on voicing in clusters. For them, we maintain that AGREE is lower ranked than, though very close to, *[_] /+voice, and due to stochastic evaluation, the candidate showing RVA wins in 25% of the cases. This is illustrated with the perception grammar in Example (16), where the first row gives the ranking values of the constraints that result in the correct percentages of winning forms (assuming an evaluation noise of 2.0). These ranking values were calculated in Praat with an OT grammar that learnt the constraint ranking based on 100,000 tokens drawn from an input distribution with the respective percentages (with the Gradual Learning Algorithm, Boersma & Hayes 2001). The ranking between the last two constraints in Example (16) depends on the actual selection points at evaluation time, even though their position on the ranking scale is fixed (98.43 >> 96.54; as indicated with the solid line between them). Due to this variation, we did not use violation marks for the possibly fatal violations of these two constraints.

(16) Perception of a voiceless plosive in an assimilating context by P3 and P4

ranking values:	105.03	105.03	100.00	98.43	96.54
[aˈ_‿^{d}a]	*[‿] /–voice/	*[^{d}] /–voice/	*[^{t}] /+voice/	*[_] /+voice/	AGREE
75% ☞ /ap.da/					*
25% ☞ /ab.da/				*	
/ab.ta/	*(!)	*(!)		*	*

[11] Speech perception involves several grammatical and non-grammatical factors, such as speech rate, order of presentation, prosodic context, semantics, lexical context and other top-down influences, working memory and attention (including uncontrolled influences, such as tiredness, emotional distress, psychopharmacological interventions), exposure to other languages, hearing loss, etc. As made clear above, we controlled for many of these factors: order of presentation via token randomisation, prosodic context via the stability of stress position, semantic, lexical and other top-down effects via the use of nonce words, hearing loss via dedicated questions and other language exposure via the exclusion of a speaker fluent in Czech, a variety that has RVA (Dvoržák 2010) and may thus influence the participant's performance. Apart from these, we need to abstract away from some of these factors, but we believe that this idealisation is standard practice in scientific studies.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

Tableau (17) (Example (17)) is the perception grammar of P7, who gave 93% 'b'-responses to pD words. For this participant, we assume that he was more guided by the structural restriction of his language, and, therefore, has a reverse ranking of the relevant STRUCTURAL and CUE constraints, and a larger distance between the two, mirroring the observed performance (ranking values were calculated as above):

(17) Perception of a voiceless plosive in an assimilating context by P7

ranking values:	105.74	105.74	100.00	99.18	95.07
[a͡ _‿ ^{d}a]	*[‿] /–voice/	*[^{d}] /–voice/	*[^{t}] /+voice/	AGREE	*[_] /+voice/
7% ☞ /ap.da/				*	
93% ☞ /ab.da/					*
/ab.ta/	*(!)	*(!)		*	*

We also observed variation in the production experiment that was not reflected in our formalisation up to now. The production process formalised in Examples (8) and (9) predicts that RVA always applies. In our production experiment (Section 2.2), only four participants showed this systematic application, the remaining seven participants producing 67% to 80% assimilated forms. The variation in the behaviour of these seven participants can be accounted for by assuming that, in their grammar, AGREE is ranked close to IDENT[voice], and that due to stochastic evaluation, the candidate violating AGREE, in that, the nonassimilated form, can win. This is illustrated in Example (18a), representing the production of |paɪ̯z+ɛ:r|, as in Example (3d) by P2, P3, P6 and P9 (80% RVA), and Example (18b), representing the same form produced by P10 (67% RVA) (calculations of the percentages were performed as above and are based again on an evaluation noise of 2.0).

(18a) (Non)application of RVA in phonological production by P2, P3, P6 and P9

ranking values:	107.45	100.00	97.45	95.09
|paɪ̯z+ɛ:r|	*V_{weak}	IDENT[voice]$_V$	AGREE	IDENT[voice]
/paɪ̯ˈzɛ:r/	*!			
20% ☞ /ˈpzɛ:r/			*	
80% ☞ /ˈbzɛ:r/				*
/ˈpsɛ:r/		*!		*

(18b) (Non)application of RVA in phonological production by P10

ranking values:	106.68	100.00	97.32	96.02
|ˈpaɪ̯z+ˈɛ:r|	*V_{weak}	IDENT[voice]$_V$	AGREE	IDENT[voice]
/paɪ̯zˈɛ:r/	*!			
33% ☞ /pzˈɛ:r/			*	
67% ☞ /bzˈɛ:r/				*
/psˈɛ:r/		*!		*

