RESEARCH   **Open Access**

Production and perception of velar stop (de)voicing in European Portuguese and Italian

Daniel Pape[1*] and Luis M T Jesus[1,2]

Abstract

Speech production and speech perception studies were conducted to compare (de)voicing in the Romance languages European Portuguese (EP) and Italian. For the speech production part, velar stops in two positions and four vowel contexts were recorded. The voicing status for 10 consecutive landmarks during stop closure was computed. Results showed that during the complete stop closure voicing was always maintained for Italian, and that for EP, there was strong devoicing for all vowel contexts and positions. Both *language* and *vowel context* had a significant effect on voicing during stop closure. The duration values and voicing patterns from the production study were then used as input factors to a follow-up perceptual experiment to test the effects of *vowel duration*, *stop duration* and *voicing maintenance* on voicing perception by EP and Italian listeners. Perceptual stimuli (VCV) were generated using biomechanical modelling so that they would include physically realistic transitions between phonemes. The consonants were velar stops, with no burst or noise included in the signal. A strong language dependency of the three factors on listeners' voicing distinction was found, with high sensitivity for all three cues for EP listeners and low sensitivity for Italian listeners. For EP stimuli with high voicing maintenance during stop closure, this cue was very strong and overrode the other two acoustic cues. However, for stimuli with low voicing maintenance (i.e. highly devoiced stimuli), the acoustic cues vowel duration and stop duration take over. Even in the absence of both voicing maintenance during stop closure and a *burst*, the acoustic cues vowel duration and stop duration guaranteed stable voicing distinction in EP. Italian listeners were insensitive to all three acoustic cues examined in this study, with stable voiced responses throughout all of the varying fully crossed factors. None of the examined acoustic cues appeared to be used by Italian listeners to obtain a robust voicing distinction, thus pointing to the use of other acoustic cues or combination of other cues to guarantee stable voicing distinction in this language.

Keywords: Voicing maintenance; Devoicing; Cue weighting; European Portuguese; Italian; Speech production; Speech perception

Introduction

Human perception does not rely only on a single perceptual cue, but also rather on a combination of different cues to guarantee a stable and robust perceptual outcome. One phonological contrast can be triggered by multiple acoustic cues or by a combination of these cues.

Several acoustic cues (separately or combined) form the perceptual construct of a voicing distinction between voiced and voiceless stops. For the perception of stop voicing in general, voice onset time (VOT) is seen as one of the major acoustic cues in a variety of languages, the other major cues being stop duration and adjacent vowel duration [1-5]. For stops in medial position, the classical definition of a voiced stop is *voicing during closure* [6]: intervocalic stops usually have voicing during closure in a number of languages, which influences perception [6,7]; however, other acoustic cues like *stop duration* [7] and preceding *vowel duration* [8] strongly contribute to voicing distinction. Adding to this complexity, Li et al. [9] also showed that a number of other acoustic cues can be extracted from natural speech which may also play a role as perceptual cues. Regarding

---

* Correspondence: danielpape@ua.pt
[1]Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal
Full list of author information is available at the end of the article

the interplay of several acoustic cues, Francis et al. [7] studied English simultaneous variation and cue weighting in the stop voicing distinction. They showed that if the stop closure duration in the word *rabid* is increased to more than 70 ms, then listeners hear the word *rapid*, but *only* when there is no voicing, i.e. in the absence of voicing maintenance during stop closure. Bailey and Summerfield [10] also showed in English that in fricative stop syllable initial clusters (e.g. *star*, *spar* or *scar*), all of the acoustic differences measured in natural production are able to provide correct stop perception results. No single cue was necessary, and many different cues in combination were sufficient for a correct perception. Redundant cues often enter into trading relationships that increase the magnitude of certain cues while decreasing the magnitude of others, i.e. the listener decision is the result of a certain cue trading of the perceptual system. The complexity increases when taking into account language differences; Oglesbee [11] showed that, in a multidimensional stop categorisation task, a comparison of listeners from different languages results in different preferences for the important perceptual cues, thus pointing to a language dependency for both selection and weighting of available stop contrast cues.

Even for various languages of the same family like Italian, Spanish and European Portuguese (EP), the role of acoustic cues and cue trading is rather difficult to predict. For example, Italian shows a stable *voicing maintenance* (the presence of voicing throughout the complete stop closure) for all word-medial phonologically voiced stops, as shown by Shih et al. [12] in a cross-linguistic database comparison. EP, in contrast, has completely different voicing maintenance for their phonologically voiced stops. Recent findings [13] showed that EP phonologically voiced stops often have no discernible burst, and more importantly, most of the EP phonologically voiced stops in speech production studies [13,14] are produced as devoiced. Thus, it would be reasonable to assume that the absence of voicing maintenance during stop closure in combination with the missing burst would lead EP listeners to identify the phonologically voiced consonants as voiceless. However, it is also possible that other cues take over to guarantee a robust perceptual outcome. Speech production results for EP point to the hypothesis that stop distinction would not be all that different from other languages, since both stop durations [5,13-16] and preceding vowel durations [14] are significantly different for the contrast of phonologically voiced and voiceless stops.

One aim of this study was to disentangle the described controversies of EP by examining the presence of several acoustic cues in the production of EP stops and compare the actual use of these acoustic cues in the perception of EP voicing distinction and to shed light on the interplay of the different cues for intervocalic medial velar stops, all in the absence of a facilitating burst. The second aim of this study was to contrast EP production and perception data from (matched) Italian production and perception data to find similarities and/or differences in the voicing behaviour of these two Romance languages. In other words, we were interested in comparing production and perception data from these two languages to find out how the perceptual system evaluates a voiced/voiceless distinction, which of the available acoustic cues are chosen and how a possible interplay of several cues is ordered.

