Modeling sound merger in north metropolitan French with BiPhon-OT

Linguistics Bachelor's Thesis

University of Amsterdam

Tristan Chopinez - 13075098

Supervisor: Dr. Suki Yiu

26 / 06 / 2023

Acknowledgements

I am extremely grateful to Dr. Suki Yiu, who reached out to me and helped me understand the scope of the process as I was beginning. She trusted my ideas and guided me through my exploration.

I am also grateful to Cesko Voeten, who introduced me to some of the central topics of this thesis and encouraged my initial exploration of these subjects.

Thanks also to my classmates, who were always willing to review my work and give me feedback on my ideas and methods.

Finally, a special thank you to my parents, who believed in me with this project and all others before it, and without whom none of this would have been possible.

Abstract

In this thesis, the phonological merger of two nasal vowels in north metropolitan French is simulated through a bidirectional phonetics and phonology model. Using optimality theory as a medium, the simulation accurately predicts the merger in north metropolitan French, as well as a different development in Quebecois and Belgium French. Native north metropolitan French speakers are tested with a word reading task and a minimal pair discrimination task in order to investigate additional predictions from the model. Results reveal that the model can correctly predict multi-stepped sound change, but also suggests that it cannot account for linguistic variation within individuals.

Table of Contents

1 - Introduction

North metropolitan French is a broad term referring to the colloquial dialects of French from the north of France. Like most dialects of the language, it contains a set of phonemic nasal vowels: vowels whose airflow passes through the nasal cavity as well as the oral, and which are meaningfully differentiated from vowels whose airflow passes only through the oral cavity. The purpose of this thesis is to use these unusual sounds as a means to examine and potentially challenge the functionality of existing sound processing models. Specifically, the intent is to work with a model which purports to simultaneously simulate sound production and perception within single individuals: Optimality Theoretic Bidirectional Phonology and Phonetics.

1.1 A modeling system

Optimality Theoretic Bidirectional Phonology and Phonetics (BiPhon-OT) is a model originally developed by Paul Boersma (2006), which uses several layers of bidirectional optimality theoretical interactions in order to simultaneously describe language perception and production in individuals. Specifically, Boersma (2006, 2007, 2008, 2009, 2011) focuses on the layers which determine the sound systems of speakers; developing a model which can take a series of morphemes (already organized in a grammatical manner) and turn them into a correct-to-language sound pattern, and just as well take in a sound pattern and break it down into its underlying morphemic structure. This model utilizes several layers of representations, 5 of which are related to sound processing. The top one is the morpheme layer, which connects the sound system to the rest of the grammar, and is part of the 'semantic representations', others of which are not relevant to sound systems. Underneath are the phonological representations. The first of these is the underlying form (represented as |sound|), which represents the most basic (underlying) phonological shape for a given assortment of morphemes. The second is the surface form (represented as /sound/), which takes into account overarching phonological rules. Underneath again are the phonetic representations. The first of these is the auditory form (represented as [[sound]]), which represents the target sound-wave of the form. The second phonetic (and final overall) representation is the articulatory form (represented as [sound]), which represents for an individual speaker the specific articulation of a form. Importantly, this last layer is present in production (top down) but not in perception (bottom up). The model considers that speakers aim for an ideal acoustic form which they must produce using articulatory means, and that listeners can ignore articulation since they receive input which is already in acoustic form.

This model, as I have described it, is the basis for BiPhon in general, and can be applied to neural network models with BiPhon-NN (Boersma, 2019; Boersma, Benders, & Seinhorst, 2020; Boersma, Chládková, Benders, 2022). In BiPhon-OT, optimality theory is used as the connector between these representations. Each layer is connected on each side to a group of constraints, which establish how one representation connects to the ones around it. For example, cue constraints determine in perception how the acoustic signal (acoustic form) can be broken down into a set of meaningful and discreet phones (surface form), and in productions they determine what sound signal (acoustic form) a phonologically established series of phones (surface form) should be transformed into.

