languages

MDPI

Article

Prosodic and Segmental Aspects of Pronunciation Training and Their Effects on L2

Silvia Dahmen [1,*], Martine Grice [2] and Simon Roessig [3]

[1] Herder-Institut, Universität Leipzig, 4275 Leipzig, Germany
[2] IfL-Phonetik, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany
[3] Department of Linguistics, Cornell University, Ithaca, NY 14850, USA
* Correspondence: silvia.dahmen@uni-leipzig.de

Abstract: Some studies on training effects of pronunciation instruction have claimed that the training of prosodic features has effects at the segmental level and that the training of segmental features has effects at the prosodic level, with greater effects reported when prosody is the main focus of training. This paper revisits this claim by looking at the effects of pronunciation training on Italian learners of German. In a pre-post-test design, we investigate acoustic changes after training in learners' productions of two features regarded as prosodic and two features regarded as segmental. The prosodic features were the pitch excursion of final rises in yes–no questions and the reduction in schwa epenthesis in word-final closed syllables. The segmental features were final devoicing and voice onset time (VOT) in plosives. We discuss the results for three groups (with segmental training, with prosody training, and with no pronunciation training). Our results indicate that there are positive effects of prosody-oriented training on the production of segments, especially when training focuses on syllable structure and prosodic prominence (stress and accent). They also indicate that teaching segmental and prosodic aspects of pronunciation together is beneficial.

Keywords: second-language learning; second-language acquisition; second-language teaching; pronunciation instruction; prosodic training; production; intonation; syllable structure; final devoicing; epenthetic schwa

Citation: Dahmen, Silvia, Martine Grice, and Simon Roessig. 2023. Prosodic and Segmental Aspects of Pronunciation Training and Their Effects on L2. *Languages* 8: 74. https://doi.org/10.3390/languages8010074

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 22 March 2022
Revised: 21 February 2023
Accepted: 21 February 2023
Published: 6 March 2023

1. Introduction

Phonetic-phonological competence of L2 learners is commonly assessed by categories such as (foreign) accentedness, intelligibility and comprehensibility, for example in the Common European Framework of Reference for Languages (CEFR, but see also Derwing and Munro 1997; Thomson 2017). The CEFR states that the goal of pronunciation instruction is not to achieve a native-like pronunciation but rather to speak in a way that does not impair communication (Council of Europe 2020; Chun and Levis 2020). This implies that while a learner's utterance can be heavily influenced by the their first language (foreign accent), it may still be easily understood by native speakers (Derwing and Munro 2015, p. 5), so the more important aspects of pronunciation for successful communication are that the listener can identify what has been said and the message the speaker intends to communicate (intelligibility) without investing excessive effort into the process of understanding (comprehensibility). Studies on native-speaker perception of L2 speech have indicated since the 1980s that prosodic features play an important role in comprehensibility and intelligibility and that teaching prosodic aspects leads to improvements in both prosodic and segmental features of pronunciation, while the converse has not been shown for segmental training (Anderson-Hsieh et al. 1992; Munro and Derwing 1995; Derwing et al. 1998; Gordon and Darcy 2016). Nonetheless, Derwing and Munro argue that the findings of such studies do not imply that only prosodic features should be taught (Derwing and Munro 2015, p. 9), as segmental errors can also lead to misinterpretations of utterances and add to the perception of foreign accent. However, these claims are probably true only for target

languages such as English and German. Recent studies have shown different patterns for the effects of segmental and prosodic influence on the strength of a perceived foreign accent and comprehensibility when the target language is a tone language (Yang et al. 2021) or when the native language of the listeners/raters is different from the target language (Kaunzner 2015, 2018). Yang et al. (2021) examined the effects of prosodic and segmental deviations in L2 utterances in Mandarin Chinese and found that native Chinese listeners' ratings of foreign accent and comprehensibility were influenced by segmental rather than prosodic correctness. Kaunzner (2015, 2018) compared comprehensibility ratings for L2 German utterances of Italian learners for native German, Polish, and Italian listeners/raters and found that only the German listeners rated utterances with prosodic deviations as less comprehensible than utterances with segmental deviations, while the Polish and Italian listeners were instead influenced by segmental deviations. In addition, more-recent findings (e.g., Ulbrich and Mennen 2016; van Maastricht et al. 2021) have indicated that there is a strong interplay between segmental and prosodic features when native listeners rate speech for intelligibility, comprehensibility, and degree of perceived foreign accentedness, where some prosodic features affect native ratings more than others. Research involving English speech manipulated such that native prosody was mixed with non-native segments and vice versa revealed that native listeners' ratings of foreign accentedness depended on both segmental and prosodic deviances and that the impact of prosody depended on the nativeness of the segments: non-native prosody on native segments led to the perception of a weaker foreign accent than on non-native segments, and native prosody on non-native segments led to a stronger perception of foreign accent than on native segments (Ulbrich and Mennen 2016). In a study involving native listener judgements of Spanish learners' L2 Dutch utterances, speech data were manipulated such that a combination of rhythmic or intonational patterns or the speech rate of L1 Dutch speakers was transferred to original learners' utterances. The results showed a stronger influence of intonation on perceived foreign accentedness and comprehensibility when it was the only native feature transferred, while a syllable-timed rhythm (as in Spanish) and a slow speech rate had no such effects (van Maastricht et al. 2021). Thus, the question whether and to what extent it is prosodic or segmental features that mostly affect comprehensibility and perceived foreign accentedness is not as clear as previous research has indicated.

While there is a large number of publications on the general effectiveness of pronunciation instruction (see Saito and Plonsky (2019) for a discussion on intervention studies conducted until 2017), only a few studies have examined the effects of prosodic training on L2 production of segmental features and of segmental training on L2 production of prosodic features. Among these, Missaglia (1999a) found that Italian learners' production of German vowels improved more for a group that received training focused on prosody than for a group that received segmental instruction. While the segmental training consisted of a common set of discrimination and production tasks for German vowels, mixed with articulation exercises, she used the contrastive prosody method (CPM) for her prosodic training. In this method, learners are first made aware of their native language features, such as the rules for sentence-stress or word-stress placement, and of the phonetic features used to mark prominence. This awareness enables them to detect the differences between their L1 production and that of native speakers of the target language and to adapt their production accordingly. The basic assumption behind the method is that in order to know how to produce L2 features, learners need to know explicitly what the corresponding features are in their L1 and what they have to change to correctly produce an utterance in the target language. Learners are treated as bilinguals who are able to make use of their L1 competence in order to improve their L2 productions (Missaglia 1999b, 2007). Common tasks within the CPM are comparing utterances of native speakers to the same utterances produced by L2 speakers and describing the differences, or deliberately producing utterances in the target language with prosodic features of L2 speakers and then changing those features to approximate L1 production. Missaglia's CPM training included stress placement and intonation, including how to produce deaccentuation. Since the CPM

training also included the effects of deaccentuation on the phonetic realisation of vowels, it is unsurprising that vowel production improved for the group receiving this training. The distinction between prosody and segments is difficult to uphold here in that both stress and accentuation have cues that are linked to the production of segments.

Li et al. (2022) examined training effects of embodied prosodic training (involving hand gestures) on the pronunciation skills of Catalan learners of French. They found that embodied prosodic training has positive effects not only on perceived foreign accentedness ratings but also on F2 values of front rounded vowels.

In a larger-scale study on Italian learners of German, Dahmen (2013) compared the results of segmental training (including vowel length, VOT for plosives, and final obstruent devoicing) to those of prosodic training (including intonational focus marking, rhythmic syllable reduction, and syllable structure) for two training groups and a control group of L2 German learners from Northern Italy. Both trainings were based on a method described by Dieling and Hirschfeld (2000), which includes perception and production tasks. For the perception, learners are usually first introduced to a phonetic or phonological feature by listening to utterances that focus on the respective feature. An introductory task for the length contrast in German vowels, for example, could be listening to a story about animals at the zoo, where the teacher first names only those animals whose names contain stressed long vowels and then animals whose names contain stressed short vowels. The learners would not be expected to know all the words, but they should be able to say that there are differences in the vowels between the two sets of names. Further listening tasks include discrimination of contrasting sounds, stress patterns or intonation contours, using minimal pairs and identification exercises in which the learners are presented with speech stimuli and have to signal which of the stimuli contain a certain sound, stress pattern, or intonation contour. Other identification tasks involve detecting rules such as final obstruent devoicing.

For the production part, simple listen-and-repeat exercises are combined with articulation exercises and with tasks involving hand gestures or other visual support. Further production exercises progress from simple repetition to free production. The comparison of all three groups in the study showed that both training groups improved on both the segmental and the prosodic levels but that the group receiving the prosody training improved in more aspects than the group with segmental training. Training effects were assessed for VOT in alveolar plosives /t/ and /d/, final obstruent devoicing, and the quantity and quality of German long versus short vowels for the group that received training labelled as 'segmental', and for rhythmic reduction in unstressed syllables, syllable structure (the realisation of word-final codas and avoidance of epenthetic vowels), and prosodic marking of corrective focus for the group that received training labelled as 'prosodic'. During the study, other aspects of L2 German were also trained, namely the intonation of yes–no questions and answers, as well as stress and accent (word and sentence stress) for the so-called prosody group and the pronunciation of German r-sounds as well as /h/ versus the glottal stop in syllable onsets for the so-called segment group. The training effects in these areas were not assessed.

In this paper, we revisit some of the data collected during the training project that was the basis for Dahmen (2013), using state-of-the-art statistical analyses and making the results more accessible by presenting them in English. We also revisit the terms 'prosodic' and 'segmental', since many features are traditionally assigned to one of the two categories, although they have effects on both. We report in detail on two features that were assigned to the prosodic level and two that were assigned to the segmental level in (Dahmen 2013). The two 'prosodic' categories are the intonation of yes–no questions[1] (cf Section 3) and syllable structure, more specifically the production of epenthetic schwa after word-final consonants (cf Section 4). The 'segmental' categories are final obstruent devoicing (cf Section 5) and VOT in fortis plosives (cf Section 6).

These four features all contribute considerably to the intelligibility, and ultimately to the comprehensibility, of L2 speech. Intonation is crucial for signalling sentence modality because, even in German, questions can often be fragments that are not necessarily syn-

tactically marked as interrogative. The production of epenthetic schwa can lead to the perception of an extra syllable, which in turn can be interpreted as a suffix (such as the plural form in nouns), thus leading to problems at the grammatical level. Although the absence of final obstruent devoicing does not in itself create lexical confusions, the voiced consonant may be followed by epenthesis, leading to the same problem, that of being interpreted as an extra syllable. VOT, especially a lack of aspiration, can lead to lexical confusions, especially if these are in stressed syllables, where the aspiration in German is enhanced. Although language is highly redundant and minimal pairs can often be distinguished by virtue of the context in which they occur, intelligibility and comprehensibility are improved if the listener does not have to deal with conflicting information from the context and the pronunciation. These considerations were the motivation for investigating the effects of training on these four aspects of pronunciation.