This paper presents novel results on how particular languages select and weight available acoustic cues to achieve a robust perceptual outcome. This question is particularly relevant in the light of the results of Oglesbee [11], who showed that a comparison of listeners from different languages leads to different preferences for important perceptual cues, thus pointing to a language dependency for the selection and weighting of several stop contrast cues.

Method
Production study
Acoustic (MK 224 microphone and MV 181A preamplifier, Cirrus Research, Hunmanby, UK) and electroglottography (EGG) (EG2-PCX, Glottal Enterprises, Syracuse, NY, USA) signals of EP and Italian $C_1V_1C_2V_2$ clusters in the carrier sentence (EP, *Diga $C_1V_1C_2V_2$ outra vez*/'Say $C_1V_1C_2V_2$ again'; Italian, *Dite $C_1V_1C_2V_2$ ogni dì*/'Say $C_1V_1C_2V_2$ once more') were recorded (PMD 671 Solid State Recorder, Marantz, Eindhoven, The Netherlands; 16 Bits, 48 kHz) in a soundproof environment. As can be seen from the carrier sentences, both preceding and following vowels of the target $C_1V_1C_2V_2$ items were very similar for the two languages, thus excluding influences from differing vowel or consonant context. Both the velar stops /k g/ and the four vowel contexts /i e o a/ were pairwise identical for the target item (e.g. 'kaka' or 'gigi'), sentence stress was laid on the $C_1V_1C_2V_2$ pseudoword, and lexical stress was set to its first syllable. Thus, we obtained two different consonant positions (intervocalic *initial* and *medial* position). For both languages, each item was repeated nine times in randomised order by six native speakers with university education. The EP speakers were from Central Portugal (mean age 25 years, recorded at the Speech, Language and Hearing Laboratory (SLHlab), University of Aveiro, Portugal), and the Italian native speakers were from the Veneto region in Northern Italy (mean age 40 years, recorded at the Instituto di Scienze e Tecnologie della Cognizione (ISTC), Padova, Italy). Speech rate was held constant during all recordings; correct realisations of all items were supervised by a

trained phonetician (both languages). One trained phonetician manually labelled the following landmarks:

- Onset and offset of the neutral lead-in vowel (preceding the target word)
- Onset of first stop (V_1) burst, if present
- Onset and offset of the first target vowel (V_1)
- Onset of the second stop (C_2) burst, if present
- Onset and offset of the second target vowel (V_2)

Similar to [12], we computed the time-dependent voicing status for all stops, sampled at 10 equidistant points throughout the complete stop duration. The first point was set to the beginning of the stop closure (i.e. the preceding vowel offset), whereas the 10th point was set to the following vowel onset. To determine the voicing status for each point a PRAAT 5.2 [17] automatic voicing detection algorithm (*AC pitch extraction* algorithm with the settings *voiceless decision* = 0.55 and *silence threshold* = 0.1) was used and then manually checked against the synchronised EGG signals.
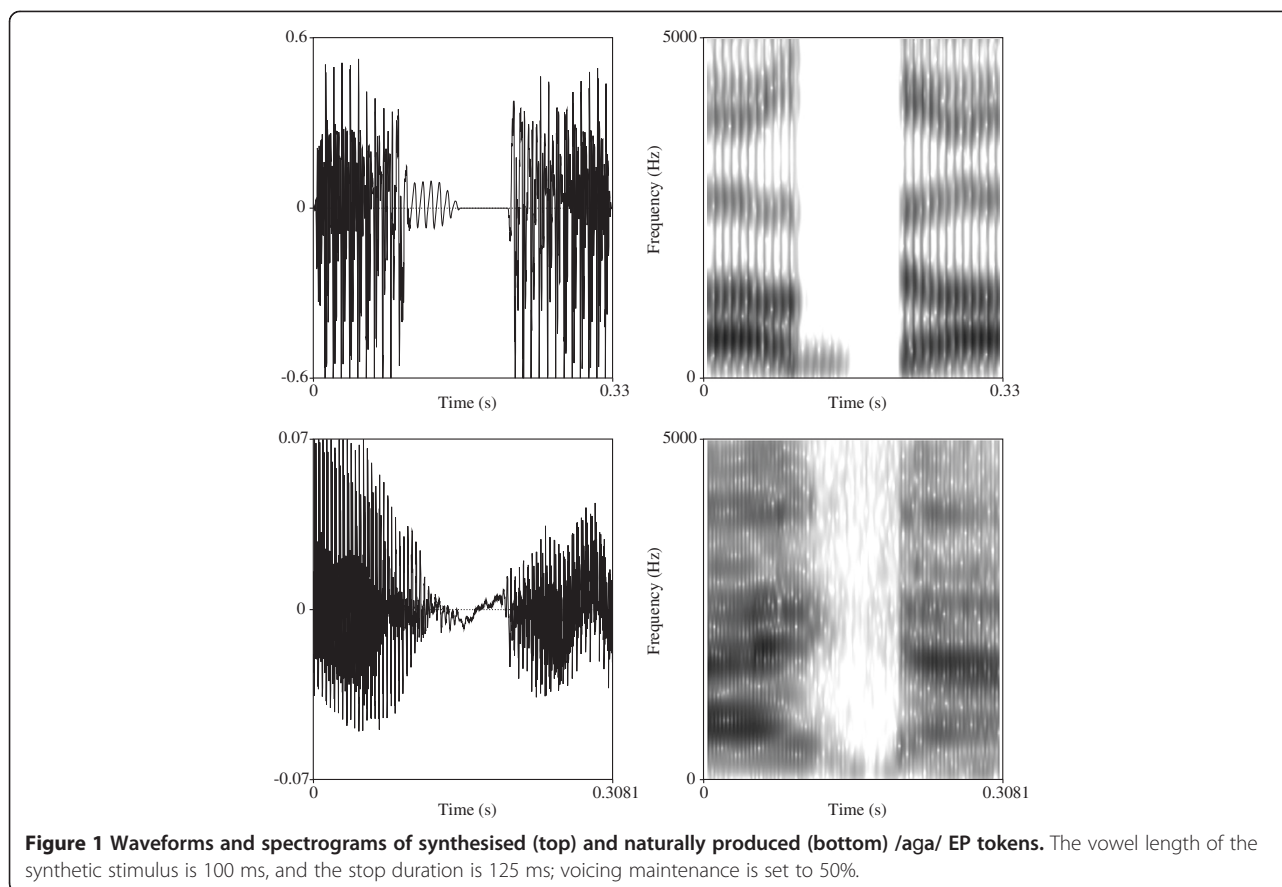
Perception study

