HAL
open science

Constructing Physically Realistic VCV Stimuli for the Perception of Stop Voicing in European Portuguese

Daniel Pape, Jesus Luis M.T., Pascal Perrier

To cite this version:

Daniel Pape, Jesus Luis M.T., Pascal Perrier. Constructing Physically Realistic VCV Stimuli for the Perception of Stop Voicing in European Portuguese. PROPOR 2012 - International Conference on Computational Processing of the Portuguese Language, Apr 2012, Coimbra, Portugal. pp.338-349. hal-00684224

HAL Id: hal-00684224

https://hal.science/hal-00684224v1

Submitted on 21 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constructing physically realistic VCV stimuli for the perception of stop voicing in European Portuguese

Daniel Pape[1], Luis M.T. Jesus[1,2], Pascal Perrier[3]

[1] Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal
[2] School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal
[3] DPC/Gipsa-Lab, UMR CNRS 5216, Grenoble INP, Grenoble, France
{danielpape, lmtj}@ua.pt;
{Pascal.Perrier}@gipsa-lab.grenoble-inp.fr

Abstract. In this book chapter we present the generation of physically realistic stimuli with a biomechanical speech production model, with the aim to produce perceptually appropriate VCV sets for the European Portuguese (EP) voicing distinction. The duration measures necessary for the biomechanical model were extracted from an extensive EP speech production database, recorded for this aim. The same database was used to generate realistic voicing extinction contours for the perceptual continuum. To assess the realistic accuracy of the biomechanically generated stimuli, we compared the biomechanical stimuli set to linear interpolation between articulatory targets, traditionally used for speech synthesis.

Keywords: biomechanical modelling, perceptual cues, cue weighting, European Portuguese, voicing perception

1 Introduction

The work outlined in this book chapter consists of perceptual stimuli modelling as part of a research project on the importance of *voicing maintenance* in both speech production and perception in European Portuguese (EP) compared to other languages. For velar stop perception, we used and compared extracted voicing patterns and durational values from real speech productions (see Pape & Jesus 2011) in a matched cross-linguistic speech perception study, with the aim to examine the actual use and interaction (cue weighting) of the perceptual cues vowel duration, consonant duration and voicing maintenance. The speech material generated for the perceptual experiments consisted of biomechanically modelled stimuli acoustically synthesized with a three mass vocal fold model. The biomechanical modelling has the main advantage that all obtained tongue movements, trajectories and phoneme targets are comparable to natural speech, but with the additional possibility to manipulate all important temporal and glottal source parameters while maintaining articulatory realism. In sum, 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.

1.1. Perceptual Cues for Stop Voicing

For speech production, phonological voicing distinction is defined as the presence or absence of vocal fold vibrations during consonant production (Jakobson et al. 1952). For speech perception, a stop voicing distinction is mainly based on Voice Onset Time (VOT) (Lisker & Abramson 1964, 1967). Cross-linguistic differences in voicing perception are captured by changes in the VOT boundaries, i.e., by the location of the identification boundaries between voicing categories on a VOT continuum (Hoonhorst et al. 2009). In languages with three voicing categories (voiced, voiceless and voiceless aspirated) the mean VOT boundaries are located around ±30ms. Two-category languages differ in the nature of their voicing categories. In languages with a voiceless aspirated contrast the boundary is at +30ms (Lisker & Abramson 1970), whereas for languages with voiced/voiceless contrast without aspiration the boundary is 0ms (for Spanish: Williams 1977; for French: Serniclaes 1987). Infants below six months of age raised in an English environment are sensitive to both VOT boundaries (±30ms, 0ms), although only the positive VOT boundary is phonological in English (Aslin et al. 1981) or the 0ms VOT in the other language (French: Hoonhorst et al. 2009; Spanish: Lasky et al. 1975).

VOT is one of the most dominant cues for characterising stops in a number of languages, but a number of additional perceptual cues are found to influence the perception of voicing: consonant and adjacent vowel duration (Luce & Charles-Luce, 1985; Jessen 1998; Cuartero 2002; Viana 1984), and loudness (Repp 1979), among others. These other cues distinguish voicing, in combination with VOT but also without VOT, i.e., when VOT is ambiguous or missing. Further, the literature shows (Morrison 2005; Escudero et al. 2009) that human perception does not rely only on a single perceptual cue, but rather on a combination of different cues to guarantee a stable and robust perceptual outcome. Taking into account the variety of perceptual cues for stop voicing, the question arises how different languages weight the available cue to achieve robust perception. However, few studies (Francis et al. 2000) attempted to study the simultaneous variation and cue weighting for stop voicing distinction, and (to our knowledge) no studies examined this cue weighting framework for stop voicing in a cross-linguistic context. The last point is of utmost importance when one takes into account speech production differences, for example, in voicing maintenance between different languages (see, e.g., Solé 2011 for Spanish vs. English, and Pape et al. (submitted) for EP vs. German and Italian). Given these cross-linguistic differences, the question arises whether these are also reflected in the perception, and if so, how these differences influence the cue weighting of the available cues.