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

4. Alternative accounts

The main alternative theoretical accounts for the influence of phonological alternations on the process of speech perception are Durvasula & Kahng (2015, 2016); Durvasula et al. (2018) and Daland et al. (2019). They are inspired by Bayesian models of speech perception and conceive of perception as REVERSE INFERENCE, by which the listener identifies 'the best estimate of the intended underlying representations of the utterance given their phonological/phonetic knowledge and the acoustics of the utterance' (Durvasula et al. 2018: 1).[12] They all build on data collected in rigorous experimental settings, provide excellent descriptions of the phenomena they deal with and, crucially, make clear that to understand speech perception, we need to integrate top-down phonological expectations and bottom-up acoustic properties. From this point of view, they are thus comparable to our approach (as also stated by Daland et al. 2019), but they also differ from the model we propose in several respects. Despite the relevant results they obtain, we think that these differences suggest that an approach along the lines we developed in this paper might represent a step forward with respect to previous work.

In their work on illusory vowel perception by Korean speakers, the Durvasula & Kahng (2015) study shows that the quality of the epenthetic vowel depends on language-specific phonological processes, therefore providing evidence for a role of phonology in speech perception. They claim that the presence of a phonological vowel deletion process, formalised as /V_1/ → [∅], supports the inference of the inverse process in speech perception, formalised as [∅] → /V_1/ (p. 390). For Korean /ɨ/, they propose the phonological rules in Example (17):

(17) /ɨ/-deletion rules (in production)

ɨ → ∅ /__ V	/k^hɨ+ədo/ → [k^hədo]	'although (it is) big'
ɨ → ∅ / V__	/k^ha+ɨnɨ/ → [k^hani]	'because we go'

Such vowel-deletion processes are argued to 'increase the *global* probability of reverse inference to [ɨ] when there is no vowel correspondent in the acoustic token' (Durvasula & Kahng 2015: 390). The presence of such processes is, thus, one of many factors contributing to the retrieval of the underlying form. Despite the plausibility of this proposal, Durvasula and Kahng do not provide an account of other factors, such as the phonological context, nor, most importantly, a quantification of the influence of the relevant rules on the calculation of the posterior probability, which hampers the possibility of formulating testable predictions. As shown in Section 3, we maintain that our model represents a step forward with respect to Durvasula and Kahng's because it allows for the explicit formalisation and quantification of the influence of the relevant factors.

[12] We excluded the extensive work that exists on so-called perceptual compensation, where listeners undo context-induced coarticulation to retrieve the correct lexical entry (Gaskell & Marslen-Wilson 1996; Mitterer & Blomert 2003, a.o.), as, differently from the one discussed in the paper, this process is a phonetic one (e.g. it is not triggered by phonotactic restrictions), and is influenced by lexical access (whereas we used nonce words).

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

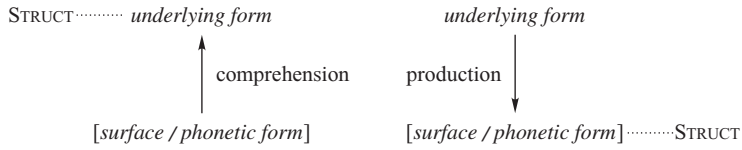

Figure 7
A two-level model for production and comprehension.

Another, more serious problem of their formalisation is the lack of an explicit distinction between phonetic and surface representations. Though they mention the role both of phonological patterns and of phonetic characteristics in the process of speech perception, their formalisation only involves two levels of representation, resulting in an architecture such as the one in Figure 7.

A conflation of phonetic and surface phonological representations in a formal model has several drawbacks compared to the three-level account proposed in Section 3, especially if the model builds on standard OT assumptions (which is admittedly not the case of Durvasula and colleagues). Firstly, a two-level model makes it impossible to distinguish between phonetic and phonological processes, as both apply in the same mapping. A conflation of the two would result in wrong predictions, as easily illustrated with Emilian RVA. Recall that RVA cannot be triggered by sonorants but only by obstruents. While obstruents are phonologically specified for [±voice], sonorants lack a voicing specification, despite displaying vocal fold vibration. A phonetic account of RVA referring to vocal fold vibration/ the presence of a voice bar would, therefore, incorrectly predict that also sonorants trigger RVA. On the other hand, a phonological account where the feature [±voice] spreads due to a phonological restriction (AGREE) correctly describes the process.[13]