These four features also provide clear evidence of the difficulty in upholding the prosodic–segmental dichotomy. For example, even in an aspect of pronunciation that could be regarded as clearly prosodic, i.e., the intonation of yes–no questions, a rise or complex pitch movement can lead to schwa epenthesis or the lengthening of a vowel, both of which are usually treated as segmental (Grice et al. 2015, see discussion in Section 4 below). This is referred to as tune–text interaction, indicating that the intonation and the segmental structure cannot be treated separately. A clearer case in our investigated features is the pronunciation of word-final consonants. This is not only segmental but also prosodic. This is because obstruent devoicing is related to syllable structure: an error in syllable structure, e.g., the epenthesis of schwa in *Rad* 'bike' [rad.də], leads to a possible resyllabification, in addition to other adjustments, such as the lengthening of the plosive (transcribed as a geminate) and possibly the shortening of the vowel. This resyllabification runs the risk of removing the (syllable final) context for the devoicing of <d> to apply. Voice onset time is not purely segmental either: it depends on the temporal coordination of laryngeal and supralaryngeal gestures, and it interacts with syllable prominence, such that the strength of plosive aspiration depends on whether the syllable is lexically stressed or accented (e.g., Lisker and Abramson 1967; Jessen and Ringen 2002; Savino et al. 2015; Lein et al. 2016).

Given these interactions, our research question is concerned with how far each of these features of L2 speech can improve with targeted explicit training. Specifically: (1) How successful is training in intonation and syllable structure (suppressing epenthesis) and does it affect the production of individual consonants? and (2) How successful is training in final devoicing and VOT of voiceless plosives and does this training affect the production of syllable structure and intonation?

2. Materials and Methods

2.1. Subjects and Recordings

The data were recorded during a training project in Germany, one day *before* and one day *after* each training phase. The recordings were conducted in a quiet room using a mobile DAT recorder and head-mounted microphones. The trainings took place in Bischofswerda (Saxonia) as part of a training camp for students from all over Italy who were preparing to take part in the German language diploma (*Deutsches Sprachdiplom der Kultusministerkonferenz*) for the level B2/C1 of the Common European Framework of Reference. The training camp consisted of two phases of 10 days each, in which different groups of students took part in the courses. In the following, we give details on the speakers in the groups.

In the first phase of the training project, students attended courses on reading and listening comprehension as well as on oral and written communication. During the first phase, 8 students (3 male, 5 female) from one school class in Turin were recorded. They were 17 or 18 years old at the time of the recordings and had learned German for 3.5 to 7 years. They reported no German relatives or friends and thus used German only in the classroom. They did not receive any pronunciation training during the duration of the project. Therefore, this group is the *control group* in the present study.

In the second phase of the training project, the reading and listening comprehension group was split in two subgroups, which took turns attending reading/listening comprehension and pronunciation training. Students recorded were from Montagnana and Turin. The groups undertook training in what was referred to as either segmental or prosodic aspects of pronunciation. The groups are heretofore referred to as the *segment group* and the *prosody group*, respectively. The segment group consisted of 13 subjects altogether, 7 from Turin (2 male, 5 female) and 6 from Montagnana (1 male, 5 female). The prosody group consisted of 12 subjects, 6 from Turin (all female) and 6 from Montagnana (1 male, 5 female). All subjects in the test groups were between 17 and 19 years old, had learned German for 4 to 5 years, and used German only in the classroom at the time of the recordings. More information about the training is given in the next section.

The students were randomly assigned to the training groups. The metadata of the students do not indicate any systematic differences in pronunciation competence between groups. Differences between the groups before training are most likely due to individual factors not controlled for in this study. The analysis presented here concentrates on differences between the time point before and the time point after training rather than absolute differences between groups.

2.2. Speech Materials

The speech materials presented in this article consist of read sentences as well as semi-spontaneous utterances. The semi-spontaneous utterances were yes–no questions (cf Section 3) elicited in specially designed card games. We first give an overview of the read sentences and explain the card games below. The following sentences were used in the study:

(1) *Dina gab Elmar ein neues Rad.* ('Dina gave a new bike to Elmar')
(2) *In der gelben Hütte lebte ein großer Hund.* ('In the yellow hut lived a big dog')
(3) *Tina gab Hanna einen guten Rat.* ('Tina gave good advice to Hanna')
(4) *Die billigen Hüte waren ganz schön bunt.* ('The cheap hats were pretty colourful')
(5) *Helga spielte einmal Tennis.* ('Helga once played tennis')

The sentences were presented to the students in random order to reduce the chance of their identifying the minimal pairs. For the occurrence of word-final epenthetic vowels (cf Section 4), we examined the target words *Rad*, *Hund*, *Rat*, and *bunt* (sentences 1 to 4). *Rad* and *Rat* (sentences 1 and 3) were the target words for measuring final obstruent devoicing (cf Section 5). For VOT (cf Section 6), we looked at *Tina* and *Tennis* (sentences 1 and 5).

The card games were played in pairs. The cards in this game depicted day-to-day objects in different colours. The participants had the task of collecting cards with the same colour or the same object by exchanging cards with their fellow player. To initiate the exchange, participants formulated a yes–no question, e.g., *hast du einen gelben Teller?* (English: 'do you have a yellow plate?'). This question was followed by the answer, and if desired, the card was exchanged.

The materials used in the analyses of the different phenomena will be described in the respective subsections to make them more accessible to the reader for the interpretation of the results.

2.3. Training

During the training phases, the control group received 90 min of reading and listening comprehension training per day. This course was taught by the same teacher as the pronunciation training classes to rule out a teacher effect. The test groups received 45 min of pronunciation training per day. Each pronunciation training session contained perception and production exercises for the respective segmental or prosodic areas, usually with one or two new phenomena introduced in each session and then repeated in the following sessions. For instance, the segment group engaged in discrimination and production exercises for long versus short vowels and for aspirated versus unaspirated plosives in the first session, and then in the second session, they engaged in production exercises for both and for a

first introduction to final obstruent devoicing. The prosody group received training on sentence intonation, nuclear accent placement (sentence stress) and focus marking, word stress, rhythm (reduction in unstressed syllables), and syllable structure. The segment group received training in aspirated plosives, final obstruent devoicing, the long-short and tense/lax distinction in German vowels, consonantal and vocalised realisations of <r>, the fricative allophones [ç] and [x] of orthographic <ch>, word-initial /h/ versus glottal stop, and front rounded vowels. The students were asked not to exchange pronunciation exercises between the groups, and their teachers reported when they did. For that reason, two subjects that had originally been recorded had to be excluded from the study. These speakers were not included in the study (they are thus not part of the speaker sample described in the previous subsection). The training sessions for the areas relevant to the present study are briefly described below.

2.3.1. Intonation of Yes–No Questions (for Results, cf Section 3)

Only the prosody group received training in the intonation of yes–no questions. To make the participants aware of the high final rise in German yes–no questions, the teacher wrote questions such as *ist das ein Tisch? hast du ein Buch? kennst du München?* ('is this a table?', 'do you have a book?', 'do you know Munich?') on a board and drew lines over the sentences to indicate at which point and to which extent the intonation contour rose while the participants listened to the questions and identified the rise in pitch and in the line drawn over the sentence. Next, other questions of the same type were presented in oral and written form, and the students drew their own lines to represent the intonation contours they perceived. The point at which the contour starts rising in German (i.e., the accented syllable) was identified by the group, and a rule was formulated. Again, yes–no questions were used to apply the rule (task: find the syllable where the rise starts). This task was combined with oral production exercises and with hand gestures that imitated the rising pitch contours. The use of hand gestures in combination with oral output has been found to enhance L2 production of both segmental and prosodic features (e.g., Baills et al. (2022); Li et al. (2020)). Other production tasks included dialogues of the form *hast du [Objekt]?* ('do you have [object]?')–*ja/nein* ('yes/no'), where each participant asked others for a matching object on a card, knowing that there were pairs of identical cards. Similar tasks had one participant at a time choose an object from a set of possible objects (e.g., an orange, a banana, a book, a newspaper etc.), the others asking questions such as *kann man es essen? ist es gelb?* ('can you eat it? ', 'is it yellow?') to find out which object the candidate had chosen. Hand gestures were used during production throughout the training phase.

2.3.2. Avoiding Word-Final Epenthetic Vowels (for Results cf Section 4)

The first step in the training of participants of the prosody group was to make them aware that they had produced epenthetic vowels after words ending in consonants, e.g., *Tisch, Stuhl, Blatt* ('table, chair, leaf'). Recordings of participants were played, and all cases of epenthetic schwa were pointed out by the teacher. As word-final schwa is a very common grammatical marker in German (orthographically represented by <-e>), word pairs such as *Tisch–Tische* ('table–tables') were presented as auditive stimuli to make the participants aware that epenthetic schwa can lead to the perception of unintended grammatical forms by German native listeners. In order to avoid word-final schwa epenthesis, participants were asked to produce words ending in fricatives, e.g., *Tisch*, and lengthen the final consonant for as long as they could, in order to prevent the reflex of adding a vowel. Subsequently, the final consonant was shortened (where the teacher indicated via a hand gesture when to stop producing the consonant, thus indicating the duration of the sound) until a normal duration was reached. For word-final plosives, as in *Blatt*, participants were asked to lengthen the aspiration of the plosive, first driving small balls of paper over a table with the force of the aspiration and then shortening it until the appropriate duration was achieved. In following sessions, words with more-complex codas were used for similar tasks, e.g., *eins, einst, Herz, Herbst* ('one, once, heart, fall'). In these tasks, the participants had to 'build up'

the words sound by sound in order to carefully pronounce all consonants in the complex codas. Another productive exercise included the oral production of the above-named word pairs of the type *Tisch–Tische*, with a special focus on the different pronunciations of each member of a word pair.

2.3.3. Final Obstruent Devoicing (for Results cf Section 5)

In order to be made aware of the rule of final obstruent devoicing in German, the segment group was first presented with orthographic stimuli, focusing on the graphemes <b, d, g>. For example, in the sentence *Sabine ist sehr hübsch und lieb* ('Sabine is very pretty and kind'), they were asked to first find all graphemes <b> and then listen to a recording of the sentence and mark all instances of <b> being pronounced as [p]. The same procedure was carried out for other sentences, including words with <b,d,g> in the onset and coda positions. After this identification exercise, the rule for final obstruent devoicing was formulated in written form and then applied to other words, e.g., *Korb* ('basket'), *Land* ('country'), and *Tag* ('day'). In the next step, the graphemes <s> and <v> were treated in the same fashion. As a productive exercise, singular and plural forms of nouns ending in <b, d, g, s, v> were pronounced by the participants, focusing on the change in pronunciation of these graphemes when they change their position within syllables. For instance, in *Tag*, <g> is pronounced [k], but in the plural *Tage*, it is pronounced [g]. For word-final plosives <b, d, g>, participants held a sheet of paper before their mouths and produced aspiration strong enough to move the paper. For word-final fricatives <s, v>, they put a finger on their larynxes to feel whether their vocal folds were vibrating for words such as *Haus* ('house'), where there should be no vibration during the final consonant, versus *Häuser* ('houses'), where there should be.