1.2. Modelling Physically Realistic Perceptual Stimuli

One important point to take into account when designing a valid multidimensional cue weighting experiment are the transitions between the phoneme targets: For example, velar stops show a strong articulatory forward loop (see, e.g., Mooshammer et al. 1995), additionally the shape of the loops differs for voiceless and voiced consonantal targets (Brunner et al. 2011). Assuming that these loops could play a

perceptual role, in the design of multidimensional perceptual stimuli one preferably aims for the most realistic synthesis model, with the aim to not disregard possible important influences on the perceptual system. However, it is not well understood how the temporal changes of articulatory pattern and thus the resulting articulatory trajectories are processed by the perceptual system. Thus, the particular model that is used to define and build a multidimensional stimuli space for perceptual experiments should be able to generate realistic articulatory movements, while simultaneously allowing for independent parametrical control of perceptual parameters, such as closure duration, vowel duration and transition duration. That is, in order to generate adequate perceptual stimuli for a cross-linguistic study on voicing distinction it is important to take into account the time-varying articulatory changes while producing a realistic phoneme chain with adequately realistic transitions. The most realistic perceptual stimuli would consist of naturally recorded speech, but for this condition one cannot control for the presence or lack of possible perceptual cues, therefore introducing an unwanted perceptual bias into the experiment.

There is no shortage of different models for articulatory motivated synthesis (Maeda 1990; Teixeira et al. 2005; Birkholz et al. 2010). However, very few of these models can claim to respect adequate articulatory movements. As soon as articulatory realistic transitions between the targets are required, only biomechanical models are currently capable of producing the required physically realistic articulatory trajectories. Thus, the use of a biomechanical modelling is the best compromise to guarantee highly realistic perceptual stimuli and independently control the varying parameters. Taking all these considerations into account, we used the advanced 2D tongue model described in Perrier et al. (2003) for the generation of suitable stimuli.

According to the biomechanical model described in Payan & Perrier (1997) and the improved version in Perrier et al. (2003), the tongue trajectories are obtained by activating (and thus deforming) different tongue muscles (Posterior and Anterior Genioglossus, Hyoglossus, Styloglossus, Verticalis, and Inferior longitudinalis) in a 2D Finite Element biomechanical model of the vocal tract, controlled on a target-to-target basis. The necessary vocal tract contours were extracted from X-ray data. The target for each phoneme is specified as a set of motor commands for each muscle. The movements between targets are achieved by a constant time shift of the motor commands between successive target values (as described in Perrier et al. 1996), resulting in time-varying vocal tract shapes (i.e., one complete 2D vocal tract shape each sampling period). Following, each of the obtained mid-sagittal vocal tract shapes is then converted to a time-dependent area function (Perrier et al. 1992). Then, the area functions were acoustically synthesized with a reflection-type line analog of the vocal tract (Story et al., 2000). Vocal folds oscillations are generated and controlled with a numerical implementation of the three mass model designed by Story & Titze (1995) based on lumped-elements (Titze & Story, 2002).

The biomechanical model was chosen because it has been shown to accurately account for articulatory trajectory shaping (Perrier et al. 2003) and velocity profiles (Payan & Perrier 1997) in the mid-sagittal plane, all possibly important for consonant perception (Sussman et al. 1998, Perrier & Fuchs 2008). Further, the model respects the relations between curvature and speed, which is found to be important in the correct perception of movements in general (Viviani & Stucchi 1992), and also for articulatory movements and thus speech (Perrier & Fuchs 2008). As can be seen in

figure 3 for the movement of the different tongue nodes, the biomechanical model accurately accounts for the articulatory forward loops observed in natural velar stop productions. These articulatory loops – and thus the naturalness of the model – mark an important difference between simple kinematic modelling (i.e., interpolating between consecutive articulatory target positions) and realistic biomechanical models. As described, the obtained accuracy of the transitions in the biomechanical model is important for our aims of the perceptual experiments, requiring that the resulting stimuli being as realistic as possible, but with the simultaneous independent control of all important parameters.