Thirty two native EP participants from Central Portugal and 10 native Italian participants from the Veneto region listened to stimuli in a soundproof room (EP, SLHlab Aveiro; Italian, ISTC Padova). They did not receive course credit or financial compensation for their participation. None of the listeners reported speech or hearing problems. Open headphones with a linear frequency response (Sennheiser HD 600, Wedemark Wennebostel, Germany) connected to the internal headphone output of a notebook computer (no other processes running and all networking interfaces disabled) were used. Listeners' responses were collected by means of mouse clicks placed on two screen buttons. The loudness of the stimuli presentation was held constant across all listeners to a comfortable level. The sampling frequency of the presented stimuli was 48,000 Hz at 16 bits.

The speech material generated for the perceptual experiments (extensively described in [18]) consisted of biomechanically modelled VCV stimuli [19] acoustically synthesised frame by frame with a parametric model of the vocal tract [20] driven by a three-mass vocal fold model [21]. The reason for using biomechanical modelling in contrast to, for example, a Klatt synthesiser, lies in the ability of the biomechanical models to generate physically realistic trajectories between consecutive phonemes. In other words, articulatory trajectories are not linearly interpolated, as is normally the case with other synthesis approaches. Research on trajectories has shown that the characteristics of curved paths are explained by anatomical factors and muscle mechanics, for arm movements [22-24] as well as for speech

movements [19,25,26]. Biomechanical modelling, in contrast to other synthesis approaches, has the advantage that all obtained tongue movements, trajectories and phoneme targets are comparable to natural speech. This allows the manipulation of glottal source parameters while maintaining articulatory realism. Thus, the use of biomechanical modelling is the best compromise to guarantee highly realistic perceptual stimuli, without the risk of missing hidden perceptual cues (which cannot be controlled for) when using manipulated natural speech. In other words, the use of the biomechanical model has the main advantage that all obtained tongue movements, trajectories and phoneme targets are comparable to natural speech (see for example the modelling of articulatory loops), with the additional possibility to manipulate relevant temporal and glottal source parameters while at the same time maintaining articulatory realism. Thus, the use of biomechanical modelling is the best compromise to guarantee highly realistic perceptual stimuli and to independently control parameters such as duration, transition and targets. Figure 1 shows the comparison of synthesised waveforms and spectrograms of the generated /aga/ stimulus (top) with the /aga/ item as produced by an EP speaker (bottom).

Three different factors known to influence the perception of stop voicing were examined: stop duration, contextual vowel duration and voicing maintenance during stop closure. Each factor was laid out in a continuum with several levels and was combined with all levels of the other factors (i.e. fully crossed and non-adaptive design). The extremes of the continua (perception experiments) were determined according to the values of the speech production study described before:

1. Stop duration: mean durations (rounded to the closest decimal) of the voiceless and phonologically voiced velar stops /k g/ in the vowel contexts /a o/ were taken as the limits of the stop duration continuum, i.e. 100 ms (mean of the voiced stop) and 150 ms (mean of the voiceless stop). One intermediate value (125 ms) was introduced.
2. Vowel duration: mean durations (rounded to the closest decimal) of the preceding vowels /a o/ of the voiceless/voiced velar stops /k g/ were taken as the limits of the vowel duration continuum, i.e. 70 ms (mean of the preceding vowel of the voiceless velar stop) and 130 ms (mean of the preceding vowel of the voiced velar stop). One intermediate value (100 ms) was introduced.
3. Voicing maintenance: the voicing maintenance continuum was defined by the two endpoints fully voiced and fully devoiced/voiceless. For the intermediate values, five conditions were defined (12.5%, 25%, 37.5%, 50% and 75%) at which the stop

Figure 1 Waveforms and spectrograms of synthesised (top) and naturally produced (bottom) /aga/ EP tokens. The vowel length of the synthetic stimulus is 100 ms, and the stop duration is 125 ms; voicing maintenance is set to 50%.

voicing ceases and thus the stop devoicing begins (and remains until its offset). The unequal step sizes result from the hypothesis that the perceptual differences would be smaller towards higher voicing percentages of the stimuli, so smaller step sizes for higher voicing percentages were excluded to obtain a reduced total number of stimuli. The fully devoiced condition denotes different underlying control mechanisms than the voiceless condition [27,28], although the result, i.e. the voicing maintenance, is identical in both conditions.