2.3.4. Voice Onset Time (for Results cf Section 6)

The segment group was first presented with written words present in German and Italian (and English), namely *Pizza* and *Taxi*. Participants were asked to pronounce the words in their Italian form, then the teacher pronounced them in the German way, with aspirated plosives. After thus making the participants aware of the difference in the production of plosives in German and Italian, the next step was a discrimination task with minimal pairs, such as *Pass–Bass* ('passport–bass'), *Tank–Dank* ('tank–thanks'), or *Karten–Garten* ('cards–garden'), where they indicated which of the words of a word pair they had heard. The term 'aspiration' was introduced, and the different use of voicing versus aspiration in Italian and German was explained. The need for the aspiration of fortis plosives in German was explained by the fact that unaspirated [t], for example, can be perceived as [d] by German listeners, which might result in misunderstandings. In order to obtain a strong aspiration, the participants were asked to hold a sheet of paper in front of their mouths and make it move by producing a puff of air after the release of the plosives. This was repeated for a great number of German words with initial [t^h, p^h, k^h]. Additionally, a card game was played during which the participants had to find words with matching initial sounds written on cards. For example, the words *Pass* and *Polizei* ('police') would be a match, but *Pass* and *Bass* would not be. In order to receive the cards of a matching pair, the participants had to pronounce the words loudly, and the other players decided whether aspiration was produced in the correct places.

2.4. Overview of Groups and Training

To provide a better overview of the methodology used in this study, Table 1 lists all speaker groups with their origin and a summary of the training they received.

Table 1. Overview of training groups.

Group	Number of Speakers	City/Region	Training
Control group	8 (3 male, 5 female)	Turin (all)	No pronunciation training
Segment group	13 (3 male, 10 female)	Montagnana (6 speakers) Turin (7 speakers)	Final obstruent devoicing; aspiration of fortis plosives
Prosody group	12 (1 male, 11 female)	Montagnana (6 speakers) Turin (6 speakers)	Intonation of yes–no questions; epenthetic vowels

2.5. Statistical Analyses

The data were statistically modelled with Bayesian mixed models. For tutorial introductions of Bayesian statistics with phonetic data, see Vasishth et al. (2018), Roettger and Franke (2019), and Nalborczyk et al. (2019). Bayesian statistics were carried out because they are known to provide reliable results, even for small samples (van de Schoot et al. 2015). The models were fit with brms 2.16.3 (Bürkner 2018) in R 4.1.2 (R Core Team 2021). The package brms ('Bayesian regression modelling with Stan') implements an interface to Stan to compute Bayesian models via Markov chain Monte Carlo (MCMC) sampling (Carpenter et al. 2017). All models were checked for convergence by ensuring that they did not exhibit *Rhat* values larger than 1.00. The model fit was visually inspected by using predictive posterior check plots. To assess the training effects, we examined the differences between the posterior distributions before and after training by employing the hypothesis function of the brms package. Throughout the analysis, we used tidyverse 1.3.1 for data processing (Wickham et al. 2019). For plotting, we used ggplot2 3.3.5 (Wickham 2016).

3. Training Effects on Magnitude of Question Rises

In this section, we examine the effects of training on the final rise in yes–no questions. Both German and Italian commonly have final rises in such questions. Refer to Appendix A for an overview of the native patterns in the two languages. In this comparison, it becomes evident that final yes–no question rises in Italian are *smaller* in magnitude than those in German.

Moreover, we ask whether the magnitude of the rise produced by Italian learners of German is similar to their L1, i.e., whether learners exhibit smaller rise magnitudes in their L2 because of influences from their L1 before training. We can investigate how this element of their L2 changes through training and whether the three training groups, namely control, segment, and prosody, show different training outcomes with respect to the question rise. The reader is reminded that only the prosody group received explicit training on question intonation (see Section 2.4).

3.1. Data

The data analysed here were elicited with a card game specifically designed for this task. The players ask for cards with specific colour-object combinations (*do you have a blue coffee pot?* German: *hast du eine blaue Kanne?*). Each player has a tableau in front of them depicting specific colour-object combinations in two rows of eight numbered positions. In addition, each player has a stack of cards designating positions 1 to 8. At the beginning of one move, a player draws a position card (e.g., position 3) and looks up the colour-object combination in this position in the upper row of the tableau (e.g., green plate). The player then formulates a question for this specific colour-object combination, e.g., 'in position 3, do you have a green plate?' (German: *in Position 3, hast du einen grünen Teller?*). The other player looks up the position in the lower row of their tableau and produces an answer. The answer can be 'yes' (German: *ja*) or 'no, I have <alternative>', where <alternative> stands for a different colour-object combination, e.g., *no, I have a green ball* (German: *nein, ich habe eine grüne Kugel*). The colour adjectives were *blaue/blauen* 'blue', *gelbe/gelben* 'yellow', *graue*/grauen 'grey' and *grüne/grünen* 'green'. The object nouns were *Kanne* 'coffee pot', *Teller* 'plate', *Gabel* 'fork', and *Kugel* 'ball'.

In total, 317 recordings entered the analysis. Of these recordings, eight were excluded because the questions lacked a final rising movement. As a result, the magnitude of 309 final question rises could be assessed. An example contour of one question is given in Figure 1C. This instance is taken from the recordings before training.

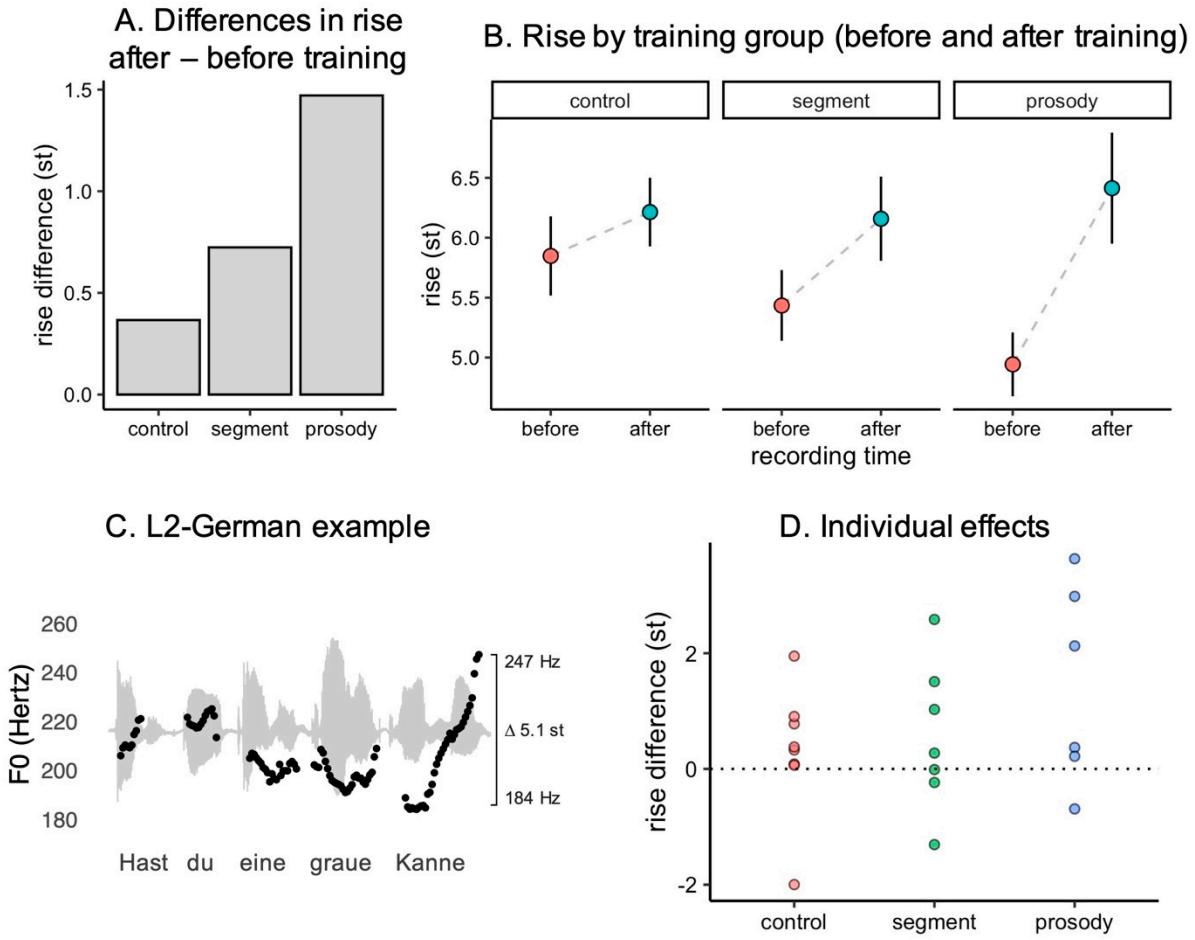

Figure 1. Differences in means before and after training (**A**), means and SE before and after training (**B**), example contour from one L2 speaker (**C**), separate differences in rise for speakers (**D**), where each dot corresponds to one speaker.

3.2. Analysis and Results

Table 2 gives the means and standard deviations of the final rise for the three training groups before and after training. In addition, the last column represents the difference between the mean before the training and the mean after the training.

Table 2. Results for the final rise of the three training groups in semitones (st).

Training	Time Point	Final Rise Mean (st)	Final Rise SD (st)	Difference (st) (Mean After–Mean Before)
Control	Before	5.85	2.17	0.37
	After	6.21	2.29	
Segment	Before	5.43	2.05	0.72
	After	6.16	2.63	
Prosody	Before	4.94	1.87	1.47
	After	6.41	3.25	

The results are illustrated in Figure 1A,B. Panel A shows the differences in means before and after training (mean before minus mean after). First, the differences in all groups are positive. This means that all groups adjust their final question rises to make them larger. The largest change is obtained by the prosody group, the smallest change by the control group. The segment group is situated in between these two poles. Panel B shows the means with standard errors before and after training. The slope of the dashed line illustrates the change within each of the groups between the two recording time points. In addition, it can be observed in this plot that the prosody group is not only the group with the largest improvement after training but also the group that exhibits the lowest values before training.

The statistical model used rise magnitude as the dependent variable. The fixed effects were time of recording (before or after training) and training type (control, segment, prosody), as well as the interaction between time of recording and training type. The model included random intercepts for speakers and by-speaker random slopes for the effect of recording time. In addition, the model used random intercepts for the nouns that the rise is realised on (e.g., *Teller*, *Kanne*, . . .).

We used a normally distributed prior probability distribution (prior) with a mean of 0.0 and a standard deviation of 1.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, we used a Student's t distribution with degrees of freedom of 3.0, a median of the data as a mean of the distribution and a standard deviation of 2.5 (ν = 3.0, μ = median of the variable, σ = 2.5). As priors of the standard deviations of the random intercepts and slopes and as the residual standard deviation of the model, we used a Student's t distribution (ν = 3.0, μ = 0, σ = 2.5). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions (η = 1). MCMC chains were run for 7000 iterations, with 3500 warmup iterations at four chains, resulting in a total of 14,000 posterior samples used for inference.