For the generation of the tongue contours, the biomechanical model needs the phoneme sequence identity (e.g., /aka/), the holding phase of each phone (e.g., 100ms) and the transition time from one phone to the other. Since all values for the biomechanical model are defined at the muscle command level that means that, e.g., the required holding phase does not correspond to the acoustic duration of a phone, but rather to a combination of transition and holding time (see section 2.1).

2 Method

2.1. Parameterisation of the Biomechanical Tongue Model

The biomechanical model, as described in Perrier et al. (2003), can be controlled with two different approaches as the input to the generation of the tongue contours: Inverse Synthesis; directly by entering the corresponding *lambda* commands for the different muscles[1]. For the EP stimuli generation we used the Inverse Synthesis approach. This Inverse Synthesis is implemented as follows: (1) Convex target regions have been defined in the (F1,F2,F3) acoustic space to specify the spectral characteristics of the elementary sounds of EP ; (2) relations between motor commands and formant values have been learned in the form of a radial basis functions model based on a large number of simulations with the biomechanical model; (3) for a given sequence of phonemes, the Inverse Synthesis Model finds the sequence of motor commands that minimizes the global change in motor commands while making sure that the successive target regions are reached at with the right timing. For more details about the procedure see Perrier et al. (2005). It is optimised by the inclusion of real speech formant and position data and their variance, thus Inverse Synthesis guarantees correct articulatory targets, which were extracted from cross-linguistic EMMA data. The biomechanical model's Inverse Synthesis approach accepts the input of phoneme

[1] To obtain input of raw muscle data, the model accepts the input of *alpha* values for each target and muscle. These values contribute to specify muscle activation together with feedback information about current muscle length and length change rate. Muscle tissues stiffness is increased as a function of muscle activation. The more force is applied, the stiffer the muscle fiber remains. This possibility can be used to generate preferred articulatory targets and trajectories not already predefined, i.e., to force model to generate non-existing or impossible articulatory targets.

identity, phoneme duration and transition duration to generate from this information the resulting appropriate motor commands and corresponding muscle forces.

Figure 1 shows a block diagram illustrating the process of controlling the biomechanical model with the EP values extracted from the EP speech production database described in section 2.2. For our perceptual experiments in EP, the VCV stimuli were generated by the biomechanical model by defining the following parameters as the input:

- identity of the context vowels adjacent to the velar stop (e.g., /aka/ or /iki/);
- vowel duration of the preceding and following vowel (from 70ms to 130ms);
- stop closure duration (from 50ms to 150ms);
- voicing maintenance during closure (from fully voiced to fully devoiced).

We modelled the first three factors based on the durational values for EP as described in section 3.1. For this reason, we chose the vertices and step sizes of the parameters in accordance with the durational values shown in figure 2. Table 1 shows all parameters and the corresponding values used as orthogonal factors for the perceptual experiments.

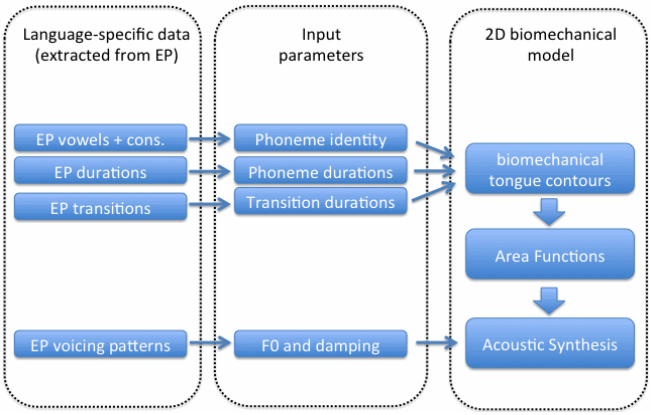

Fig. 1. Block diagram illustrating the process of integrating the EP production data into the biomechanical model.

Table 1. Factors and step size for the generation of the perceptual stimuli. The steps are determined in accordance with the durational values given in section 3.1.

steps	vowel duration (ms)	consonant duration (ms)	consonant voicing (%)
1	70	50	0 (fully devoiced)
2	100	75	12.5
3	130	100	25
4		125	37.5
5		150	50
6			75
7			100 (fully voiced)

While the parameters *vowel duration* and *consonant duration* are controlled during the synthesis of the biomechanical tongue contours (by selecting appropriate *t_hold* and *t_transition* values for consonants and vowels), the parameter *consonant voicing* (or *voicing maintenance*) is generated during the following acoustic synthesis of the stimuli with the Story et al. (2000) model (see section 3.3).