BiPhon-OT has been used to successfully describe morpho-phonological phenomena. For example, van Leussen (2020) uses the model to describe how the French adjective 'bonM/bonneF' (*good*) can be produced starting with a single morpheme to multiple possible phonetic forms ([bɔn], [bɔ̃]), depending on the gender and phonological form of its head noun; and inversely how these forms can be perceived as being the same morpheme. Below (figure 1) are these different forms in their different stages of production/perception (excluding the articulatory form).

	Masculine, vowel-initial	Masculine, consonant-initial	Feminine, vowel-initial	Feminine, consonant-initial
Morphemes	bon-M; acteurM	bon-M; mariM	bon-F; amieF	bon-F; voitureF
Underlying Form	|bɔn+Ø#aktœʁ|	|bɔn+Ø#maʁi|	|bɔn+ə#ami|	|bɔn+ə#vwatyʁ|
Surface Form	/.bɔ.nak.tœʁ./	/.bɔ̃.ma.ʁi./	/.bɔ.nə.a.mi./	/.bɔ.nə.vwa.tyʁ./
Acoustic Form	[[bɔnaktœʁ]]	[[bɔ̃maʁi]]	[[bɔnami]]	[[bɔnvwatyʁ]]

Figure 1: the different forms of the French adjective 'bon/bonne' and their different representation levels, as described by van Leussen (2020).

To highlight how successful the model is, it should be pointed out that the feminine vowel-initial form was not part of van Leussen's work, but has been derived here successfully and accurately through its several stages of representation using the same layered constraint sets with which the other forms were derived in the original paper.

1.2 BiPhon-OT and sound change

BiPhon-OT is a functional model to describe a stable sound system within an individual, but it can perhaps go beyond that by showing how the formation of a grammar within an individual - and specifically the interaction of the output from adult speakers with the developing grammar of first language acquiring children - can lead to sound change. Hamann (2009) suggests that when acquiring a sound grammar for their L1, children must use constraints in such a manner that allows them to analyze and then replicate the input they receive. Infants therefore begin with receiving raw acoustic (auditory) data, and must create a system of cue constraints which interprets data into a functional surface form representation. This same process must then happen going up further layers of representation, as well as down to test the constraint sets regarding their functionality in production. Hamann (2009) argues that if a child interprets phonetic cues differently from what was intended by the adults providing input, then sound change can occur. She gives the examples of [ut] and [yt]. As a result of articulatory and sensorimotor constraints, older speakers of the Ohalaish language produce /ut/ with a low F2 within the vowel, and a high F2 at the transition from vowel to consonant. The /t/ creates a coarticulatory space within the acoustic output, in which F2 is raised. Older speakers interpret this raising as a cue marking a back rounded vowel followed by a coronal consonant, but younger speakers interpret instead the presence of these distinct F2 values throughout the duration of the vowel as it being a front rounded vowel /y/. From the same input, different cue interpretations (as codified through cue constraints) lead to different surface forms from one generation to the next. In this thesis, the functionality of such sound change modeling will be tested using a modern sound change in the nasal vowel system of north metropolitan French (NMF) as a case study.