We are interested in the differences in posterior distributions between the recording time points (before vs. after) in each group to assess the evidence for an improvement in the groups. Therefore, we calculated the posterior distribution of the differences before and after training (after minus before). We report the estimated difference β, the standard error of the estimate (SE), the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is position $Pr(\beta > 0)$. The parameter β indicates how large the model estimates the difference in rise magnitude between the two recording time points. $Pr(\beta > 0)$ indicates how certain we can be that the difference between before and after training is indeed positive, i.e., that the rise indeed became larger during training. Table 3 presents the results of the statistical model. The table shows that the estimate of the differences is largest in the prosody group (1.20). The 90% CI does not include zero, and $Pr(\beta > 0)$ is 0.99. Given the model and the data, we can conclude that this constitutes strong evidence for an increase in the rise magnitude from before training to after training. The other two groups also yield positive estimated differences, where the estimate for the segment group is larger. However, for both training groups, the 90% CI includes zero, and $Pr(\beta > 0)$ is only 0.84. Hence, given the model and the data, the evidence for a positive difference (or an increase in rise magnitude during training) is much weaker.

Table 3. Results of the Bayesian mixed model regarding the difference in rise magnitude between recording time points (after training minus before training) in the three groups.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	Pr(β > 0)
Control	0.41	0.41	−0.26	1.07	0.84
Segment	0.47	0.49	−0.33	1.27	0.84
Prosody	1.20	0.52	0.35	2.04	0.99

An interesting question in the context of the training effects on the final rise magnitude is whether all subjects behave in a uniform way. Figure 1D gives insights into the development of the individual language learners in the groups. In this plot, each dot corresponds to one subject. The *y*-axis shows the differences between the recording time points before and after training (after minus before), just like Figure 1A for the whole group. It can be observed that there is indeed a considerable amount of variation among the individuals. While most subjects show a positive difference, i.e., a larger rise after training, a minority of subjects exhibit the reverse pattern or a difference close to zero. This is particularly true for the segment training group. In addition, we can see that the training groups overlap to a certain extent: not all individuals in the prosody group yield larger rise differences than all individuals in the segment or control group. However, in the prosody group, there are some speakers who yield much larger differences, and the only speaker who reverses the pattern is close to zero.

In addition, there are differences between the groups before training. In Figure 1B, we observe smaller rise magnitudes for the prosody group and the segment group compared with the control group at the recording time point before training. As outlined in the methods section, however, the metadata of the students do not indicate any systematic differences between the groups. It is also beyond the scope of this paper to assess whether the magnitude of the improvement during the training is causally linked to the base level before training.

3.3. Interim Summary and Discussion

In this section, we analysed the rise magnitude of yes–no question rises, and how it develops in the three training groups under discussion. In all training groups, we see some kind of increase in rise magnitude after training. Our analysis has demonstrated that these differences are largest for the prosody group, and our statistical modelling has provided strong evidence for a positive change in only this group. We have also shown, in addition to the general trend of an increase in the rise magnitude and the group differences, that there is considerable individual variation.

In Appendix A, we compare similar questions produced by German and Italian native speakers in their L1s. These results show that Italian L1 yes–no questions exhibit considerably smaller rise magnitudes than their German counterparts. The learners' results presented in this section seem to range in between the two extremes, with a tendency towards the German realisation pattern after training in the prosody group.

An interesting point to consider is whether the observed rise magnitudes can be explained by a phonetic or phonological transfer effect from the L1 to the L2 (Mennen 2007). At first glance, it may appear to be a clear phonetic effect. Both languages have a rising question intonation that can be described as a combination of low accent L* followed by a high or rising boundary tone. The phonetic implementation of the height of the final tonal target appears to differ across the languages, and Italian learners of German may transfer their phonetic knowledge about the final rise to their L2 German. However, we gain a different perspective from a closer look at the phonological descriptions of intonation contours in both languages. In German, a typical nuclear yes–no question contour is one that is best described as L* H-^H%, with an H intermediate phrase boundary tone and an upstepped ^H% intonation phrase boundary tone (Grice and Baumann 2002). This contour is characterised by the rise towards an extra-high final pitch. Given enough syllables between the L* and the end of the phrase, a plateau occurs. The L* H-^H% contour contrasts with L* L-H% which is said to be used to convey indignation or for answering the phone (Grice and Baumann 2002).

For Italian, as Savino (2012) points out, there is considerable variation in the realisation of question contours in the different varieties of the language, and each variety has multiple intonation patterns in its inventory. In Savino's study, the final rise is not predominant for the Turin region that the speakers of the present study were from, although it was found in around 15% of (information-seeking) polar questions. However, for other Northern Italian

varieties, such as Bergamo and Milano, she identifies a rising contour as predominant and describes it as H+L* L-H% (representing not only the final rise but also a preceding fall, which we are not concerned with here). Although Savino's intonation contours were obtained from task-oriented dialogues, the task (a map task) was different from the card game used in the current study and could have affected the distribution of different contours. What is important here is that both Savino's study and our results (from a considerably smaller sample) show that the final rise is available to the speakers as an option and is part of their intonational repertoire. Consequently, we may hypothesise that these speakers of Italian map their H+L* L-H% onto the German L* L-H% contour. In this light, the outcome observed in this study can be seen as the result of a phonological transfer of the boundary tone sequence, in which L-H% is used instead of the German native H-^H% with the higher final target. This hypothesis needs to be tested in future research. In doing so, it would be interesting to investigate how the final part of the contour is realised over different numbers of syllables in both languages and compare it with the learners' productions.

4. Training Effects on the Reduction in Epenthetic Vowels

A striking characteristic of the Italian pronunciation of words ending in a consonant is the epenthesis of a word-final vowel. As native Italian words usually end in vowels, epenthesis is usually found in loan words such as *tennis* [ˈtɛnːisːə] (Sluyters 1990). However, epenthesis is not present across the board. Inter alia, it appears to depend on factors such as the metrical structure of the word (more often if the final syllable is stressed), the voicing of the final consonant (more often when the final consonant is voiced), and the intonation contour (more often with rises and complex contours) (Grice et al. 2015). Unsurprisingly, epenthesis is also found in the pronunciation of Italian learners when they speak German. This subsection investigates the effects of explicit instruction in the prosody group in syllable structure, concentrating on words with a final consonant. This training aimed at both making the learners aware of their production of epenthetic vowels and reducing them by focusing on producing the word-final consonants without a following vowel. The segment group did not receive any information or instruction on word-final epenthetic vowels, but they did receive training on final obstruent devoicing (cf Section 5). As this training also focuses on the word-final consonant, it may have also had an effect on the production of epenthetic vowels, at least for words ending in consonants that undergo final devoicing in German.

4.1. Data

In order to assess the training effects of both the explicit syllable structure training that the prosody group received and the (implicit) segmental training of final obstruent devoicing, we separately focus on words ending in <t> and those ending in <d> because the word-final <d> is prone to be interpreted and produced as a voiced stop by Italians. Before the training, all groups produced epenthetic vowels in both conditions, but not consistently within groups and not to the same extent between groups. The control group produced the smallest number of epenthetic vowels before and after the training, followed by the segment and prosody groups (cf Table 4). With regard to the two conditions, epenthetic vowels were more often produced in words with a final <d> than those with a final <t> by both training groups, but not in the control group.

The data analysed here were recordings of two words ending in <d>, specifically *Rad* 'bike' and *Hund* 'dog', and two words ending in <t>, specifically *Rat* 'advice' and *bunt* 'colourful'. They were produced in sentences (1) to (4) below with the target words (here underlined) accented and in sentence-final position. This position leads to accentual and phrase-final lengthening, but this effect is constant across conditions. They were read aloud by all 33 subjects (target words are underlined, cf Section 2.1):

(1) *Dina gab Elmar ein neues Rad.*
(2) *In der gelben Hütte lebte ein großer Hund.*
(3) *Tina gab Hanna einen guten Rat.*
(4) *Die billigen Hüte waren ganz schön bunt.*

The sentences were interspersed with fillers during the recordings. Each sentence was produced three times by members of the control group and five times by members of the training groups. In total, 192 word realisations of the control group (8 speakers × 4 words × 3 repetitions × 2 recording times), 480 word realisations of the prosody group (12 speakers × 4 words × 5 repetitions × 2 recording times), and 520 word realisations of the segment group (13 speakers × 4 words × 5 repetitions × 2 recording times) entered the analysis. All word realisations were analysed in Praat to detect the presence of epenthetic vowels, and each occurrence was counted. The percentage of word realisations with epenthetic vowels of the data set was calculated and compared for all groups before and after the training sessions. The percentage values before and after training for both sets of words and all three groups are presented in Table 4, together with the differences between recordings made before and after training.

Table 4. Epenthetic vowel results.

Rad and Hund			
Training	**Time Point**	**% Epenthetic Vowels**	**Difference (%) (% after–% before)**
Control	Before	14.58	−2.08
	After	12.50	
Segment	Before	64.62	−25.38
	After	39.23	
Prosody	Before	84.17	−16.67
	After	67.50	
Rat and Bunt			
Training	**Time Point**	**% Epenthetic Vowels**	**Difference (%) (% after–% before)**
Control	Before	14.58	−10.42
	After	4.17	
Segment	Before	41.54	−7.69
	After	33.85	
Prosody	Before	67.50	−20.00
	After	47.50	

4.2. Analysis and Results

All groups show reductions after training in the percentage of epenthetic vowels for both sets of words, but to different extents when we compare groups and target words. The control group has the lowest values before and after training for both sets of target words. The reduction is larger for words ending in <t>. The segment group exhibits a massive reduction for words ending in <d> and only small improvements for words ending in <t>. The prosody group improves in both sets of words, slightly more for words ending in <t>. The improvements in the reduction in epenthetic vowels are illustrated in Figure 2.

For the statistical analysis, epenthetic vowel (yes/no) entered the model as a binary dependent variable for each set of words: (1) *Rad* and *Hund* and (2) *Rat* and *bunt*. The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well the as the interaction between the two variables. The model included random intercepts for speakers and by-speaker random slopes for the effect of time of recording.

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, a Student's *t* distribution was used with degrees of freedom

of 3.0, a mean of 0.0, and a standard deviation of 2.5 (ν = 3.0, μ = 0.0, σ = 2.5). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student's t distribution (ν = 3.0, μ = 0.0, σ = 2.5). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions (η = 1). The model ran with four MCMC chains for 4000 iterations.

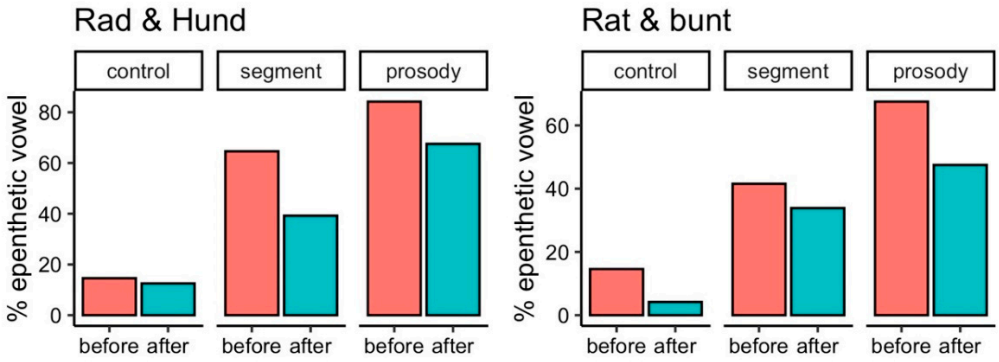

Figure 2. Percentages of epenthetic vowel occurrence for training groups before and after training.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after minus before training) in terms of log odds. We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that epenthetic vowels are reduced after training. $Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is.