In sum, for each *vowel-stop-vowel* item 15 fully factorized stimuli are generated (3 *vowel_durations* paired with 5 *consonant_durations*), which then are acoustically synthesized with 7 different *consonant_voicing* curves. The phoneme-to-phoneme transition time is set to a standard value of 60ms obtained from the EP database.

2.2. EP Real Speech Corpus

To obtain the holding and transition phase as the input into the biomechanical model it was necessary to measure durational values in the acoustic space from real EP speech data. For this aim, we computed durational measures from an extensive speech production database (see Pape & Jesus 2011) generated for this purpose. The database was recorded for 4 EP speakers, consisting of /CV_1CV_2/ items in the frame sentence *Diga CVCV outra vez*, with all EP stops and fricatives /p b t d k g f v s z ʃ ʒ/ with the (identical) preceding and following (V_1,V_2) four vowel contexts /i e o a/. Thus, the consonants can be compared in initial (**C**V_1CV_2) and medial (CV_1**C**V_2) position. Each item was randomly repeated 9 times. Vowels and consonant boundaries are defined on base of the onset and offset of stable formant structure. We concentrate here on velar stops, the consonant modelled later in different contexts.

3 Results

3.1. Durational Measures from the EP Real Speech Corpus

Figure 2 shows the durational measures for the preceding vowel and velar stop in medial position. As can be seen, the consonant duration is higher for voiceless stops. Further, the vowel preceding the consonant is nearly twice as long for the voiced stop when compared to its voiceless counterpart. Thus, our EP data show clear differences for both preceding vowel and consonant duration between voiceless and voiced velar stops. The consonant durations are in line with EP durations found by Veloso (1995a, 1995b) and Delgado Martins (1975). With respect to the expected intrinsic vowel duration differences (with low vowels being longer than high vowels (Lehiste 1970)) our database confirms this tendency for the preceding vowel duration.

We computed an ANOVA with the duration measures as dependent variable and the consonant identity /k g/ as factor. For both the initial and the medial consonant position, the consonant duration was highly significantly (initial: $p < 0.001$; medial: $p < 0.001$). Further, the preceding vowel duration was highly significant for the medial consonant position ($p < 0.001$) but not for the initial position ($p = 0.54$). The duration

of the following vowel was not significant for the medial position (p = 0.35). For initial position, nothing can be concluded about the following vowel duration, since the following vowel for the initial position is the preceding vowel for the medial position, so it is not clear which of the consonants influences the durational cues.

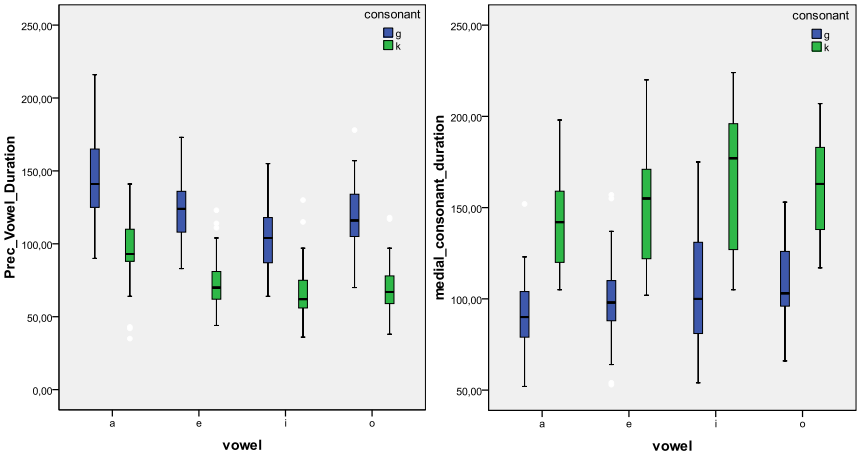

Fig. 2. Boxplots of the preceding vowel duration (left panel) and velar stop duration (right panel) for medial position (all values in milliseconds), with the neighbouring vowel context on the x-axis. The voiced stop is shown in darker shading, the voiceless stop in lighter shading. The voiced and voiceless character of each item is defined by its phonological status.

3.2. Tongue Contour Trajectories for the EP Stimuli

For the biomechanical EP VCV stimulus (velar consonant, 100ms vowel duration, 125ms consonant duration), figure 3 shows the trajectories of selected tongue nodes for the production of /aka eke iki oko/ stimuli. All four plots clearly confirm that the biomechanical model correctly reproduces the forward articulatory loop that can be observed for natural velar stop production (Mooshammer et al. 1995).