Secondly, a conflated representation does not allow to accurately define the involved auditory cues and their interaction with phonological restrictions, and, hence, fails in explicitly weighting the relevance of auditory cues compared to the phonological knowledge. We showed in Examples (14), (15) and (16) that separating phonetic and surface phonological forms allow to: (i) explicitly refer to the auditory information in the input; (ii) explain how this auditory information is mapped onto phonological categories; (iii) explain how this mapping is influenced by structural restrictions and (iv) how listeners can differ in the weight that they give to specific perceptual cues and structural restrictions.

[13] A similar argument is made by Loporcaro (2015), who dismisses approaches conflating phonetics and phonology on the basis of the development of two diachronic changes that occurred in Romance, in that /a/-fronting in (Old) French and /ɛ/-diphthongisation in (Old) Tuscan. It is fair to point out that neither Durvasula & Kahng (2015, 2016), nor Durvasula et al. (2018) deal with RVA, nor with the phenomena discussed by Loporcaro (2015). We are thus not specifically criticising their analysis, but we want to stress that any theory with only two levels runs into problems when dealing with data that show a difference between phonetic and phonological processes.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

Furthermore, in a two-level model, perception and comprehension (i.e. lexical access) all have to be accounted for in one step from phonetic to underlying form, which leads to the problem that the structural restrictions triggering phonological processes would necessarily have to tackle different types of representations in production and comprehension. As shown in Figure 7 with STRUCT, these restrictions would hold on the surface/phonetic form in production but on the underlying form in comprehension. The two-level model would, hence, require two identical but formally independent restrictions, whereas in a three-level model, such as BiPhon, one and the same phonological restriction applies to the surface phonological form in both processing directions.

Finally, a two-level model predicts that the results of production and perception experiments should perfectly align, as the relevant constraints driving the mapping between the two levels would be the same in both directions. However, as shown above, such misalignments can be observed (cf. Boersma & Hamann 2009: 12–33; Daland et al. 2019: 826–827 for illustrations from loanword adaptation). In the three-level model that we are employing, the production and perception do not have to align. The relevant structural constraint for our voicing assimilation – AGREE – is a constraint on the phonological surface form, and, therefore, interacts with different constraints in the two processing directions: with FAITH constraints in production and CUE constraints in perception. If some individuals put more weight on a specific CUE constraint in perception, this will not influence their phonological production, where the CUE constraint does not play a role.

Daland et al. (2019) differ from the Bayesian reverse inference approach employed by Durvasula and colleagues by explicitly distinguishing three levels of representation. Daland et al. (2019: 858) state that whereas their analysis of illusory vowel perception in Korean speech is 'essentially the same as [the BiPhon analysis] offered in Boersma & Hamann (2009)', it goes beyond the latter in two respects. The first is that their analysis is 'probabilistic, and is therefore well-suited to handle the variability that is ubiquitous in perceptual experiments' (p. 859), though Daland et al. themselves note that variability can be straightforwardly dealt with by stochastic OT (as shown here in Section 3.3), and that this difference is thus 'not theoretically crucial' (p. 858). The second aspect in which their model is deemed better is that it 'explicitly links the output of a probabilistic model with behaviour in both discrimination and identification experiments' (p. 859). This is allowed by the so-called linking assumptions:

(19) Linking assumptions by Daland et al. (2019: 858)
   i. Discrimination of two acoustic sequences will be poor when there is a unique phonotactically licit parse which provides a good acoustic match to both sequences.
   ii. Discrimination will be good otherwise.
   iii. Identification corresponds to the highest likelihood parse, regardless of whether it includes an excellent acoustic match.