1.3 The state of nasal vowels in NMF

Academic studies observing the details of nasal vowels in NMF are surprisingly uncommon. Standard French descriptions appear to be the basis for descriptions of the system in non-governmental sources, which then sometimes detail deviations in practice from the standard. Standard descriptions are therefore a good point to begin an exploration of the topic. According to l'Académie de Normandie, a French governmental publication, the French language contains 4 phonemic nasal vowels: /ɑ̃/ /ɔ̃/ /ɛ̃/ /œ̃/. This official description can be found reflected in learning materials, such as Anderson (1968). Such descriptions are, for the most part, standard, and do not pertain to a particular variety of the language. Academic studies offer an insight on the realization of these 4 vowels in the dialects. Delvaux, Metens & Soquet (2002), for example, use a case study of 4 Belgian native speakers to suggest that in modern speech some slightly different acoustic properties can be found for these vowels, but find nonetheless that the standard descriptions accurately represent a 4 way paradigm of one back vowel, one low vowel, and two front vowels differentiated by rounding. Nicholas, Fagyal & Carignan (2019), however, in a study comparing perception of nasal vowels between NMF and Quebec French (QF), make the claim that the former does not distinguish /œ̃/ from /ɛ̃/ and ignore /œ̃/ entirely as a result. This proposed discrepancy between NFM and other varieties indicates a sound change, either a merger or a split. Similarly, De Mareüil, Adda-Decker & Woehrling (2007) make the claim that NMF speakers realize the phoneme /œ̃/ identically to /ɛ̃/, in a study comparing the vowel systems of NMF and southern metropolitan French (SMF). They base this claim on Malécot & Lindsay (1976), who directly observe and describe this phenomenon among speakers from the Paris Region, finding almost exclusively 3 nasal vowel systems. They describe this as a merger of two of the historical vowels, which were once distinguished only by rounding (one was spread, the other rounded) but which in the NMF of Paris now both have neutral rounding. Their proposition is well supported by external evidence: the official description of the sounds, alongside the continuing existence of a 4 nasal vowel system in QF, SMF, Belgian French (BF), and possibly still in some NMF speakers; suggests that '/ɑ̃/ /ɔ̃/ /ɛ̃/ /œ̃/' is indeed the older system, and that the simultaneous existence of 3 nasal vowel systems in some speakers is a results of merger (diachronically 4 -> 3) rather than a split (3 -> 4). This fact is further indicated to by the usage of different standardized spellings for /œ̃/ (um, un, eun) and /ɛ̃/ (in, im, ain, ym, yn, aim) (Anderson, 1968; Malécot & Lindsay, 1976). All sources of evidence therefore suggest that the merger account must be correct, (and it is likely for this reason that multiple studies assert the merger as fact without themselves providing evidence for it). It is this merger which will be used as a case study in order to test the effectiveness of BiPhon-OT and cue constraint reanalysis as a source of sound change.

1.4 Research question

'Can generational reinterpretation of cues in BiPhon-OT, as described in the introduction, serve as a functional and accurate model of the merger of phonemes /ɛ̃/ and /œ̃/ in north metropolitan French?'

The following approach will be taken in order to answer this question: First a theoretical model will be developed to describe the merger using generational constraint recreation in BiPhon-OT. If this endeavor is successful, then the answer will be partially answered as we will have a functional model to describe the merger. In such a case, there will be a second experimental portion which will aim at collecting diverse data capturing different stages of merger. Such data would then be compared to the model's predictions, in order to test its accuracy.

2 - Theoretical basis for a model

2.1 Phonetic factors of sound change

The differentiating feature between /ɛ̃/ and /œ̃/ is one of rounding. Therefore, in order to model the merger of these sounds, it is necessary to investigate the causes of rounding loss. Hayes et al. (chapter 4, 2004) suggest that loss of rounding differentiation can be attributed to factors of height and backness. Specifically, they suggest that high vowels differentiate rounding more than low vowels (and that in some languages back vowels differentiate rounding more than front vowels, but that this effect is not as universal). They base this claim on two previous studies. Linker (1982) demonstrates in several languages (importantly French is among them) that the articulatory factors which create roundedness in vowels are more pronounced in high vowels than in low vowels to a significant degree. Stevens (1998) demonstrates that the effect of rounding is more acoustically distinctive in high rather than low vowels. Hayes et al. (chapter 4, 2004) suggest that this means that the perception of rounding in low vowels is weaker than the perception of rounding in high vowels, as is supported by a study by Terbeek (1977). This last finding has been suggested to be the case specifically in French by Hirsch et al. (2003) who show that there is a stronger anticipation of rounding in the following vowel based on coarticulation at the end of a consonant in high vowels rather than low ones.

All of these findings correspond to Malécot & Lindsay (1976)'s study, which suggests that the merger between /ɛ̃/ and /œ̃/ is not a case of the elimination /œ̃/ in favor of /ɛ̃/, but rather a neutralization of both of their rounding features, such that /ɛ̃/ went from spread to unspread and /œ̃/ from rounded to unrounded. If it can be suggested that these vowels have both been undergoing lowering, then it would follow that rounding may become less distinctive and eventually become lost as a feature, causing the merger as Malécot & Lindsay (1976) observe it.