We compared our modelled biomechanical stimuli with target-to-target synthesis approaches. These approaches are the standard in articulatory synthesis. Examples can be found in the Maeda (1990) approach, the 3D articulatory synthesizer (Birkholz et al., 2010), the SpeechTrainer (Kröger 2003) and SAPWindows (Teixeira et al. 2005). For all these synthesis approaches, a number of articulatory targets are defined, and trajectories are achieved by a linear interpolation between targets. Figure 4 shows the illustration of the difference between several points on the tongue contour in the biomechanical model (in black) and targets for a target-to-target synthesis approach (dots in light colour). As can be seen in the right panel, the definition of one articulatory consonant target (here the highest point of the tongue contour) would results in very different articulatory tongue trajectories when comparing the biomechanical model (solid lines) and target-to-target approaches (dashed lines). It has to be noted that the differences between biomechanical and other synthesis approaches are less notorious when two consonantal targets are considered (e.g., the beginning and end of velar contact), however this approach is rarely used (Pape et al.

2011). Thus, the target-to-target approaches traditionally used for articulatory synthesis suffer from unrealistic modelling of the articulatory trajectories between targets, and are therefore unable to successfully model the articulatory loops seen in real speech articulatory data. Only the biomechanical model is able to reproduce these loops, as shown for all vowels and tongue nodes in the trajectories in figure 4.

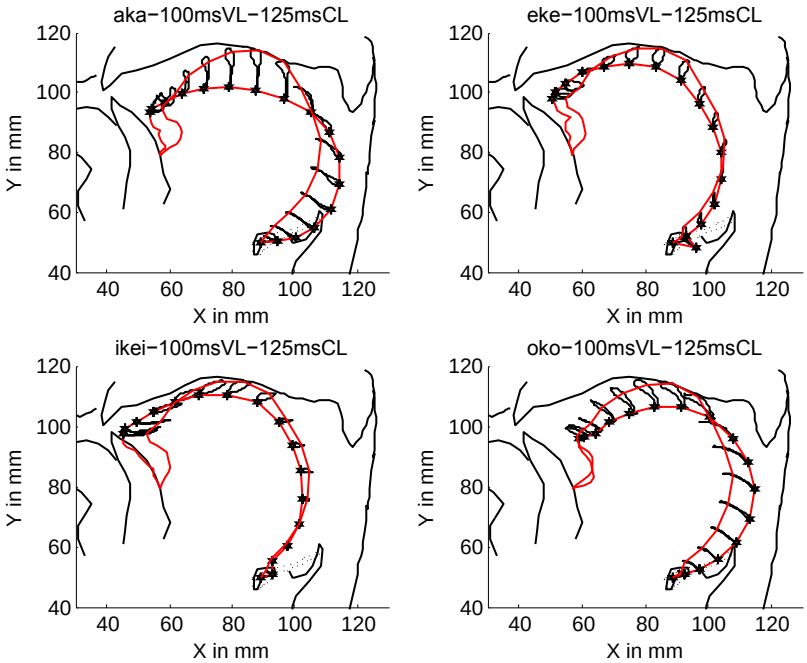

Fig. 3. Trajectories of selected tongue nodes for /VkV/ stimuli for the vowel contexts /a e i o/. The lines in light colour show the vowel and velar stop target positions.

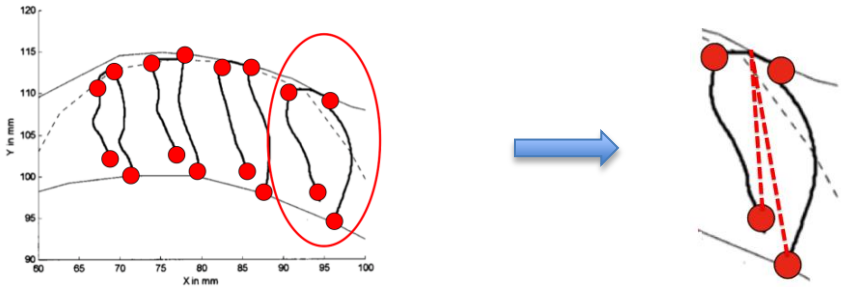

Fig. 4. The left panel shows four tongue surface trajectories for the /aka/ stimulus generated with the biomechanical model (in black). The dots in light colour exemplify the articulatory targets for a target-based synthesis. The right panel shows the differences in one tongue node trajectory between the biomechanical model (solid lines) and a given linear target-to-target trajectory (dashed lines, articulatory target at the highest point of the tongue contour).