The results are presented in Table 5. For the set *Rad* and *Hund*, the results show a robust reduction in epenthetic vowels only in the segment group, with $Pr(\beta < 0) = 0.99$. For the set *Rat* and *bunt*, the results indicate a robust reduction in epenthetic vowels only in the prosody group, with $Pr(\beta < 0) = 0.98$.

Table 5. Statistical results for epenthetic vowels.

Rad and Hund					
Training Group	**β**	**SE**	**90% CI Low Boundary**	**90% CI High Boundary**	**Pr(β < 0)**
Control	−1.13	1.56	−3.80	1.28	0.77
Segment	−2.46	1.05	−4.28	−0.84	0.99
Prosody	−0.77	1.37	−2.87	1.54	0.73
Rat and Bunt					
Training Group	**β**	**SE**	**90% CI Low Boundary**	**90% CI High Boundary**	**Pr(β < 0)**
Control	−2.07	1.44	−4.53	0.20	0.93
Segment	−0.58	0.62	−1.62	0.41	0.84
Prosody	−1.33	0.64	−2.4	−0.31	0.98

4.3. Interim Summary and Discussion

In this section, we analysed the effects of two trainings on the realisation of word-final plosive codas with regard to the occurrence of epenthetic vowels. The results show that the prosody training was effective for both sets of words, but the effects are robust only for words ending in orthographic <t>, not for orthographic <d> (as in *Rad* and *Hund*). The voicing of final consonants plays an important role in the occurrence of final epenthetic vowels in Italian (Grice et al. 2015), which is reflected in our data set. Words ending in

voiced consonants (even if the voicing is the result of a spelling-based pronunciation) exhibit more cases of vowel epenthesis than do words ending in voiceless consonants. Consequently, the syllable structure training in the prosody group can result only in a reduction in epenthetic vowels in words ending in an orthographic voiced consonant when this consonant is interpreted as devoiced by the learners. This means that schwa epenthesis is best combined with an explicit training of final obstruent devoicing. The segment group that received explicit instruction in final obstruent devoicing shows a robust reduction in vowel epenthesis only for those words in which devoicing occurs, but not for others. This means that the segmental training had a positive effect for one set of words, probably due to the focus on the syllable coda and explicit instructions to produce final plosives with aspiration (precluding schwa epenthesis). However, the effects are not transferred to the other words with final consonants if these are voiced, so this does not constitute an improvement in the production of syllables in general. The results for all groups show that both trainings are effective but that they should be combined. We will next look at final devoicing in order to find out whether the syllable structure training of the prosody group had any effects on the production of (orthographically) voiced plosives.

5. Training Effects on Final Obstruent Devoicing

Final obstruent devoicing refers to a phonological phenomenon occurring in syllable codas in German words. Plosives and fricatives that are underlyingly voiced become voiceless in that position, so the word *Rad* 'wheel' is pronounced [ʁaːt], while the plosive is voiced when it is in syllable-initial position as in the plural form *Räder* ['ʁɛːdɐ]. German spelling does not reflect these differences, so learners interpret graphemes that usually represent voiced obstruents as such (Hayes-Harb et al. 2018). In Italian, obstruents usually occur in syllable codas when they are part of a geminate consonant, e.g., *fredda* ['frɛd.da] 'cold', and there is a voicing distinction in that position (e.g., *fretta* ['frɛt.ta] 'hurry'). As a consequence, Italian learners tend to pronounce German *Rad* as [rad.də]. In this section, we investigate the effects of explicit training of final devoicing of plosives as conducted with the segment group (see Section 2.3). The other groups did not receive any information or instruction on final obstruent devoicing, but the prosody group received training focusing on word-final consonantal codas and the avoidance of an epenthetic vowels (cf Section 4). This may have led to more awareness of the syllable coda and even an improvement in final devoicing.

5.1. Data

Final devoicing is a neutralisation process, although many studies claim that this neutralisation is incomplete because German natives produce word pairs such as *Rad–Rat* ('bike'–'advice') slightly differently (e.g., Roettger et al. 2014). However, the training was based on complete neutralisation, so this is what the learners aimed to achieve. The data analysed here are part of the data set described in Section 4. Here, we look at only the word pair *Rad–Rat* that we elicited as described above in the carrier sentences (cf Section 2.1):

(1) Dina gab Elmar ein neues Rad.
(3) Tina gab Hanna einen guten Rat.

In total, 48 repetitions for each target word were elicited from the control group (8 speakers × 3 repetitions × 2 recording times), 130 for the segment group (13 speakers × 5 repetitions × 2 recording times), and 120 for the prosody group (12 speakers × 5 repetitions × 2 recording times). All the word tokens were annotated in Praat, and the values were automatically extracted. The parameters examined here are the duration of the vowel and consonantal closure intervals and the duration of voicing during the closure interval for *Rad*. In order to achieve a neutralisation effect, these parameters should become more similar for the two target words after the training.

5.2. Analysis and Results for Neutralisation of Vowel Duration in L2 German

What we are interested in here is the absolute distance between the vowel /a:/ in *Rat* and the vowel /a:/ in *Rad* in terms of duration. We ask whether the distance becomes smaller, i.e., whether the vowels of *Rat* and *Rad* become more similar after the training. Because we are not dealing with a parameter on the level of one utterance but rather the relation between different productions, we first calculate the mean for each vowel for each speaker. That is, for each speaker, we calculate the mean duration of /a:/ from *Rat* and the mean duration of /a:/ from *Rad*. Next, we calculate the *absolute* distance between these durations. Table 6 presents the absolute distance between the vowels in both target words in milliseconds for each group before and after the training and the standard deviation as well as the changes in that distance after the training.

Table 6. Results for vowel duration distance.

Training	Time Point	Mean Vowel Duration Distance (ms)	SD Vowel Duration Distance (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	22.92	16.61	−4.50
	After	18.42	12.36	
Segment	Before	19.63	12.73	3.22
	After	22.85	17.97	
Prosody	Before	11.42	12.65	3.91
	After	15.34	11.36	

The statistical model used vowel duration distance as the dependent variable. The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well the as the interaction between the two variables. The model included random intercepts for speakers (note that because we took the speaker means, there are only two observations per speaker—before training and after training).

We used a normally distributed prior with mean of 0.0 and standard deviation of 10.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercept, a Student's t distribution was used with degrees of freedom of 3.0, a mean of 15.3, and a standard deviation of 15.8 ($\nu = 3.0$, $\mu = 15.3$, $\sigma = 15.8$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student's t distribution ($\nu = 3.0$, $\mu = 0.0$, $\sigma = 15.8$). The model ran with four MCMC chains for 6000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that the distance between the vowel of *Rat* and the vowel of *Rad* was reduced during training, while a positive estimate indicates a growth of the distance between the two vowels and hence the opposite of neutralisation. $Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is. The results are given in Table 7. They show no reliable effect in any group.

The results show that all groups produce different vowel durations for the two target words, so they distinguish between them by means of vowel duration. There are only very minor changes after training, however, and statistical analysis showed that none of the changes were robust. Thus, there are no training effects for this parameter.

Table 7. Statistical results for vowel duration distance.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	Pr(β < 0)
Control	0.51	3.98	−6.04	6.97	0.45
Segment	2.35	3.76	−3.9	8.43	0.26
Prosody	2.24	3.89	−4.04	8.66	0.28

5.3. Analysis and Results for Neutralisation of Closure Duration in L2 German

Another way of assessing neutralisation effects of the training is to look at the absolute distance between the closure duration of /t/ in *Rat* and *Rad*. We ask whether the distance becomes smaller, i.e., whether the consonants of *Rat* and *Rad* become more similar after the training. Because, as with vowel duration, we are dealing with the relation between different productions, we first calculate the mean for each closure duration for each speaker. That is, for each speaker, we calculate the mean closure duration of *Rat* and the mean closure duration of *Rad*. Next, we calculate the absolute distance between these durations.

Table 8 presents the distance between closure durations in both target words in milliseconds for each group before and after the training and the standard deviation as well as the changes in that distance after the training. Again, a negative value for the difference would indicate that the distance between the vowel of *Rat* and the vowel of *Rad* was reduced during training.

Table 8. Results for closure duration distance.

Training	Time Point	Mean Closure Duration Distance (ms)	SD Closure Duration Distance (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	22.05	16.59	8.79
	After	30.83	18.82	
Segment	Before	32.17	21.58	2.75
	After	34.92	33.02	
Prosody	Before	20.64	16.18	5.89
	After	26.52	14.38	

For the statistical analysis, we used a model with closure duration distance as the dependent variable. The fixed effects were time of recording (before training or after training) and training type (control, segment, or prosody), as well the as the interaction between the two variables. The model included random intercepts for speakers (note that because we took speaker means, there are only two observations per speaker—before training and after training).

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercept, a Student's *t* distribution was used with degrees of freedom of 3.0, a mean of 23.1, and a standard deviation of 16.8 (ν = 3.0, μ = 23.1, σ = 16.8). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student's *t* distribution (ν = 3.0, μ = 0.0, σ = 16.8). The model ran with four MCMC chains for 6000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that the distance between the closure duration of *Rat* and the closure duration of *Rad* was reduced during training.

$Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is. The results, displayed in Table 9, show no reliable effect in any group.

Table 9. Statistical results for closure duration distance.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	Pr(β < 0)
Control	4.71	4.64	−3.11	12.27	0.15
Segment	2.41	5.09	−6.13	10.69	0.32
Prosody	5.54	4.92	−2.67	13.59	0.13

We can see that all groups distinguish between the two target words by means of closure duration at both time points. The changes after the training phase are minor, and according to our statistical analysis, none of them are robust, so no training effects are visible for this parameter either.

5.4. Analysis and Results for Reduction in Voicing during Closure in L2 German

As there are no measurable effects on vowel or consonant duration, we now look at vocal fold activity during the closure interval. Here, we look only at *Rad* because there was no voicing during closure in *Rat*. We measured the total duration of the consonant closure and that of the interval during which there was vocal fold vibration within the closure interval and calculated the percentage of voice during closure. Table 10 shows the mean percentages of voice during closure for all groups before and after the training as well as the standard deviation and the changes after the training. Negative values for the difference between the percentage of voicing during closure before and after training indicate an improvement. The results show improvements in all three groups but major changes only for the segment group.

Table 10. Voicing during closure results.