A three-factor design with $3 \times 3 \times 7$ levels of the corresponding continuum was thus used in the perception experiment (see Table 1). In this experiment, all possible combinations of contextual vowel duration (70, 100 and 130 ms) with all combinations of stop duration (100, 125 and 150 ms) and all voicing maintenance steps (0%, 12.5%, 25%, 37.5%, 50%, 75% and 100%) were tested. The experiment was performed in two different vowel conditions (/a/ and /o/). Five repetitions of the complete stimuli set for each of the listeners were played in randomised order. In summary, a total of 630 stimuli (three vowel durations × three *closure durations* × seven voicing maintenance conditions × two *vowel* identities × five

repetitions) were generated. The average time to perform the task was 20 min. There was a practice session of 25 stimuli prior to the beginning of the main experiment. Listeners were informed that they would hear synthetic VCV items, and their task was to identify whether the consonant was /g/ or /k/ (forced choice). Speed of response was emphasised, asking listeners to respond as quickly and accurately as possible. Stimulus repetition was not possible. *Alvin* v 1.27 [29] open source software for stimulus and visual presentation was used. The computer screen for the *identification task* showed two buttons (labelled *g* and *k*) at identical distances around a *next* button at the screen centre. After selecting their response, listeners had to click on the next button to

Table 1 Three fully crossed factors for the perceptual study: vowel duration, stop duration and percentage of stop voicing maintenance

Factor	Continuum levels
Stop duration (ms)	100, 125, 150
Vowel duration (ms)	70, 100, 130
Voicing maintenance (%)	0, 12.5, 25, 37.5, 50, 75, 100

The 0% and 100% voicing maintenance correspond to a fully devoiced and a fully voiced stimulus, respectively.

proceed, thus placing the cursor at the exact centre of the screen before the next stimulus presentation (guaranteeing identical distances for the two answer possibilities). All button options and accompanying text were written in Portuguese for EP listeners and in Italian for Italian listeners in order not to confuse listeners' internal language presentation. The placement of all buttons was rotated 180° for one half of the participants, thus counterbalancing biases of horizontal movement and listener preference.

Statistical analysis
Production study
For the statistic validation of the (production) voicing patterns, we chose to examine the three central landmarks (i.e. point 5, point 6 and point 7) as the dependent variables for the statistical analyses. The obstruent onset (point 1) and offset (point 10) are voiced by definition (i.e. vowel formants offset and onset), but the more central points can be regarded as a valid representative to examine significant voicing differences during stop closure. To statistically analyse the devoicing behaviour at these landmarks, a series of logit models (function lmer [30]) with mixed effects were run in the R environment [31]. The logit models are based on binomial distributions (z-scores, generalized linear mixed model, GLMM). This allows modelling based on binary decisions [32,33], since a binary voicing decision (either *voiced* or *voiceless/devoiced*) is obtained for each of the 10 consecutive landmarks (point 1 to point 10). The devoicing occurrences of the three central points of the phonologically voiced velar stop /g/ (dependent variables) were analysed with a $p < 0.05$ significance threshold for effects of the factors *language* (EP, Italian), *consonant position* (initial, medial) and *vowel context* /i e o a/, and their interactions. All numerical fixed factors were centred (z-transformation). *Speaker* was chosen as random factor.

Perception study
A series of logit models with mixed effects was run to statistically analyse the listeners' response patterns. Again, logit models based on binomial distributions (z-scores, Generalized Linear Mixed Model) were used to model binary decisions [33] by our listener responses (i.e., the listeners' /g/ or /k/ response). The dependent variable was the listener response; fixed factors were language (EP, Italian), stop duration (100, 125 and 150 ms), contextual vowel duration (100, 125 and 150 ms) and voicing maintenance percentage during stop closure (0%, 12.5%, 25%, 37.5%, 50%, 75%, 100%) and their interactions. All numerical fixed factors were centred (z-transformation). *Listener* was chosen as random factor.

Results
Production study
Table 2 presents the mean durations for the medial position of the production database (initial position is not given due to the occurring word boundary). It can be seen that the preceding vowels are longer for Italian compared to EP, but velar stop durations are almost identical for the two languages. Furthermore, for both languages, the classical duration pattern can be observed: vowels are longer when preceding voiced stops in contrast to voiceless stops, and stop durations are longer for voiceless stops as compared to voiced stops.
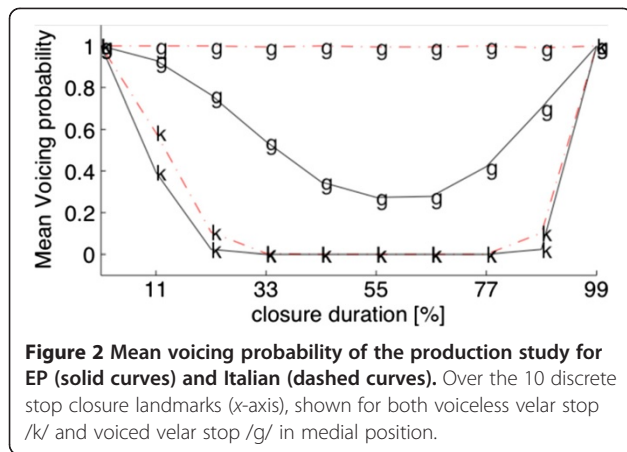
The GLMM logit models computed on the devoicing behaviour of the phonologically voiced velar stop showed that the effect of language is significant for the three central landmarks of the stop closure ($z = 7.9$, $z = 7.3$, $z = 6.4$; all, $p < 0.001$) with higher voicing for Italian than EP. The effect of vowel context was also significant ($z = 2.3$, $z = 4.5$, $z = 3.6$, $p < 0.01$), but only when comparing the open vowel /a/ to the close/close-mid vowels /i e/, with lower voicing for the open vowel. There was no significant effect between initial and medial consonant position. No interaction between the significant factors was found. Figure 2 presents the complete voicing profiles over all EP and Italian speakers for both voiceless and voiced velar stop in medial position.

When comparing the two languages, it can be seen that, throughout the complete stop closure, both the EP and Italian voiceless consonants /k/ show nearly identical extinction of voicing. In contrast, the phonologically voiced velar stop /g/ shows for EP a substantial devoicing throughout the complete stop closure, whereas the Italian /g/ maintains voicing, i.e. no devoicing is present for Italian. The observations throughout the complete stop closure confirm the statistical results performed on the central parts of the closure. Figure 3 shows the interactions between the factors *language* and *contextual vowel*: While for Italian no effect of contextual vowel height can be observed (i.e. no devoicing), EP shows higher devoicing for /a/ as compared to /i e/. Further,

Table 2 Durations in medial stop position (all values in ms)

|	EP	Italian
/g/		
Preceding vowel duration	133 (26)	192 (46)
Velar stop duration	97 (19)	103 (15)
/k/		
Preceding vowel duration	72 (22)	151 (36)
Velar stop duration	153 (22)	146 (28)

Mean durations (standard deviation in brackets) over all EP and Italian speakers (pooled over the three vowel contexts) in medial stop position (all values in ms).