3.3. Consonant Voicing Contours

To obtain the shapes of the velar stop consonant voicing curves we computed the mean curve over all repetitions and vowel contexts for each speaker of the EP database. Figure 5 shows the mean voicing probability estimate (9 repetitions, four vowels) computed for 10 consecutive equidistant landmarks throughout normalised consonant duration (see Pape & Jesus 2011). As can be seen, none of the voiced velar stops shows a linear decrease in voicing probability. Instead, the voicing decreases slower at the onset/offset of the curves, but faster during the mid.

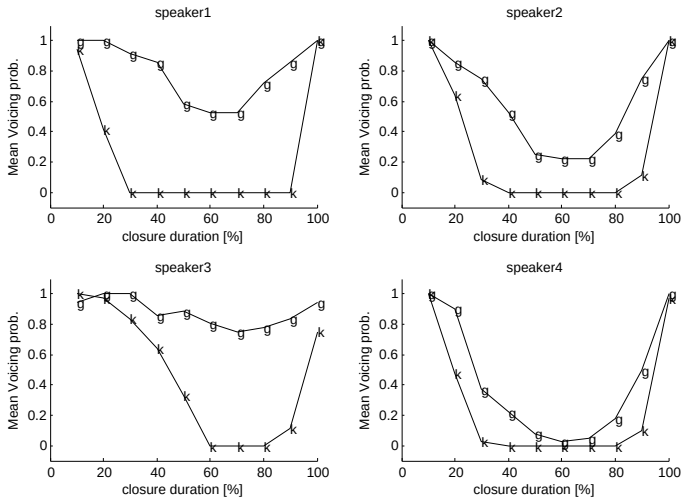

Fig. 5. Mean voicing probability curves (over all vowel contexts and repetitions) for the EP velar stops during the normalised stop closure. Each panel corresponds to one EP speaker of the database, each line represents the mean of 36 items (9 repetitions x 4 vowel contexts).

Due to this non-linear behaviour throughout the consonant closure, we chose to model our EP biomechanical stimuli by applying more realistic non-linear amplitude damping throughout the consonant closure. This was achieved by multiplying a weight to the consonant closure of the acoustically synthesized stimulus (i.e., the acoustic synthesis of the biomechanical tongue contours). In acoustic terms, this stimulus would correspond to the acoustic prototype of the /VgV/ biomechanical synthesis without amplitude damping due to the closure of the vocal tract and glottal damping (Pape et al. 2011). The obtained amplitude damping for the generation of the different consonant voicing contours was generated by applying a weighting function to the amplitude of the /VgV/ biomechanical stimulus. The shape of the weighting function was set as a cosine function ranging from 0 to π/2. The length of the cosine decrease was set to one quarter of the complete consonant duration. It was held identical for all *consonant_voicing* curves (steps given in table 1). The percentages of the *consonant_voicing* steps from table 1 define the landmarks where the cosine weighting reaches zero (π/2). Further, to avoid possible audible distortions due to jumps from full devoicing (zero) to full voicing amplitude we modelled the last 10ms of the consonant closure by weighting the fully voiced consonant with a sine function

ranging from 0 to π/2. Figure 6 shows the modelled normalised six different weighting curves of the *consonant_voicing* steps (125ms stop duration, f_s=48kHz). The procedure is illustrated for the 50% *consonant_voicing* curve (step5): At the consonant onset, no damping is applied to the acoustic synthesis output. Ranging from the first quarter until the consonant duration midpoint the amplitude is damped with a cosine weighting. The fully devoiced status is maintained until 10ms prior to consonant offset, where the 10ms sine-shaped increase to the consonant offset begins. The fully devoiced condition (step1), was modelled by a 5ms cosine decrease at consonant onset and identical 10ms sine rise at consonant offset. The right panels in figure 6 show oscillograms of the acoustic synthesis for four different *consonant_voicing* steps with the overlaid weighting function envelopes.