Indeed, there is good reason to believe that these two nasal vowels may be subject to forces which cause a lowering effect. Beddor, Krakow & Goldstein (1986) (henceforth, Beddor et al.) describe this fact using a description of the unique acoustic properties of nasal vowels. They, relying largely on previous studies (Fant, 1960; Fujimura & Linqdvist, 1971), suggest that a nasal vowel is marked (among other things) by a split in its oral equivalent's F1 as a consequence of the interaction between the nasal and oral tracts. Nasal vowels feature what Beddor et al. refer to as the FN and the F1'. The F1' is closest to a regular F1, having a somewhat higher frequency than in oral counterparts, a somewhat dampened amplitude peak, and its overall amplitude being somewhat more spread out over the surrounding frequency region. The FN is a new nasal formant, with a much lower amplitude peak, and whose position is importantly not as dependent on the oral qualities of the vowel (height, backness, rounding) as is the F1'. This fact is relevant to the acoustic shape of nasal vowels, because it means that depending on the particular value of the F1' (which, as in oral vowel F1s, is affected by height), either the F1' or the FN may be the lowest frequency peak in the vowel. High vowels have relatively low F1 values and therefore the F1' in a high nasal vowel is its true first formant (for the sake of clarity, let's call this lowest formant the LF). Low vowels have high F1 values and therefore the FN in a low nasal vowel is its LF. Beddor et al. (1986) give examples of these, using synthetic oral and nasal vowels (figure 2).

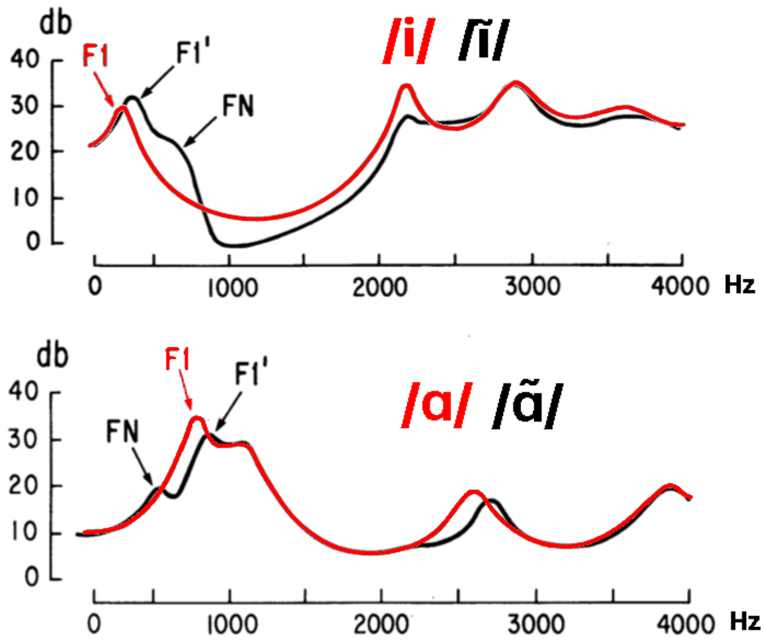

Figure 2: edited by the author, based on figures found in Beddor et al. It depicts synthetic oral (red) and nasal (black) vowels. High vowels /i/ are at the top, low vowels /ɑ/ are at the bottom.

A pattern emerges when comparing oral and nasal vowels which share their other qualities. In the high vowels, the oral F1 is lower than either of the nasal vowel formants. In low vowels, the oral F1 is lower than the nasal F1', but higher than the nasal FN. Beddor et al. connect this to experiments which show that speakers of languages without phonemic nasal vowels perceive nasalized vowels in non-nasalizing contexts as either more central (Wright 1980) or lower (their own experiments) than they are. They further connect all of this to a cross-linguistic trend of nasal vowel phonemes either becoming lower (e.g. Maithili, Portuguese, Shiriana, Yuchi) or more central (e.g. Breton, Nama, Seneca) over time (which effect is being observed can be ambiguous if it is high vowels which lower to a more central position: e.g. Bengali, Ewe). It is significant that 2 simultaneous effects seem to exist, as it reveals two possible interpretations of acoustic input leading to nasal vowel height change. Understanding this allows for the development of a model which can account for these different interpretations, and in the case of NMF, predict a particular one, as opposed to the other.