Figure 2 Mean voicing probability of the production study for EP (solid curves) and Italian (dashed curves). Over the 10 discrete stop closure landmarks (*x*-axis), shown for both voiceless velar stop /k/ and voiced velar stop /g/ in medial position.

the observed pattern of overall higher voicing for Italian compared to EP (see Figure 2) is maintained in all four vowel contexts /i e o a/.

Perception study

Figure 4 shows the results of the perception experiment for the languages EP (top panels) and Italian (bottom panels). In Figure 4, the effects of the three factors (vowel duration in different panels, stop duration as different lines in each panel, voicing maintenance on the
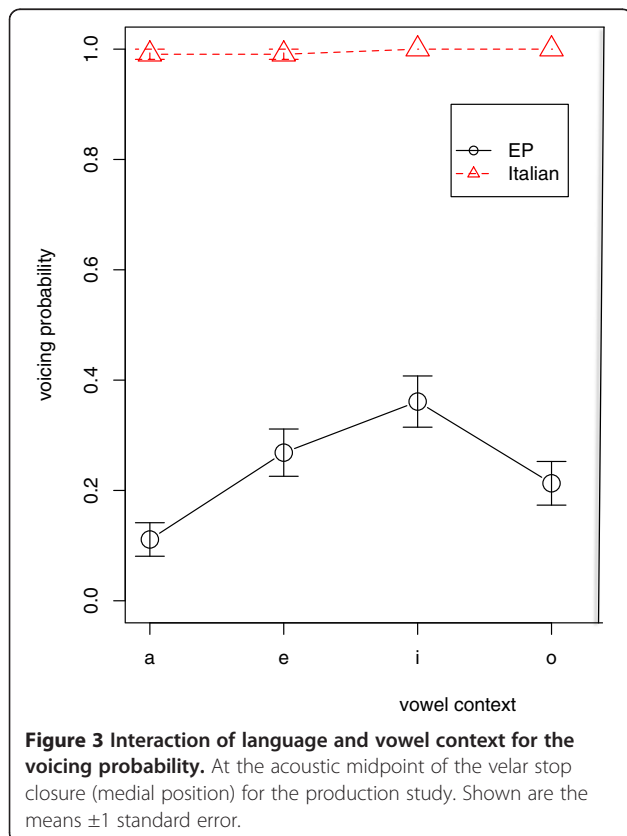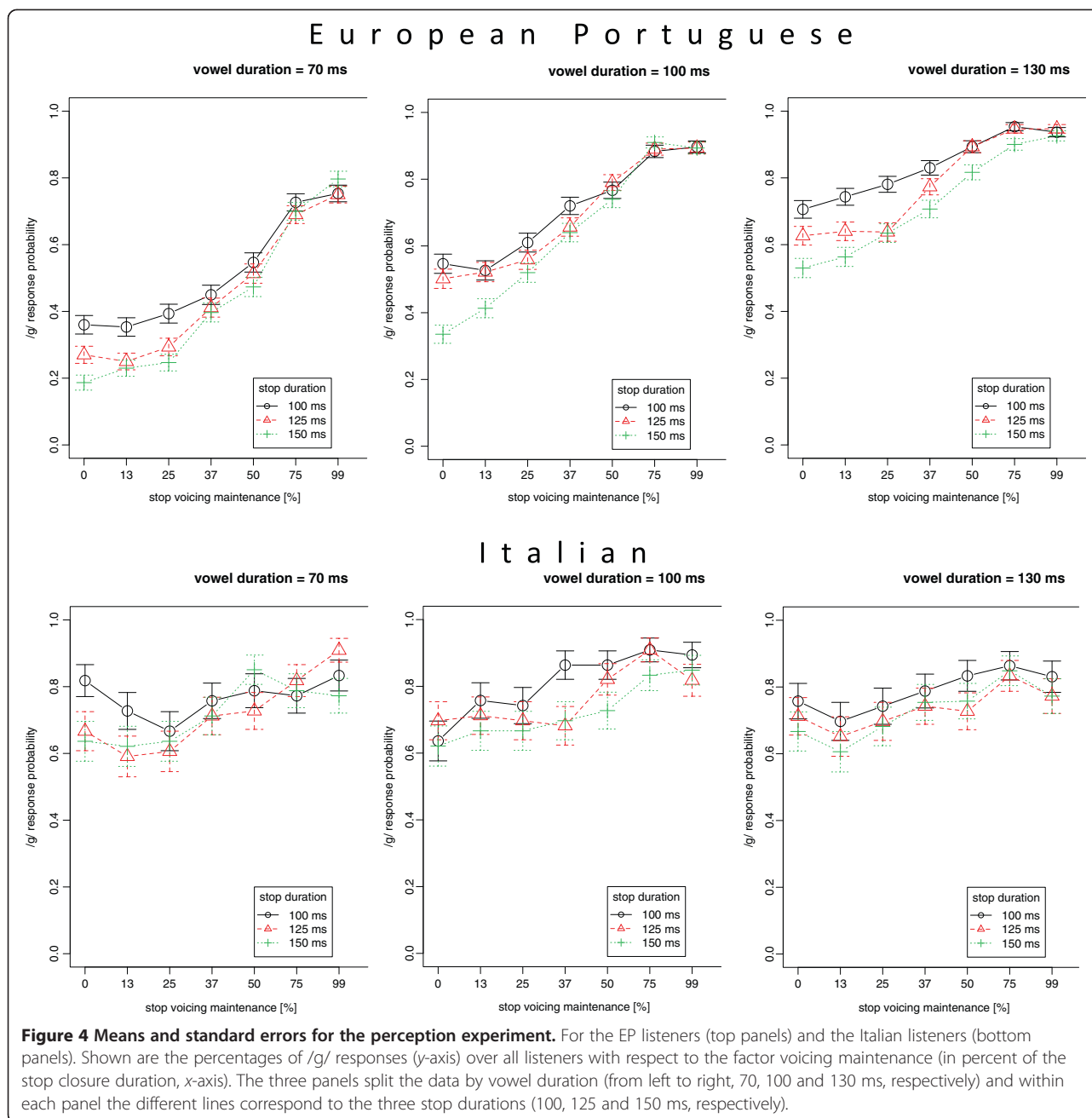
x-axis) on the mean percentage of listeners' voiced /g/ responses over all listeners are shown.