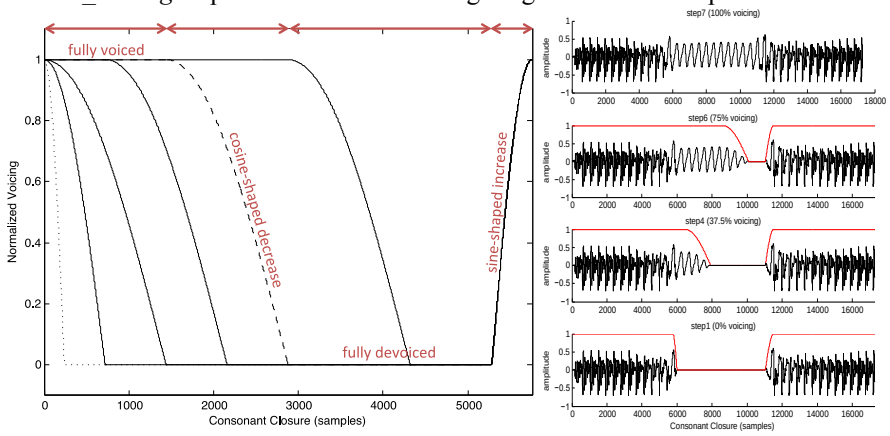

Fig. 6: The left panel shows modelled devoicing curves (step1 to step6 in table 1) for a 120ms stop closure. The voicing extinction curve for step5 (full devoicing at consonant mid) is shown as a dashed line, the fully devoiced condition (step1) is shown as a dotted line. The four different voicing statuses throughout the consonant closure for step5 are given in light colour, with arrows corresponding to the duration. The right panel shows the oscillograms for four steps of the consonant voicing continuum: top and bottom panel show the vertices (fully voiced and fully devoiced); second and third panel show the partly voiced conditions (second panel step6 = 75% voiced; third panel step4 = 37.5%voiced).

4. Conclusions

For a perceptual experiment examining European Portuguese (EP) stop voicing we modelled physically realistic stimuli by means of a biomechanical model. The input to the biomechanical model, i.e., phoneme identities, phoneme durations, transitions between phonemes and voicing extinction curves, were generated from an extensive EP real speech corpus. The resulting EP biomechanical tongue contours comply with articulatory and aerodynamic laws, e.g., articulatory loops observed for velar stops. The models (voicing curves) for the generation of different *devoicing* conditions were based on the voicing curves in the EP database. The extraction of the EP velar stops devoicing slopes showed a nonlinear behaviour in real speech.

Acknowledgements

This work was partially supported by the Portuguese Fundação para a Ciência e a Tecnologia (FCT), Portugal (grant SFRH/BPD/ 48002/2008).

References

Aslin, R., Pisoni, D., Hennessey, B., Perry, A.: Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. Child Development 52, 1135-1145 (1981)

Birkholz, P., Kröger B., Neuscheafer-Rube, C.: Articulatory synthesis and perception of plosive-vowel syllables with virtual consonant targets. In: Proc. Interspeech 2010, 1017-1020 (2010)

Brunner, J., Perrier, P., Fuchs, S.: Supralaryngeal control in Korean velar stops, Journal of Phonetics 39, 178-195 (2011)

Cuartero, N.: Voicing Assimilation in Catalan and English. Ph.D. Thesis, Universitat Autònoma de Barcelona, Barcelona, Spain (2002)

Delgado Martins, M.R. Vogais e consoantes do Português: estatística de ocorrência, duração e intensidade. Bol. de Filologia 14, 1-11 (1975)

Escudero, P., Benders, T., Lipski, S.: Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. Journal of Phonetics 37, 452-466 (2009)

Francis, A., Baldwin, K., Nusbaum, C.: Effects of training on attention to acoustic cues. Attention, Perception & Psychophysics 62, 1668-1680 (2000)

Fuchs, S., Perrier, P.: On the complex nature of speech kinematics. ZAS papers in Linguistics 42, 137-165 (2005)

Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., Serniclaes, W.: French native speakers in the making: From language-general to language-specific voicing boundaries, Journal of Experimental Child Psychology 104, 353-366 (2009)

Jakobson, R., Fant, C., Halle, M.: Preliminaries to Speech Analysis. MIT Press, Cambridge (1952)

Jessen, M.: Phonetics and phonology of tense and lax obstruents in German. John Benjamins, Amsterdam (1998)

Kröger, B.: Ein visuelles Modell der Artikulation. Laryngo-Rhino-Otologie 82, 402-407 (2003)

Lasky, R., Syrdal-Lasky, A., Klein, R.: VOT discrimination by four to six and a half month old infants from Spanish environments, Journal of Experimental Child Psychology 20, 215-225 (1975)

Lehiste, I.: Suprasegmentals. MIT Press, Cambridge (1970)

Lisker, L., Abramson, A.: Cross-language study of voicing in initial stops. Word 20(3), 384-422 (1964)

Lisker, L., Abramson, A.: Some effects of context on voice onset time in English stops. Language and Speech 10, 1-28 (1967)