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

The first two assumptions in Example (19) refer to results of discrimination tasks, the third to results of identification tasks. In the latter, 'highest likelihood parse' refers to the parse that is phonotactically best, which does not need to be an excellent acoustic match. In our account, this corresponds to the winning output candidate, which violates the fewest high-ranked STRUCTURAL, as well as several (lower ranked) CUE constraints. Daland et al.'s account is thus consistent with ours. However, while both proposals discuss the conditioning role of acoustic cues and their interaction with phonotactic constraints (Daland et al. refer, e.g. to burst release, frication noise and the associated [+noisy] feature and to phonotactic constraints), Daland et al. do not provide an actual Bayesian implementation of the interaction of these factors (they list which factors should be integrated in the Bayesian theorem to account for the behaviour of an idealised listener but do not include real values; cf. Daland et al. 2019: 857). We maintain that our model improves on this, as it explicitly formalises the most relevant cues (Esposito 2002), the weighting between them and their interaction with phonotactic constraints as OT constraints, namely, as well-defined theoretical devices that interact with each other in a predictable way in a three-level architecture that, crucially, accounts for perception as well as for production and fits the collected data. Thus, while we see how a Bayesian approach can be thought of being extensionally similar to ours (especially given our stochastic implementation), we are skeptical about the fact that the former could replicate the bidirectionality of our model (though we are by no means claiming that this is impossible).

Furthermore, note that, when discussing the positioning of their experimental findings within a general theory of speech perception, Daland et al. (2019: 857) claim to 'adopt the proposal of Durvasula & Kahng (2015) that the "parse" the listener wishes to recover consists of a lexical representation (i.e. a UR)'. As discussed in our paper, we maintain that, in speech perception, the listener first recovers an SR, and maps this surface representation (SR) to a underlying representation (UR) (these two steps can be performed simultaneously but essentially with an intermediate SR). This is an important aspect, because the constraints involved in these two steps are not identical: in the first step, CUE constraints interact with STRUCTURAL constraints, whereas in the second step, STRUCTURAL constraints interact with FAITHFULNESS constraints. This suggests that, when the speaker recovers the relevant UR, the CUE constraints play no role and predicts that one and the same grammar can produce production-perception mismatches (as we observed in our participants, cf. the end of Section 2, and modelled in our formalisation, cf. Section 3.3).

Finally, going back to the linking assumptions of Daland et al., note that the notion of acoustic match hinges on the conception that listeners of a language can easily judge whether something is a poor match or a good match. This is not the case. In contrast to phoneme identification, which listeners continuously do, judging the similarity between speech sounds is a metalinguistic task that listeners are not used to performing. Daland et al. correctly state that BiPhon does not allow to link the behaviour in discrimination and identification experiments. We consider

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

this, however, to be a strength rather than a weakness, since BiPhon is designed to model speakers and listeners of languages, not participants of metalinguistic tasks.

5. Discussion and Conclusion

The present study showed that RVA in Emilian varieties is a synchronically active process, and that it influences native speakers' perception of voiceless stops in assimilation context: participants reported to have heard a /b/ significantly more often in stimuli with a medial [p] before a voiced obstruent than in stimuli with [p] before a vowel.

The study further showed that RVA is adequately accounted for by a grammar model that takes into consideration both the production and comprehension processes and explicitly distinguish between phonetic and phonological representations, such as BiPhon. Furthermore, it was shown that reverse inference accounts of phonological influence on perception run into problems due to their conflation of surface phonological and phonetic representations, and their failure in explicitly accounting for the influence on the posterior probabilities of a given parsing of: (i) context-sensitive rules and (ii) linking assumptions. As we have shown in this paper, BiPhon remedies the shortcomings of alternative models and allows for a formalisation of the observed production and perception processes in one and the same system.

To further corroborate and refine our findings and the proposed modelling of RVA, a further set of experiments should be carried out both involving the same varieties/languages (ideally, the same speakers) and other varieties/languages.

For instance, as suggested by an anonymous reviewer, an experiment using the same set of stimuli could be carried out with participants speaking a language without RVA. Such a control group would allow us to tease apart the roles of acoustics and phonological knowledge. Along similar lines, a comparison of our results with those obtained by similar experiments carried out on different languages (e.g. Myers 2010 on English) might also be useful. However illuminating these comparative studies might be, though, we maintain that the results should not be necessarily taken at face value, as the phonetic differences between languages could make the comparison quite cumbersome (a more reliable scenario would possibly involve comparing languages that have very similar phonetic implementations but showing different RVA patterns, e.g. Warsaw and Cracow Polish; cf. Gussmann 1992; Rubach 1996; Cyran 2011; Raimy 2021).