For EP, two effects can be observed for the factor voicing maintenance. First, there is a clear bias for listeners to prefer /g/ responses, with a lack of complete /k/ responses (i.e. a 0% to 20% *voiced* response probability), but a lack of 100% /g/ responses can also be observed. Second, increasing voicing maintenance leads to an increase of voiced responses for all vowel duration and stop duration conditions. Furthermore, increasing stop duration leads to a decrease in the probability of voiced responses but mainly for low voicing maintenance values (i.e. for highly devoiced or voiceless items). Increasing vowel duration (panels from left to right) results in an increase in the /g/ response probability, except for high voicing maintenance values. Increasing the voicing maintenance thus increases the /g/ response probability for all vowel durations and stop durations, but with different magnitudes. Stop duration and vowel duration have a strong influence on voicing decisions, but this effect is limited to low voicing maintenance values. For higher voicing maintenance percentages, a ceiling effect for all stop durations and vowel durations is observed. It seems that the voicing maintenance cue is very strong here and overrides the other two acoustic cues.

For Italian, there seems to be a complete lack of effect for the two factors vowel duration and stop duration on the listeners' voicing decisions. Instead, listeners show more or less a stable voicing probability of 0.7 (voiced /g/ responses), independent of the presented vowel duration and stop duration. The factor voicing maintenance has a small influence, much less than in EP and only clearly present for the second panel (vowel duration 100 ms). It is unlikely that the reason for the absence of the voicing maintenance effect is the longer preceding vowel duration in Italian (see Table 2), since in this case, the effect should be present and stronger in the third panel (vowel duration 130 ms).

In summary, Italian listeners are not influenced by the three examined factors and show a robust voicing decision towards a voiced /g/ response across all three varying factors. In contrast, EP listeners are strongly influenced by both vowel duration and voicing maintenance but less influenced by varying stop duration.

With the aim to isolate the effect of voicing maintenance from the other two acoustic cues, in the following, we define - for both languages separately - an *ambiguous prototype*, described by intermediate durational values between voiced and voiceless velar stops. Based on the values in Table 2, for the language EP, this prototype would be characterised by contextual vowel duration of 100 ms and stop duration of 125 ms. For Italian, the stop duration would be identical (125 ms), but the preceding vowel duration increases to the maximum

Figure 3 Interaction of language and vowel context for the voicing probability. At the acoustic midpoint of the velar stop closure (medial position) for the production study. Shown are the means ±1 standard error.

Figure 4 Means and standard errors for the perception experiment. For the EP listeners (top panels) and the Italian listeners (bottom panels). Shown are the percentages of /g/ responses (*y*-axis) over all listeners with respect to the factor voicing maintenance (in percent of the stop closure duration, *x*-axis). The three panels split the data by vowel duration (from left to right, 70, 100 and 130 ms, respectively) and within each panel the different lines correspond to the three stop durations (100, 125 and 150 ms, respectively).

duration examined in this study (i.e. 130 ms). For these ambiguous prototypes, the perceptual system is not able to use the acoustic cues vowel duration or stop duration to obtain a robust voiced/voiceless distinction and therefore has to rely only on the voicing maintenance cue (burst is not present in the stimuli). If we compare the two corresponding curves in Figure 4 (top middle panel dashed line for EP, bottom right panel dashed line for Italian), it can be seen that for stimuli with voicing maintenance higher than 50% of the stop closure, the two curves are nearly identical for the two languages, showing robust /g/ responses (with a probability of 0.8).

However, with increasing devoicing of the presented stimulus (i.e. for voicing maintenance below 50%), only EP participants show an effect on listener responses, while Italian listeners maintain their stable /g/ responses. Thus, in the absence of vowel duration and stop duration cues, only EP shows an influence of voicing maintenance, while Italian listeners decisions listeners are not affected. In other words, the two languages show substantial response differences but only for higher percentages of devoicing: the more voiced the stimuli are, the more similar the resulting voicing probabilities are across the two languages.

Statistical validation of the observed differences by means of GLMM showed that all main factors language ($z = 12.1$, $p < 0.001$), with Italians reporting more /g/s than EP listeners, contextual vowel duration ($z = 37.8$, $p < 0.001$), stop duration ($z = -10.0$, $p < 0.001$) and voicing maintenance ($z = 43.6$, $p < 0.001$) had a significant effect on the listeners' voicing decision, but not the factors contextual vowel identity and *repetition* number. The following interactions were significant: voicing maintenance with language ($z = -12.2$, $p < 0.001$); voicing maintenance with stop duration ($z = 5.7$, $p < 0.001$) with higher voicing maintenance differences for short stop durations; vowel duration with language ($z = -16.2$, $p < 0.001$); vowel duration with stop duration ($z = -2.34$, $p < 0.001$).