Training	Time Point	Voicing during Closure Mean (%)	Voicing during Closure SD (%)	Difference (%) (Mean After–Mean Before)
Control	Before	65.52	38.53	−12.60
	After	52.93	41.45	
Segment	Before	70.26	36.56	−50.93
	After	19.33	32.50	
Prosody	Before	82.02	34.66	−13.77
	After	68.25	38.74	

Figure 3 shows the raw data points as a jittered strip chart (grey dots) in addition to the means (coloured thick dots). It can be observed in the plot that the distributions of the voicing during closure data substantially deviate from a normal distribution. The points are half transparent, darker areas thus indicating the clustering of data points. There are many data points with values of 0% or 100%; i.e., there are a lot of closures that are either not voiced at all or fully voiced. Therefore, a model with a normal or skewed-normal distribution would produce a bad fit of the data. Instead, we transformed the data into the range of 0 to 1 (division by 100) and fitted a Bayesian zero/one inflated beta (ZOIB) model. The ZOIB model represents a mixture of a logistic and a beta regression. Therefore, the ZOIB model is able to estimate two interesting quantities in the context of this study. First, γ, the probability that an observation is 1. Second, μ, the mean of the continuous beta distribution in between 0 and 1. The two distributional parameters were estimated along with the precision of the beta distribution ϕ and the zero/one inflation α (the probability that an

observation is either 0 or 1), but we report only the results for γ and μ (for an introductory tutorial, see Vuorre 2021). The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well as the interaction between the two variables. The model included random intercepts for speakers and by-speaker random slopes for the effect of time of recording.

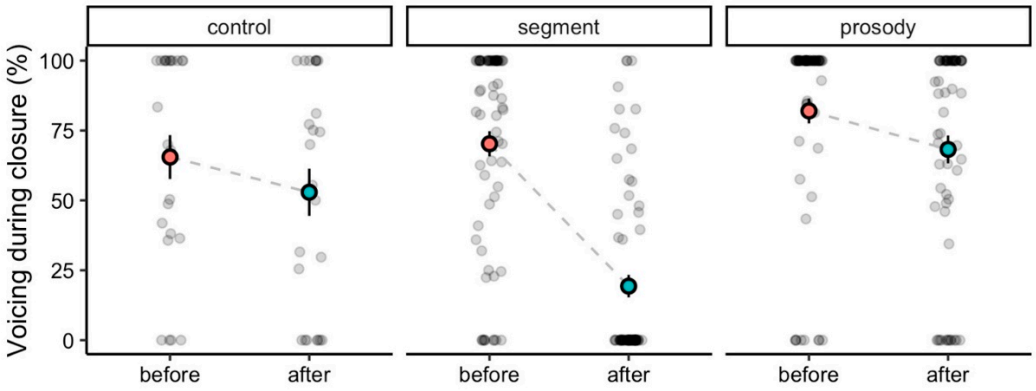

Figure 3. Voicing during closure (in %) for training groups before and after training (coloured thick points with bars: means and standard errors; grey points: raw measures).

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 1.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercepts of μ and ϕ, a Student's t distribution was used with degrees of freedom of 3.0, a mean of 0.0, and a standard deviation of 2.5 ($\nu = 3.0$, $\mu = 0.0$, $\sigma = 2.5$). As priors for the intercepts of γ and α, a logistic distribution was used ($\mu = 0$, $\sigma = 1$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student's t distribution ($\nu = 3.0$, $\mu = 0.0$, $\sigma = 2.5$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). The model ran with four MCMC chains for 8000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between recording time points (after training minus before training). We report the model estimate of the differences $\Delta\gamma$ and $\Delta\mu$, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\Delta\gamma < 0)$ or $Pr(\Delta\mu < 0)$. A negative estimate for the differences means that the voicing during closure was reduced during training. A negative difference $\Delta\gamma$ indicates that the probability of 1, i.e., full voicing, is reduced. A negative difference $\Delta\mu$ indicates that the means of the beta distribution in between 0 and 1 decreases; i.e., the relative duration of partial voicing during the closure is reduced. The results are presented in Table 11 (all estimates are in logit). There is strong evidence for a reduction in full voicings in the segment group, but not for the other groups. No group reliably reduces the mean of the beta distribution, relating to the relative duration of the partial voicings.

Table 11. Statistical results for voicing during closure.

			γ		
Training Group	**$\Delta\gamma$**	**SE**	**90% CI Low Boundary**	**90% CI High Boundary**	**Pr($\Delta\gamma < 0$)**
Control	−3.66	3.8	−10.12	1.59	0.89
Segment	−21.34	9.54	−38.63	−10.05	1.00
Prosody	−1.71	3.3	−6.05	2.91	0.80

Table 11. *Cont.*

	μ				
Training Group	**Δμ**	**SE**	**90% CI Low Boundary**	**90% CI High Boundary**	**Pr(Δμ < 0)**
Control	−0.19	0.31	−0.68	0.34	0.74
Segment	−0.27	0.31	−0.77	0.24	0.82
Prosody	−0.14	0.37	−0.74	0.5	0.66

5.5. Interim Summary and Discussion

In this section, we analysed the effects of an explicit segmental training of final devoicing, compared with an (implicit) syllable structure training, by investigating whether the subjects learned to neutralise the distinction between the words *Rat* and *Rad* by producing more-similar duration values for vowels and consonants in both words after the training. The results showed that the segmental training was not effective in that respect, which could be because the focus of the exercises was not on these aspects but was rather on the mere voicing neutralisation, i.e., the avoidance of voicing during closure and final aspiration for words such as *Rad*. Moreover, Italian learners of German encounter additional challenges when learning to modulate vowel duration in closed syllables because in their L1, closed syllables can have only a short vowel (leading to a consonant cluster or a geminate word-medially, as in [ˈfɛd.da] mentioned above).

Looking at voicing during closure, our results indicate that the segmental training was effective and led to a smaller number of word productions with fully voiced closures. In addition, as described in Section 4, there were positive effects with regard to the occurrence of epenthetic vowels in words with final (orthographically) voiced consonants. The control and prosody groups showed no reliable effects. Thus, the syllable structure training clearly had no effect on final devoicing. This once more supports the suggestion that final devoicing should be trained along with syllable structure, as syllable structure training helps to avoid epenthetic vowels, but only when the final consonant is voiceless; when the final consonant is interpreted as voiced on the basis of spelling, the training effects vanish. These results indicate that although training in final devoicing can support prosody training, the converse is not true: it is not implicitly acquired during prosody training, but it needs to be explicitly taught.

6. Voice Onset Time

German and Italian both have the plosives /p, t, k/ and /b, d, g/ in their consonant phoneme inventories, but they use different cues to distinguish between the two sets. Italian uses mainly voicing during closure (i.e., vocal fold activity during the consonant closure), whereas German uses mainly voice onset time, where /p, t, k/ is produced with a long voice lag (>30 ms) and /b, d, g/ with a short one (0–30 ms), while the vibration of the vocal folds during the consonant closure is not distinctive and generally only present when the plosive is surrounded by other voiced sounds (Jessen and Ringen 2002). The occurrence of aspirated plosives in Italian (i.e., with a positive VOT > 30 ms) is reported for some regions (Celata and Nagy 2022).

In this subsection, we examine the VOTs of all subjects from the three training groups for the word-initial plosive /t/ before and after the training phases to find out whether any changes towards longer positive VOTs are linked to the trainings that the test groups received. The segment group was explicitly made aware of the aspiration of plosives in German and of its significance for German natives to distinguish between words such as *'tennis'* and *Dennis* (a boy's name); see Section 2.3. The control and prosody groups received no explicit information or instruction on aspiration. However, the prosody group engaged in exercises for word stress, both on the phonological level (i.e., stress placement rules) and with regard to the phonetic features of word stress in German, which involve more articulatory effort and stronger air flow in stressed syllables, referred to as *Druckakzent*

(force accent). In order to generate the effort and pressure on stressed syllables, subjects were instructed to bang on the table with their fists when producing stressed syllables during the training sessions (not during the recordings). This may have had an effect on voiceless plosives in German stressed syllables, as the consonant release might have been stronger, resulting in a longer VOT. The influence of stress on VOT is reported in numerous studies (e.g., Lisker and Abramson 1967; Savino et al. 2015; Lein et al. 2016).

6.1. Data

The data analysed here were elicited in the reading tasks explained above. The target words were *Tina* and *Tennis* in the carrier sentences (cf Section 2.2):

(3) *Tina gab Hanna einen guten Rat.*
(5) *Helga spielte einmal Tennis.*

In total, 96 word realisations of the control group (8 speakers × 2 words × 3 repetitions × 2 recording times), 240 word realisations of the prosody group (12 speakers × 2 words × 5 repetitions × 2 recording times), and 260 word realisations of the segment group (13 speakers × 2 words × 5 repetitions × 2 recording times) entered the analysis.

6.2. Analysis and Results

Table 12 shows the mean VOTs for all groups before and after the training as well as the standard deviation and the difference between the mean values after the training and those before the training. All groups had already produced positive VOTs with mean values of over 30 milliseconds before the training, which shows that the subjects clearly pronounce voiceless plosives differently from their native productions, but with shorter VOTs than German natives speaking standard German (cf Kirby et al. 2020). The prosody group produced slightly shorter VOTs than the control and segment groups before the training. Positive values for the difference of mean VOTs before and after training indicate an improvement. Both test groups show longer VOTs after the training, with a slightly larger effect in the segment group. The control group exhibits a minor negative change. Figure 4 illustrates the changes in VOT for all groups.

Table 12. VOT results.

Training	Time Point	VOT Mean (ms)	VOT SD (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	33.54	21.66	−2.62
	After	30.92	14.10	
Segment	Before	34.02	17.99	9.51
	After	43.52	20.92	
Prosody	Before	28.77	16.06	6.68
	After	35.45	18.60	

For the statistical analyses, we used a mixed model with VOT as the dependent variable. The fixed effects were time of recording (before or after training), training type (control, segment, or prosody) and the interaction between the two variables. The model included random intercepts for speakers and target words, as well as by-speaker random slopes for the effect of time of recording. The model was fitted with a skewed-normal distribution to achieve a better model fit.

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, a Student's t distribution was used with degrees of freedom of 3.0, a mean of 14, and a standard deviation of 43 (ν = 3.0, μ = 31, σ = 19.3). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard

deviation of the model, we used a Student's t distribution ($\nu = 3.0$, $\mu = 0.0$, $\sigma = 19.3$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). The prior for the skewness parameter α for the skewed-normal distribution was a normal distribution with a mean of 0.0 and a standard deviation of 4.0. The model ran with four MCMC chains for 4000 iterations.

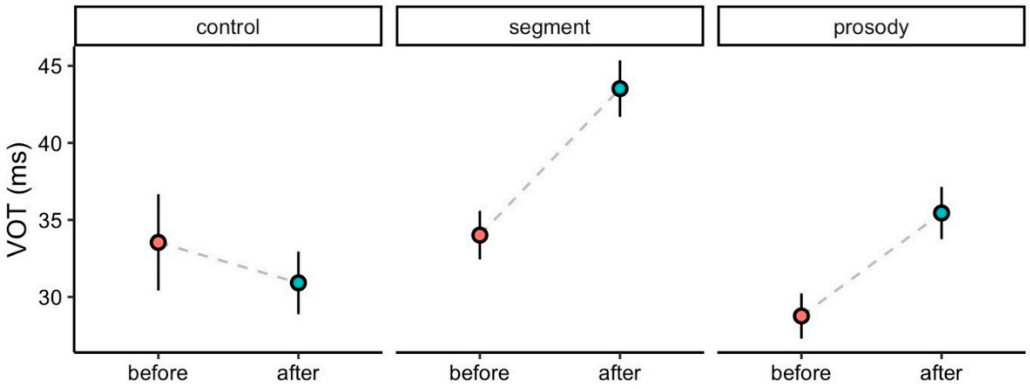

Figure 4. VOT for training groups before and after training (means and standard errors).