Lisker, L.; Abramson, A.: The voicing dimension: Some experiments in comparative phonetics. In: Proc. ICPhS, pp. 563-567 (1970)

Luce, P., Charles-Luce, J.: Contextual effects on vowel duration, closure duration, and the consonant vowel ratio in speech production. JASA 78, 1949-1957 (1985)

Maeda, S.: Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: W. Hardcastle and A. Marchal (eds.) Speech production and speech modeling, 131-149 (1990)

Mooshammer, C., Hoole, P., Kühnert, B.: On loops. Journal of Phonetics 23, 3-21 (1995)
Morrison, G.: An appropriate metric for cue weighting in L2 speech perception: Response to Escudero & Boersma (2004). Studies in Second Language Acquisition 27 (4), 597-606 (2005)
Pape, D., Jesus, L.: Devoicing of phonologically voiced obstruents: Is European Portuguese different from other Romance languages?. In: Proc. ICPhS, pp. 1566-1569 (2011)
Pape, D., Jesus, L., Hall, A.: A cross-linguistic comparison of stop and fricative devoicing, Part I: Speech production. Journal of Phonetics (submitted)
Pape, D., Perrier, P., Fuchs, S., Kandel, S.: Les trajectoires formantiques respectant les lois de la physique contribuent-elles à une meilleure perception de la parole?. In: Actes JEP, (2010)
Pape, D., Perrier, P., Fuchs, S., Kandel, S.: Does physical realism of articulatory modeling improve the perception of synthetic speech?. In: Proc. ISSP (2011)
Payan, Y., Perrier, P.: Synthesis of V-V Sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. Speech Communication 22, 185-205 (1997)
Perrier, P., Boë, L., Sock R.: Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transition with two sets of coefficients. Journal of Speech and Hearing Research 35, 53-67 (1992)
Perrier, P., Fuchs, S.: Speed-curvature relations in speech production challenge the 1/3 power jaw. Journal of Neurophysiology 100, 1171-1183 (2008)
Perrier, P., Ma, L., Payan, Y.: Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. In: Proc. Interspeech, pp. 1041-1044 (2005)
Perrier, P., Payan, Y., Zandipour, M., Perkell, J.: Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. Journal of the Acoustical Society of America 114(3), 1582-1599 (2003)
Repp, B.: Relative Amplitude of Aspiration Noise as a Voicing Cue for Syllable-Initial Stop Consonants. Language and Speech 22 (29), 173-189 (1979)
Serniclaes, W.: Etude expérimentale de la perception du trait de voisement des occlusives du français. Ph.D. Thesis, Université Libre de Bruxelles, Brussels, Belgium (1987)
Solé, M.: Articulatory Adjustments in Initial Voiced Stops in Spanish, French and English. In: Proc. ICPhS, pp. 1878-1881 (2011)
Story, B., Laukkanen, A., Titze, I.: Acoustic impedance of an artificially lengthened and constricted vocal tract. Journal of Voice 14(4), 455-469 (2000)
Story, B., Titze, I.: Voice simulation with a body-cover model of the vocal folds. Journal of the Acoustical Society of America 97, 1249-1260 (1995)
Sussman, H., Fruchter, D., Hilbert, J., Sirosch, J.: Linear correlates in the speech signal: The orderly output constraint. Behavioral and Brain Sciences 21, 241-299 (1998)
Teixeira, A., Martinez, L., Silva, O., Jesus, L., Principe, J., Vaz, F.: Simulation of human speech production applied to the study and synthesis of European Portuguese, EURASIP Journal on Applied Signal Processing 9, 1435-1448 (2005)
Titze, I., Story, B.: Rules for controlling low-dimensional vocal fold models with muscle activation. Journal of the Acoustical Society of America 112, 1064-1076 (2002)
Veloso, J.: The role of consonantal duration and tenseness in the perception of voicing distinctions of Portuguese stops. In: Proc. ICPhS, pp. 266-269 (1995a)
Veloso, J.: Aspectos da percepção das "ocusivas fricatizadas" do português. Contributo para a compreensão do processamento de contrastes alofónicos. M.Sc. Thesis, Universidade do Porto, Porto, Portugal (1995b)
Viana, M.: Etude de deux aspects de consonantisme du portugais: fricatisation et devoisement. Ph.D. Thesis, LSHA Strasbourg, France (1984)
Viviani, P., Stucchi, N.: Biological movements look uniform: evidence of motor-perceptual interactions Journal Experimental Psychological Human Perceptual Performance 18, 603-623 (1992)
Williams, L.: The voicing contrast in Spanish. Journal of Phonetics 5,169-184 (1977)