Furthermore, a set of follow-up experiments with the varieties we deal with in this paper could be carried out to control for other, possibly relevant variables. For instance, a perception experiment with the same participants and the same set of stimuli but 'p' and 'not p' as answer categories would allow us to control for and exclude a possible bias introduced by the answer categories we employed ('b' and 'not b'), whereas a set of experiments tackling RVA of /b/ in devoicing context (bT words) and both sets of answer options could help us in further refining our representational assumptions and our RVA modelling. More specifically, it would

https://doi.org/10.1017/S0022226722000457 Published online by Cambridge University Press

help us to better understand the relation between the value of the laryngeal specification ([± voice]) and its phonological activity, which could affect the modelling of RVA we proposed. We leave these experiments for future research.

Acknowledgments

We would like to thank three anonymous reviewers of Journal of Linguistics (JoL) for helping us to improve our paper; Paul Boersma for providing us with feedback on statistics; the audience of OCP16 – especially, Adam Albright, Sharon Paperkamp and Alan Prince – for giving us the possibility to discuss and refine a preliminary version of this work; Daniele Vitali for helping us with the dialectological side of the work and the speakers who took part in our experiments.

References

Baiolini, Romano & Floriana Guidetti. 2005. *Saggio di grammatica comparata del dialetto ferrarese*. Ferrara: Cartografica.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104, 591–630.

Bertoni, Giulio. 1905. *Il dialetto di Modena.* Torino: Loescher.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21, 43–58.

Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117, 1989–2054.

Boersma, Paul. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In Anton Benz & Jason Mattausch (eds.), *Bidirectional Optimality Theory*, 33–72. Amsterdam: Benjamins.

Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Andrea Calabrese & Willem Leo Wetzels (eds.), *Loan Phonology*, 11–58. Amsterdam: Benjamins.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45–86.

Boersma, Paul & David Weenink. 2017. *Praat: Doing Phonetics by Computer*. Version 6.0.25, retrieved 12 February 2017 from http://www.praat.org/.

Cavirani, Edoardo. 2015. *Modeling phonologization. Vowel reduction and epenthesis in Lunigiana dialects*. Utrecht: LOT publishing.

Churma, Daniel. 1981. Some further problems for upside-down phonology. *Working Papers in Linguistics*, Ohio State University 25, 67–106.

Coetzee, Andries. 2006. Variation as accessing 'non-optimal' candidates. *Phonology* 23, 337–385.

Coetzee, Andries. 2016. A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology* 33, 211–246.

Crosswhite, Katherine. 2001. *Vowel reduction in Optimality Theory*. New York & London: Routledge.

Cyran, Eugeniusz. 2011. Laryngeal realism and laryngeal relativism: Two voicing systems in Polish? *Studies in Polish Linguistics* 6, 45–80.

Daland, Robert, Mira Oh & Lisa Davidson. 2019. On the relationship between speech perception and loanword adaptation. *Natural Language and Linguistic Theory* 37, 825–868.

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux & Ariel Gout. 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience* 12, 635–647.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Dixit, R. Prakash. 1987. Mechanisms for voicing and aspiration: Hindi and other languages compared. *UCLA Working Papers in Phonetics* 67, 49–102.