We assess the training effects by looking at the posterior distributions for the differences between recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the standard error of the estimate (SE) the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is positive $Pr(\beta > 0)$. A positive estimate for the difference means that the VOT became longer during training. $Pr(\beta > 0)$ gives an indication of how strong the evidence for a positive estimate is.

The results are presented in Table 13. The statistical estimates show that there is strong evidence for positive differences in the segment and prosody groups regarding the VOT with a $Pr(\beta > 0)$ of 1.0 in both cases, i.e., an increase in VOT during training. There is no reliable effect for the control group ($Pr(\beta > 0) = 0.59$). All in all, the statistical results show that the segment and prosody groups increase their VOTs for /t/ during training.

Table 13. Statistical results for VOT.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	Pr(β > 0)
Control	0.48	2.29	−3.30	4.24	0.59
Segment	7.08	1.74	4.24	9.93	1.00
Prosody	5.07	1.78	2.16	8.02	1.00

6.3. Interim Summary and Discussion

We analysed the effects of explicit segmental training and (implicit) prosodic training on the production of VOT in word-initial /t/. The results show positive effects for both trainings. Thus, training the phonetic features of word stress in German clearly improves learners' VOT in fortis plosives similarly to purely segmental training. This does not mean that segmental training can be skipped for this aspect; after all, we examined only the plosives in stressed syllables here, and the effects of the prosody training might not be present in unstressed syllables. Again, a combination of both segmental training and prosodic training would be beneficial.

7. General Summary and Discussion

In this paper, we examined the effects of prosodic training in a prosodic feature (intonation) and of a prosody-oriented training in an area where prosody and segments

interact (word-final codas). The effects of segment-oriented training were assessed for final obstruent devoicing, which is linked to the syllable structure and is thus partly prosodic, and for VOT of voiceless plosives, which is regarded as a segmental feature, although the temporal coordination of laryngeal and supralaryngeal gestures is not typical of what is regarded as segmental in nature. Table 14 summarises our findings (✓ refers to a training improvement, X refers to no training improvement).

Table 14. Summary of results.

Area	Measure	Control Group	Segment Group	Prosody Group
Prosody	Final rise magnitude in yes–no questions	**X**	**X**	✓
Prosody	Epenthetic vowel	**X**	**X** (Rad/Hund)	✓ (Rat/bunt)
Segment	Final devoicing: voicing during closure	**X**	✓	**X**
Segment	Final devoicing: vowel duration distance	**X**	**X**	**X**
Segment	Final devoicing: closure duration distance	**X**	**X**	**X**
Segment	Voice onset time (VOT)	**X**	✓	✓

One result that is not at all surprising is that explicit segmental training improves the production of segments and that prosody training improves the production of prosodic features. The intonation training yielded reliable positive results for final rises in yes–no questions only for the prosody group, which is also not surprising given that there is no relation between question intonation and any of the segmental features examined here. VOT, which is dependent on the prominence of syllables (stress and accent), is a good example of a segmental area that can be influenced by prosody training. However, as noted above, we looked only at contexts in which the plosive was in a stressed (and accented) syllable, so we do not know whether the effects of the word-stress training will hold for unstressed syllables. This might be an interesting point to investigate in further research with Italian learners. Epenthetic vowels and final devoicing both focus on the word-final consonant in training. Our analyses showed that the segment-oriented training (final devoicing of /d/) had positive effects, both at the segmental level (the learners produced less voicing during closure) and at the prosodic level (they produced fewer epenthetic vowels after a word-final <d>). However, the effect does not hold for words ending in consonants that do not undergo final devoicing. For the prosody group, no reliable effects of the syllable structure training were found on final devoicing. The mere fact that the training focused on the syllable coda did not make the learners aware of final devoicing in German. The syllable coda training showed effects only for the target words ending in <t>. The voiced final consonants in the orthography of words such as *Rad* and *Hund* appeared to facilitate vowel epenthesis, analogous to the native Italian pronunciation. Thus, the lack of instruction on final devoicing prevented a positive effect for the <d>-words, at least for the small set of data examined here. More research in this area, specifically research that involves more final consonants, is needed to obtain a clearer picture.

In sum, this study provides (somewhat limited) confirmation for the previous claims, made by Anderson-Hsieh et al. (1992), Munro and Derwing (1995), Derwing et al. (1998) and Gordon and Darcy (2016), that there are positive effects of prosody-oriented training on the production of segments, but this crucially depends on the area of prosody that is being trained. In our study, the training of syllable structure and the production of prosodic prominence (lexical stress and the placement of pitch accents, which were part of the training but not of the testing) is likely to have had a greater effect on the segments than the training of intonation contours. Interestingly, also in line with the above-mentioned studies, there were no reliable positive effects of segment-oriented training on prosodic features. This was even the case when the training aimed at an area where segments and prosody interact, as is the case for final obstruent devoicing.

A limitation of this study is the small data set, making it difficult to generalise. Nonetheless, our results appear to indicate that prosodic training and segmental training are best treated in an integrated way. In particular, if aspiration is taught alongside stress and accent, aspiration can be learned in this hyperarticulated context, making the difference between L1 and L2 clearer. Moreover, there appear to be benefits in teaching final devoicing alongside syllable structure, including avoiding schwa epenthesis and thus restructuring of the word, making a "final" consonant in fact initial to a further syllable. Learning to devoice obstruents in syllable onsets instead of codas could otherwise lead to possible problems with learning to adequately produce the voicing distinction in onset position. Thus, our results support the conclusions drawn by Derwing and Munro (2015): if the segmental and prosodic levels are taught together, there is a greater likelihood of an overall beneficial outcome in pronunciation training.

Author Contributions: Conceptualization, S.D. and M.G.; methodology, S.D., M.G. and S.R.; software, S.R.; validation, S.D., M.G. and S.R.; formal analysis, S.D. and S.R.; investigation, S.D.; resources, S.D. and M.G.; data curation, S.D. and S.R.; writing—original draft preparation, S.D., M.G. and S.R.; writing—review and editing, S.D., M.G. and S.R.; visualization, S.R.; supervision, M.G.; project administration, S.D.; funding acquisition, M.G. and S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the collaborative research center SFB1252 Prominence in Language Project-ID 281511265) and project RO 6767/1-1 (Walter Benjamin program).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data and analysis scripts are available online on the Open Science Framework: https://osf.io/mfbw3/ (accessed on 2 February 2023).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix complements the substudy on the yes–no question rise magnitudes in the paper. Because both German and Italian commonly have final rises in such questions (albeit with much regional variation), we can link the effects of training to patterns of realisation in the respective languages as produced by native speakers. In this comparison, it becomes evident that Italian yes–no question rises are smaller in magnitude compared with German rises.

The data analysed here are recordings of Italian and German native speakers. We collected two data sets in order to be able to compare Italian L1 and German L1 realisation patterns. The first data set consists of recordings of Italian native speakers. It was collected at a grammar school in Turin (Northern Italy). In total, eight students (three male, five female) were recorded while playing card games specifically designed to elicit yes–no questions. These students were also later part of the training groups described in the main text of the paper (three in the prosody group, five in the segment group). The card games were played in pairs. The recordings were conducted in a quiet room in a school in Turin with a mobile DAT recorder and head-mounted microphones. The speakers were 17 to 18 years old.

The second data set contains recordings of four native speakers of German, two of them authors of this paper (S.D. and S.R.). The recordings took place in a sound-attenuated recording booth at the University of Cologne, using head-mounted microphones (recorded directly on the hard disk of a computer through an external audio interface). The speakers were aged between 22 and 35 years; two of them identified as female, two as male.

The data were elicited in a card game setting in which the subjects played in pairs. The cards in this game depicted day-to-day objects in different colours. The participants' task was to collect cards with the same colour or the same object by exchanging cards with their fellow player. To initiate the exchange, participants ask their fellow player whether

they are in possession of a specific card. For example, *do you have a green coffee pot/carafe?* (German: *hast du eine grüne Kanne?*, Italian: *hai una caraffa verde?*).

In each move, the colour or object that the player can use in their question is determined by a card from an additional stack. In one version of the game, it is the colour that is displayed by this card; in another version, it is the object. For example, if the card is green, the participant may ask *do you have a green coffee pot?* but not *do you have a blue coffee pot?* Similarly, when the card displays a coffee pot, the participant may ask *do you have a green coffee pot?* but not *do you have a green plate?* There was no visual contact between the participants; for communication, they relied solely on the auditory channel.

The German colour adjectives were *blaue/blauen* 'blue', *gelbe/gelben* 'yellow', *rote/roten* 'red',and *grüne/grünen* 'green'. In Italian, they were *azzurro/azzurra* 'blue', *giallo/gialla* 'yellow', *rosso/rossa* 'red', and *verde* 'green'. The German object nouns were *Kanne* 'coffee pot', *Teller* 'plate', *Gabel* 'fork', and *Kugel* 'ball'. In Italian, they were *caraffa* 'carafe', *piatto* 'plate', and *tazza* 'cup'. As exemplified above, the questions were of the form 'hast du eine(n) <colour> <object>?' for German and 'hai un(a) <object> <colour>?' for Italian.

From the 136 questions in the German L1 data, 13 were excluded because of hesitations, laughter, or mispronunciations; 13 were excluded because the speaker asked for two objects (in an alternative question such as *hast du eine grüne Kanne oder einen gelben Teller?* 'do you have a green carafe or a yellow plate?'). Moreover, 14 questions did not end in a simple rising intonation contour: seven of them were falling (H* L-%) and seven had a falling-rising nuclear contour (H* L-H%). Thus, for the investigation of the final rise magnitude, 96 German questions could be used. All these questions reflect the nuclear intonation pattern L* H-^H% described in Grice and Baumann (2002) for neutral German yes–no questions. An example from the data set is given in Panel C of Figure A1.

For the Italian L1 data, 110 questions were recorded from a group of eight speakers. Here, 11 questions were excluded because of hesitations, laughter, or mispronunciations; four were excluded because they were alternative questions, as in the German data described above. Of the Italian questions, 24 ended in a falling boundary tone (see Panel B of Figure A1). These were mainly by two speakers who exclusively produced rising-falling contours (L+H* L-L%). For the analysis of the final rise, these speakers had to be excluded. Hence, 71 Italian questions with a final rise elicited from six speakers remained in the data set for this measurement. The nuclear intonation contour of these questions can be described as (H+)L* nuclear accent, followed by a rising boundary tone (see Panel A of Figure A1).

In both languages, the start and end points of the final rise were annotated. Start point "L" was placed on an F0 minimum in the vowel of the syllable with the nuclear L* accent, e.g., in [a] of *Kanne* in *hast du eine grüne Kanne?* and in [u] of *azzurro* in *hai un piatto azzurro?*[2] End point "H" was placed on the F0 maximum at the end of the utterance. The rise magnitude was calculated in semitones: $12 \log_2 \left(\frac{E}{B} \right)$, where B denotes the F0 in Hz at the beginning of the rise and E denotes the F0 in Hz at the end of the rise.