Conclusions
Production study
The Italian voicing profiles generated from velar stop data are similar to those published in [12]. Data presented in this paper shows significantly higher stop devoicing for EP compared to Italian. Comparing the voicing profiles and statistical results of EP not only against those of Shih et al. [12] for Italian and Spanish but also with our own Italian data, one can observe that EP does not behave like the other Romance languages Spanish and Italian with regard to devoicing of phonologically voiced stops. This result, differentiating EP from Italian stop voicing, is backed up by the linear mixed models analysis of voicing status at the central part of the stop closure, with significant differences between Italian and EP.

A possible explanation for the devoicing differences in production could be related to the fact that, contrary to other Romance languages (e.g. Spanish and Italian), EP is classically reported [34] as being stress-timed (see however [35] for evidence that EP shows shared properties of both stress-timed and syllable-timed languages), with reduction and neutralisation characteristics similar to other stress-timed languages (e.g. German and English).

In stress-timed languages, segments between stresses have the tendency to undergo substantial changes, with strong effects on segments themselves (e.g. vowel centralisation, devoicing and reduction). In the light of speech economy theories, certain segments or acoustic cues are thus enhanced in stressed parts, thus forcing other acoustic cues to weaken. When devoicing of phonologically voiced stops occurs, it could be the case that the strengthening of other segments or cues is more important, thus leading to extinction of the (cost intensive) voicing during stop closure for phonologically voiced stops. As a result, the devoicing characteristics for the two languages examined in this study (stress-timed EP and syllable-timed Italian) differ, with EP behaving more like other stress-timed languages (e.g. German). This similarity of devoicing patterns comparing EP and German was shown for all of the phonologically voiced stops and fricatives in a cross-linguistic comparison study of devoicing [14].

Perception study
The perception experiment provides evidence for the significant role of the acoustic cues voicing maintenance, preceding vowel duration and stop duration for voiced versus voiceless identification for EP but not for Italian. Italian listeners are more or less insensitive to these acoustic cues. They show a stable voiced response for all of the presented stimuli in the fully crossed design. Thus, it seems that EP listeners are strongly influenced by the three acoustic cues, but not the Italian participants.

In this regard, the perceptual results for EP are comparable to those obtained for other languages, for example English [6]. However, the phonetic realisations of phonologically voiced stops in EP are often highly devoiced throughout the complete duration of their stop closure, as seen in the production study (Figure 2). From these production results, one could assume that for EP stop voicing perception, the acoustic cues vowel duration and/or stop durations are more important than voicing maintenance, or that all these factors interact in a complex manner that does not give voicing maintenance the strongest weight in the identification process. However, the perception results in Figure 4 show that voicing maintenance is the dominant cue *if* the presented stimuli are fully voiced and therefore the major cue triggering the voicing decision of EP listeners. If the presented stimuli are devoiced, however, other acoustic cues (vowel duration and stop duration) take over by triggering the voiced/voiceless listener responses based on the extracted phoneme durations. In other words, we encounter cue weighting among different acoustic cues in the perception of EP stop voicing. In summary, these observations support the hypothesis that voicing maintenance is a major but not a required cue for stop voicing perception in EP, indicating that there are strong interactions between the three observed acoustic cues. For Italian, no such cue weighting was found.

Another finding is the bias of all listeners towards voiced responses, as can be seen in Figure 4 even for the fully devoiced/voiceless condition and durational values prototypical of a voiceless stop. For example, over all listeners, the response probability never reaches a probability lower than 0.2 for EP, and for Italian, the probability does not drop below the 0.5 threshold, thus resulting in voiced responses from the Italian listeners only. In other words, none of the two languages showed a stable voiceless (/k/) response floor effect. This bias could be due to the missing burst in the presented stimuli and thus to the problem of extracting a stable VOT cue.

For Italian, this could explain the absence of voicing differences for all of the acoustic cues. Listeners actively use the burst, thus obtaining the pre-voicing cue (voicing with reference to the stop release). Since the burst was deliberately missing, it could be the case that Italian listeners were not able to extract this cue, thus leading to their consistent voiced response (overall 80% /g/ perception) throughout all of the three continua. VOT could be important to extract a voiceless response, thus pointing to the *short lag* versus *long lag* VOT used as a distinctive cue. However, examining the interplay and cue weighting of other acoustic cues was not the focus of this study; but based on other languages, burst and VOT have an important effect on stop voicing distinction [2,6,36]. Therefore, clearly, the results would have been different if the burst had been included among the perceptual constructs.

For EP however, the missing burst cannot explain the perception results. A large amount of EP voiced stops, and even voiceless stops, do not show a discernible burst [13,14], so in this case, it is not clear how the listeners would rely on a VOT cue in ambiguous conditions. The stimuli used in our perception experiments were designed to exclude the burst cue (i.e. to exclude both burst and voicing onset time cues) to examine the influence of voicing maintenance and durational cues on stop voicing distinction.

For EP, with the constraint of the missing burst, the results of the perception experiment show that the voicing maintenance cue is strongly used to distinguish voicing, in addition to and in combination with the vowel duration and stop duration cue. Even with the missing burst and in the absence of voicing during stop closure, EP listeners are able to make stable voiced/voiceless decisions based on vowel durations and stop durations only. However, if the stimuli to be judged are more or less fully voiced, then voicing maintenance is found to be the major cue. In this case, it overrides the vowel duration and stop duration cues and guarantees a stable voiced response of all EP listeners, even in the absence of a facilitating burst and with contradicting duration values. In summary, this study constitutes new evidence that, in the absence of a facilitating burst, multiple acoustic cues are used and combined with different cue weighting to obtain a stable stop voicing distinction for EP.

Author details
[1]Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal. [2]School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal.