Dupoux, Emmanuel, Kazuhiko Kaheki, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568–1578.
Dupoux, Emmanuel, Erika Parlato, Sonia Frota, Yuki Hirose & Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64, 199–210.
Durvasula, Karthik, Ho-Hsin Huang, Sayako Uehara, Qian Luo & Yen-Hwei Lin. 2018. Phonology modulates the illusory vowels in perceptual illusions: Evidence from Mandarin and English. *Laboratory Phonology* 9, paper 7.
Durvasula, Karthik & Jimin Kahng. 2015. Illusory vowels in perceptual epenthesis: the role of phonological alternations. *Phonology* 32, 385–416.
Durvasula, Karthik & Jimin Kahng. 2016. The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of Phonetics* 54, 15–34.
Dvořák, Vera. 2010. Voicing assimilation in Czech. *Rutgers Working Papers in Linguistics* 3, 115–144.
Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59.4, 197–231.
Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6, 110–125.
Gaskell, M. Gareth & William D. Marslen-Wilson. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 144–158.
Gaudenzi, Augusto. 1889. *I suoni, le forme e le parole dell'odierno dialetto della città di Bologna*. Bologna: Forni.
Gouskova, Maria. 2003. *Deriving economy: Syncope in Optimality Theory*. Ph.D. dissertation. Graduate Linguistics Student Association, University of Massachusetts, Amherst.
Gussmann, Edmund. 1992. Resyllabification and delinking: the case of Polish voicing. *Linguistic Inquiry* 23, 29–56.
Grosjean François. 2001. The bilingual's language modes. In Janet L. Nicol (ed.), *One mind, two languages: Bilingual language processing*, 1–22. Oxford: Blackwell.
Henderson, Jeannette & Bruno Repp. 1982. Is a stop consonant released when followed by another stop consonant? *Phonetica* 39.2–3, 71–82.
Hsu, Chai-Shune K. 1998. Voicing underspecification in Taiwanese word-final consonants. *UCLA Working Papers in Phonetics* 96, 90–105.
Huszthy, Bálint. 2016. Italian as a voice language without voice assimilation. *Proceedings of ConSOLE* XXIV, 428–452.
Inkelas, Sharon & C. Orhan Orgun. 1995. Level ordering and economy in the lexical phonology of Turkish. *Language* 71.4, 763–793.
Iosad, Pavel. 2012. Vowel reduction in Russian: No phonetics in phonology. *Journal of Linguistics* 48.3, 521–571.
John J. McCarthy & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory. University of Massachusetts Occasional Papers 18*. Amherst, Mass.: Graduate Linguistic Student Association. pp. 249–384. [Rutgers Optimality Archive 60, http://roa.rutgers.edu]
Kabak, Bariş & William J. Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50, 23–52.
Kilpatrick, Alexander, Shigeto Kawahara, Rikke Bundgaard-Nielsen, Brett Baker & Janet Fletcher. 2021. Japanese perceptual epenthesis is modulated by transitional probability. *Language and Speech* 64: 203–223.
Kim, Yuni. 2002. Phonological features: Privative or equipollent? B.A. thesis, Harvard University.
Krämer, Martin. 2000. Voicing alternations and underlying representations: The case of Breton. *Lingua* 110, 639–663.
Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.
Leben, William R. & Orrin W. Robinson. 1977. 'Upside-Down' Phonology. *Language*, 53 (1), pp. 1–20.
Liberman, Alvin, Katherine Safford Harris, Howard Hoffman & Belver Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.
Lombardi, Linda. 1995a. Laryngeal neutralization and syllable wellformedness. *Natural Language and Linguistic Theory* 13, 39–74.
Lombardi, Linda. 1995b. Laryngeal features and privativity. *The Linguistic Review* 12, 35–59.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17, 267–302.
Loporcaro, Michele. 1998. Syllable structure and sonority sequencing: Evidence from Emilian. In Armin Schwegler, Bernard Tranel & Myriam Uribe-Etxebarria (eds.), *Romance linguistics: Theoretical perspectives. Selected Papers from the 27th Linguistics Symposium on Romance Languages*, 155–170. Amsterdam: John Benjamins.
Loporcaro, Michele. 2011. Phonological processes. In Martin Maiden, John Charles Smith & Adam Ledgeway (eds.), *The Cambridge history of the Romance languages*, 109–154. Cambridge: Cambridge University Press.
Loporcaro, Michele. 2013. *Profilo linguistico dei dialetti italiani*. Bari: Laterza.
Loporcaro, Michele. 2015. *Vowel length from Latin to Romance*. Oxford: OUP.
Manzini, Maria Rita & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
Maiden, Martin & Mair Parry. 1997. *The dialects of italy*. London: Routledge.
McCarthy, John J. 2008. The serial interaction of stress and syncope. *Natural Language and Linguistic Theory* 26, 499–546.
McClelland, James & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18, 1–86.
McQueen, James M. & Anne Cutler. 1997. Cognitive processes in speech perception. In William J. Hardcastle & John Laver (eds.), *The handbook of phoneticsSciences*, 566–585. Oxford: Blackwell.
Mester, Armin & Jaye Padgett. 1994. Directional syllabification in Generalized Alignment. *ROA* 1. https://doi.org/doi:10.7282/T3HX1B11; accessed on 12/11/2022.
Miatto, Veronica, Silke Hamann & Paul Boersma. 2019. Self-reported L2 input predicts phonetic variation in the adaptation of English final consonants into Italian. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 949–953. Canberra: Australasian Speech Science and Technology Association Inc.
Mitterer, Holger & Leo Blomert. 2003. Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics* 65, 956–969.
Monahan, Philip J., Eri Takahashi, Chizuru Nakao & William J. Idsardi. 2009. Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. In Shoichi Iwasakai, Hajime Hoji, Patricia M. Clancy & Sung-Ock Sohn (eds.), *Japanese/Korean linguistics* 17, 391–405. Stanford: CSLI.
Myers, Scott. 2010. Regressive voicing assimilation: Production and perception studies. *Journal of the International Phonetic Association* 40, 163–179. doi:https://doi.org/10.1017/S0025100309990284
van Oostendorp, Marc. 1997. Style levels in conflict resolution. In Frans Hinskens, Roeland van Hout & Leo Wetzels (eds.), *Variation, change and phonological theory*, 207–229. Amsterdam: John Benjamins.
Pape, Daniel & Luis M. T. Jesus. 2015. Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech* 58, 224–246.
Passino, Diana. 2013. A unified account of consonant gemination in external sandhi in Italian: Raddoppiamento Sintattico and related phenomena. *The Linguistic Review* 30, 313–346.
Phatak, Sandeep A., Andrew Lovitt & Jont B. Allen. 2008. Consonant confusions in white noise. *The Journal of the Acoustical Society of America* 124: 1220–1233.
Polivanov, Evgenij Dmitrievič. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, 79–96.
Prince, Alan & Paul Smolensky. 1993[2004]. *Optimality Theory: Constraint interaction in generative grammar*. London: Blackwell.
Raimy, Eric. 2021. Privativity and ternary phonological behavior. In Sabrina Bendjaballah, Ali Tifrit & Laurence Voeltzel (eds.), *Perspectives on Element Theory*, 65–110. Berlin/Boston: Mouton De Gruyter.
Repp, Bruno H. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22, 173–189.
Rohlfs, Gerhard. 1966. *Grammatica storica della lingua italiana e dei suoi dialetti. Fonetica*. Einaudi: Torino.
Rubach, Jerzy. 1996. Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry* 27, 69–110.
Rubach, Jerzy. 1997. Polish voice assimilation in Optimality Theory. *Italian Journal of Linguistics* 9, 291–342.