Panel D of Figure A1 shows the results obtained from the measure of the final rise. The violin plots show the distributions of the data. The thick black dots represent the respective means of these distributions. The graph presents a clear picture, where German exhibits substantially larger final rises (13.1 st) than does Italian (5.23 st).

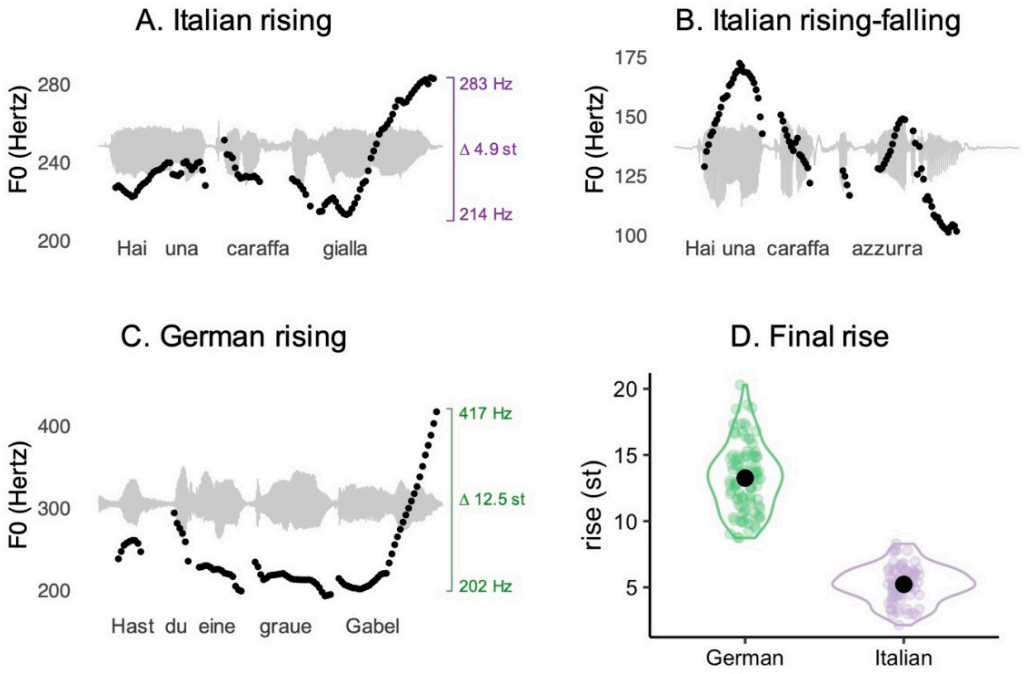

Figure A1. Example contours (**A**–**C**) and final rise magnitudes of German L1 and Italian L1 (**D**).

Notes

[1] For the intonation of questions, the data were collected during the project, but they have been analysed here for the first time.

[2] The labels deviated from this pattern only when there was no reliable F0 calculation in the vowel that was due to creaky voice.

References

Anderson-Hsieh, Janet, Ruth Johnson, and Kenneth Koehler. 1992. The Relationship between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning* 42: 529–55. [CrossRef]

Baills, Florence, Charlotte Alazard-Guiu, and Pilar Prieto. 2022. Embodied prosodic training enhances accentedness and general suprasegmental accuracy. *Applied Linguistics* 43: 776–804. [CrossRef]

Bürkner, Paul-Christian. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10: 395–411. [CrossRef]

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76: 1–32. [CrossRef]

Celata, Chiara, and Naomi Nagy. 2022. Sociophonetic Variation and Change in Heritage Languages: Lexical Effects in Heritage Italian Aspiration of Voiceless Stops. *Language and Speech. online first*. [CrossRef]

Chun, Dorothy, and John Levis. 2020. Prosody in Second Language Teaching. In *The Oxford Handbook of Language Prosody*. Edited by Carlos Gussenhoven and Aoju Chen. Oxford: Oxford University Press, pp. 619–32.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment–Companion Volume*. Strasbourg: Council of Europe Publishing.

Dahmen, Silvia. 2013. *Prosodie oder Segmente? Phonetische Untersuchungen zu Trainingseffekten bei Italienischen Deutschlernenden*. Cologne: Cologne Publication Service. Available online: https://kups.ub.uni-koeln.de/5368/ (accessed on 2 February 2023).

Derwing, Tracey M., and Murray J. Munro. 1997. Accent, Intelligibility and Comprehensibility: Evidence from Four L1s. *Studies in Second Language Acquisition* 19: 1–16. [CrossRef]

Derwing, Tracey M., and Murray J. Munro. 2015. *Pronunciation Fundamentals. Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Derwing, Tracey M., Murray J. Munro, and Grace Wiebe. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410. [CrossRef]

Dieling, Helga, and Ursula Hirschfeld. 2000. Phonetik lehren und lernen. Fernstudieneinheit 21. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 5: 3.

Gordon, Joshua, and Isabelle Darcy. 2016. The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation* 2: 56–92. [CrossRef]

Grice, Martine, and Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191: 267–98.

Grice, Martine, Michelina Savino, Alessandro Caffò, and Timo B. Roettger. 2015. The tune drives the text–Schwa in consonant-final loanwords in Italian. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.

Hayes-Harb, Rachel, Brown Kelsey, and Bruce L. Smith. 2018. Orthographic Input and the Acquisition of German Final Devoicing by Native Speakers of English. *Language and Speech* 61: 547–64. [CrossRef]
Jessen, Michael, and Catherine Ringen. 2002. Laryngeal features in German. *Phonology* 19: 189–218. [CrossRef]
Kaunzner, Ulrike. 2015. Die Verständlichkeit vorgelesener Texte durch nichtdeutsche Sprecher. Zur Sprechwirkung bei muttersprachlichen und nicht-muttersprachlichen Rezipienten. In *Aktuelle Forschungstendenzen in der Sprechwissenschaft: Normen, Werte, Anwendung, hrsg. v. B. Teuchert (= Sprache und Sprechen)*. Baltmannsweiler: Schneider, pp. 65–81.
Kaunzner, Ulrike. 2018. Das klingt sympathisch!. Selbst- und Fremdbild in der Sprechwirkung des italienischen Akzents. In *Gesprochene (Fremd-) Sprache als Forschungs- und Lehrgegenstand*. Trieste: EUT Edizioni Università di Trieste, pp. 139–55.
Kirby, James, Felicitas Kleber, Jessica Siddins, and Jonathan Harrington. 2020. Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German. Paper presented at the 10th International Conference on Speech Prosody 2020, Tokyo, Japan, May 25–28; pp. 210–14.
Lein, Tatjana, Tanja Kupisch, and Joost van de Weijer. 2016. Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–49. [CrossRef]
Li, Peng, Florence Baills, Lorraine Baqué, and Pilar Prieto. 2022. The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research*. *online first*. [CrossRef]
Li, Peng, Xiaotong Xi, Florence Baills, and Pilar Prieto. 2020. Appropriately performing hand gestures cueing phonetic features facilitates simultaneous speech imitation in an L2. Paper presented at the 7th GeSpIn Conference, GeSpIN 2020, Virtual, September 7–9; Sweden: KTH Royal Institute of Technology.
Lisker, Leigh, and Arthur S. Abramson. 1967. Some Effects of Context On Voice Onset Time in English Stops. *Language and Speech* 10: 1–28. [CrossRef] [PubMed]
Mennen, Ineke. 2007. Phonological and phonetic influences in non-native intonation. In *Non-Native Prosody. Phonetic Description and Teaching Practice*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin and New York: De Gruyter, pp. 53–76.
Missaglia, Federica. 1999a. *Phonetische Aspekte des Erwerbs von Deutsch als Fremdsprache durch italienische Muttersprachler*. Frankfurt am Main: Hector.
Missaglia, Federica. 1999b. Contrastive prosody in SLA—An empirical study with adult Italian learners of German. Paper presented at the ICPhS, San Francisco, CA, USA, August 1–7; pp. 551–54.
Missaglia, Federica. 2007. Prosodic Training for Adult Italian Learners of German: The Contrastive Prosody Method. In *Non-Native Prosody. Phonetic Description and Teaching Practice*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin and New York: De Gruyter, pp. 237–58.
Munro, Tracey M., and Murray J. Derwing. 1995. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45: 73–97. [CrossRef]
Nalborczyk, Ladislas, Cédric Batailler, Hélène Lœvenbruck, Anne Vilain, and Paul-Christian Bürkner. 2019. An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research* 62: 1225–42. [CrossRef]
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 20 February 2023).
Roettger, Timo B., and Micheal Franke. 2019. Evidential Strength of Intonational Cues and Rational Adaptation to (Un-)Reliable Intonation. *Cognitive Science* 43: e12745. [CrossRef]
Roettger, Timo B., Bodo Winter, Sven Grawunder, James Kirby, and Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43: 11–25. [CrossRef]
Saito, Kazuya, and Luke Plonsky. 2019. Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Language Learning* 69: 652–708. [CrossRef]
Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42: 23–48. [CrossRef]
Savino, Michelina, Martine Grice, and Alessandro Caffo. 2015. The Influence of Prominence on the Production of Plosives in Italian. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.
Sluyters, Willebrord. 1990. Length and Stress Revisited: A Metrical Account of Diphthongization, Vowel Lengthening, Consonant Gemination and Word-final Vowel Epenthesis in Modern Italian. *Probus* 2: 65–102. [CrossRef]
Thomson, Ron. 2017. Measurement of accentedness, intelligibility, and comprehensibility. In *Assessment of Second Language Pronunciation*. Edited by Okim Kang and April Ginther. London: Routledge, pp. 11–29.
Ulbrich, Christiane, and Ineke Mennen. 2016. When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism* 20: 522–49. [CrossRef]
van de Schoot, Rens, Joris J. Broere, Koen H. Perryck, Mariëlle Zondervan-Zwijnenburg, and Nancy E. van Loey. 2015. Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology* 6: 25216. [CrossRef]
van Maastricht, Lieke, Tim Zee, Emiel Krahmer, and Marc Swerts. 2021. The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Commun* 133: 81–90. [CrossRef]

Vasishth, Shravan, Bruno Nicenboim, Mary E. Beckman, Fangfang Li, and Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71: 147–61. [CrossRef] [PubMed]

Vuorre, Matti. 2021. Sometimes I R: How to Analyze Visual Analog (Slider) Scale Data? Available online: https://web.archive.org/web/20210902234333/https://mvuorre.github.io/posts/2019-02-18-analyze-analog-scale-ratings-with-zero-one-inflated-beta-models/ (accessed on 2 February 2023).

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. Available online: https://ggplot2.tidyverse.org (accessed on 2 February 2023).

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, and Jim Hester. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4: 1686. [CrossRef]

Yang, Chunsheng, Jing Chu, Si Chen, and Yi Xu. 2021. Effects of Segments, Intonation and Rhythm on the Perception of L2 Accentedness and Comprehensibility. In *The Acquisition of Chinese as a Second Language Pronunciation. Prosody, Phonology and Phonetics*. Edited by Chunsheng Yang. Singapore: Springer. [CrossRef]