References
1. N Cuartero, *Voicing Assimilation in Catalan and English*. Ph.D. Thesis (Universitat Autonoma de Barcelona, 2002)
2. M Jessen, *Phonetics and Phonology of Tense and Lax Obstruents in German* (John Benjamins Publishing, Amsterdam, 1999)
3. L Lisker, AS Abramson, The voicing dimension: Some experiments in comparative phonetics, in *Proceedings of the 6th International Congress of Phonetic Sciences* (Prague, 1967), pp. 563–567
4. PA Luce, J Charles-Luce, Contextual effects on vowel duration, closure duration, and the consonant vowel ratio in speech production. J Acoust Soc Am. **78**, 1949–1957 (1985)
5. MC Viana, *Étude de deux aspects du consonantisme du Portugais: fricatisation et dévoisement*. Ph.D. Thesis (Université des Sciences Humaines de Strasbourg, 1984)
6. L Lisker, "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. Lang Speech. **29**(1), 3–11 (1986)
7. A Francis, K Baldwin, C Nusbaum, Effects of training on attention to acoustic cues. Percept Psychophys. **62**, 1668–1680 (2000)
8. LJ Raphael, Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J Acoust Soc Am. **51**(4B), 1276–1303 (1972)
9. F Li, A Mennon, J Allen, A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. J Acoust Soc Am. **127**(4), 2599–2610 (2010)
10. PJ Bailey, Q Summerfield, Information in speech: observations on the perception of [s]-stop clusters. J Exp Psychol Hum Percept Perform. **6**, 536–563 (1980)
11. EN Oglesbee, *Multidimensional stop categorization in English, Spanish, Korean, Japanese, and Canadian French*. Ph.D. Thesis (Indiana University, 2008)
12. C Shih, B Möbius, B Narasimhan, Contextual effects on consonantal voicing profiles: a cross-linguistic study, in *Proceedings of the 14th International Congress of the Phonetic Sciences* (San Francisco, 1999), pp. 989–992
13. M Lousada, LMT Jesus, A Hall, Temporal acoustic correlates of the voicing contrast in European Portuguese stops. J Int Phon Assoc. **40**, 261–275 (2010)
14. D Pape, LMT Jesus, Devoicing of phonologically voiced obstruents: Is European Portuguese different from other Romance languages? in *Proceedings of the 17th International Congress of Phonetic Sciences* (Hong Kong, 2011), pp. 1566–1569
15. J Veloso, The role of consonantal duration and tenseness in the perception of voicing distinctions of Portuguese stops, in *Proceedings of the 6th International Congress of Phonetic Sciences* (Stockholm, 1995), pp. 266–269
16. MRD Martins, Vogais e consoantes do Português: Estatística de ocorrência, duração e intensidade. Boletim da Filologia **XXIV**(1–4), 1–11 (1975)
17. P Boersma, Praat, a system for doing phonetics by computer. Glot International. **5**(9/10), 341–345 (2001)
18. D Pape, LMT Jesus, P Perrier, Constructing physically realistic VCV stimuli for the perception of stop voicing in European Portuguese, in *Computational Processing of the Portuguese Language*, ed. by H Caseli, A Teixeira, A Villavicencio (Springer, Heidelberg, 2012), pp. 328–349
19. P Perrier, Y Payan, M Zandipour, J Perkell, Influences of tongue biomechanics on speech movements during the production of velar stop consonants: a modeling study. J Acoust Soc Am. **114**, 1582–1599 (2003)
20. BH Story, A parametric model of the vocal tract area function for vowel and consonant simulation. J Acoust Soc Am. **117**(5), 3231–3254 (2005)
21. BH Story, Vowel acoustics for speaking and singing. Acta Acustica. **90**(4), 629–640 (2004)
22. JR Flanagan, DJ Ostry, AG Feldman, Control of trajectory modifications in target-directed reaching. Journal of Motor Behaviour **25**, 140–152 (1993)

23. PL Gribble, DJ Ostry, Are complex control signals required for human arm movements? J Neurophysiol. **79**(3), 1409–1424 (1998)
24. PL Gribble, DJ Ostry, Compensation for loads during arm movements using equilibrium-point control. Exp Brain Res. **135**, 474–482 (2000)
25. P Perrier, S Fuchs, Speed-curvature relations in speech production challenge the one-third power law. J Neurophysiol. **100**, 1171–1183 (2008)
26. S Tasko, J Westbury, Defining and measuring speech movement events. J Speech Hear Res. **45**, 127–142 (2002)
27. LMT Jesus, C Shadle, A parametric study of the spectral characteristics of fricatives. J Phon. **30**, 437–464 (2002)
28. LMT Jesus, C Shadle et al., Devoicing measures of European Portuguese fricatives, in *Computational Processing of the Portuguese Language*, ed. by NJ Mamed, J Baptista, I Trancoso (Springer, Berlin, 2003), pp. 1–8
29. JM Hillenbrand, RT Gayvert, Open source software for experiment design and control. J Speech Lang Hear Res. **48**(1), 45–60 (2005)
30. D Bates, M Maechler, B Bolker, *lme4: Linear mixed-effects models using S4 classes, R package version 0.999375-42*, 2011
31. RDCT, "*R: A Language and Environment for Statistical Computing*," R Development Core Team (RDCT), R Foundation for Statistical Computing, 2010
32. H Quené, HVD Berg, Examples of mixed effects modelling with crossed random effects and with binomial data. J Mem Lang. **59**, 413–425 (2008)
33. HR Baayen, *Analyzing linguistic data: A practical introduction to statistics using R* (Cambridge University Press, Cambridge, 2008)
34. M Cruz-Ferreira, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (Cambridge University Press, Cambridge, 1999), pp. 126–130
35. S Frota, M Vigário, On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. Probus. **13**, 247–275 (2001)
36. R Lasky, A Srydal-Lasky, R Klien, VOT discrimination by four to six and a half month old infants from Spanish environments. J Exp Child Psychol. **20**, 215–225 (1975)