Rubach, Jerzy. 2008. Prevocalic faithfulness. *Phonology* 25, 433–468.
Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.
Strouse, Anne, Daniel H. Ashmead, Ralph N. Ohde & D. Wesley Grantham. 1998. Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America* 104, 2385–2399.
Swadesh, Morris. 1934. The phonemic principle. *Language* 10, 117–129.
Vagges, Kyriaki, Franco E. Ferrero, Emanuela Magno-Caldognetto & Cristina Lavagnoli. 1978. Some acoustic characteristics of Italian consonants. *Italian Journal of Linguistics* 3, 69–85.
van Dommelen, Wim A. 1983. Parameter interaction in the perception of French plosives. *Phonetica* 40, 32–62.
Vitali, Daniele. 2020. *Dialetti emiliani e dialetti toscani. Le interazioni linguistiche fra Emilia-Romagna e Toscana e con Liguria, Lunigiana e Umbria*. Bologna: Pendragon.
Vitali, Daniele & Davide Pioggia. 2014. *Dialetti Romagnoli. Pronuncia, ortografia, origine storica, cenni di morfosintassi e lessico. Confronti coi dialetti circostanti*. Verucchio: Pazzini.
Wetzels, Willem Leo & Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77, 207–244.
Whang, James. 2021. Multiple sources of surprisal affect illusory vowel epenthesis. *Frontiers in Psychology* 12, 1664–1078. doi: 10.3389/fpsyg.2021.677571.
Wright, Richard. 2004. A review of perceptual cues and robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically-based phonology*, 34–57. Cambridge: Cambridge University Press.
Yazawa, Kakeru, James Whang, Mariko Kondo & Paola Escudero. 2020. Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research* 36, 557–581.
Zimmerer, Frank & Henning Reetz. 2014. Do listeners recover "deleted" final /t/ in German? *Frontiers in Psychology* 5, 735. http://doi.org/10.3389/fpsyg.2014.00735

Appendix A. Duration of voice bar (ms)

Participant	/b/ [b]	Amount of tokens	/pD/ [bD]	Amount of tokens
P1	85	2	94	4
P2	76	2	66	5
P3	84	2	98	5
P4	74	2	100	3
P5	75	2	124	1
P6	83	3	94	1
P7	106	2	108	3
P8	98	3	102	1
P9	71	3	94	1
P10	57	2	93	2
P11	62	3	71	1

Authors' addresses: *(Cavirani)*
KU Leuven, Belgium
edoardo.cavirani@kuleuven.be

(Hamann)
University of Amsterdam, the Netherlands
S.R.Hamann@uva.nl