In oral vowels, vowel height is inversely connected to F1 value, such that a high F1 will lead to interpretation of a vowel a low, and a low F1 will lead to interpretation of a vowel as high. Nasal vowels force a complication of this system, by having two initial formants. Given the FN's attenuated amplitude, it is clearly distinct in quality from F1s in oral vowels. The F1' therefore is, regardless of position, the formant who's quality most closely matches that of the F1s in oral vowels. Here, it becomes necessary to define *F1* in a manner different to usual phonetic tradition. F1 will now refer to the first significant peak amplitude, as opposed to the first overall peak which this paper refers to as LF. In oral vowels, the F1 and the LF are the same, and this distinction is irrelevant. In nasal vowels, the F1 and LF are the same only if the F1' has a lower frequency than the FN. Otherwise the F1' is the F1 and the FN is the LF. Differentiating F1 and LF begs the question: which of these two is interpreted by listeners as an indicator of vowel height?

If the F1 is interpreted as the marker of vowel height, then the F1' in nasal vowels is always interpreted as the indicator of height. Due to the slight F1 frequency raising effect of nasalization, listeners with such cue interpretation will tend to hear nasal vowels as lower than their oral counterparts. This would cause the lowering effect found experimentally and cross-linguistically. Such an interpretation, therefore, is what must be hypothesized in order to predict the lowering of /ɛ̃/ and /œ̃/ in (older forms of) NMF.

If the LF is interpreted as the marker of vowel height, then which nasal vowel formant is interpreted as the marker of height depends on which has the lower frequency. When the F1' is lower, then it is the marker of height, and much as in the 'F1 = height' interpretation, the vowel will be interpreted as lower than its oral counterpart by speakers. This will also be the case if the F1' and FN are of near identical values; or if the FN is just lower than the F1' and just higher than the equivalent oral vowel's F1. All of these occur in central and high vowels. When the nasal vowel is lower, having an FN lower than both its F1' and its oral equivalent's F1 (see figure 2 above), then the nasal vowel will be interpreted as higher than its oral equivalent. High vowels will tend to lower and low vowels will tend to raise, meaning that such an interpretation would cause the centralizing effect found experimentally and cross-linguistically. QF shows evidence of an effect which such an acoustic interpretation would cause, with many speakers raising /ɛ̃/ to something closer to /ẽ/ (Carignan, 2011).

2.2 Proposition

What I propose above is exactly what Delvaux, Metens & Soquet (2002) found. Although their study only featured 4 participants as a case study, and although those participants were Belgian French (not NMF) speakers who all differentiated the 2 vowels, the details of the realizations of these vowels for each of them were suggestive of the phenomenon I have described. Comparing the speakers' vowels to the expected values of these vowels, they found consistent articulatory lowering and accompanying F1 raising, and found that both vowels also had lowered F2 values (which can be explained by an articulatory backing accompanying the lowering). Importantly, they found that /ɛ̃/ had undergone more F2 lowering than /œ̃/, making the vowels more similar acoustically. This latter fact is what would be predicted if, alongside articulatory backing, /ɛ̃/ became less spread (causing F2 lowering, which adds to the F2 lowering caused by backing) and /œ̃/ became less rounded (causing F2 raising, which demishes the net effect of backing). As has been mentioned, this was found in BF rather than NMF, but there is reason to believe both dialects may be prone to the same general forces. In fact, the acoustic and articulatory phenomena might be the same in all dialects, but other dialects may have other factors which cause them to resist this. In SMF, as shown by De Mareüil, Adda-Decker & Woehrling (2007), there is a tendency for nasal vowels to denasalize, and for a nasal consonant to appear after it. This would shield it from the nasal vowel lowering effect, and therefore from loss of rounding differentiation. In QF, as suggested above (Carignan 2011), there is a centralizing effect in front nasal vowels, indicative of an LF=height interpretation of nasal acoustics, and which certainly would not allow for the rounding loss which can be observed in low vowels. For these reasons, although BF is not under investigation, and although evidence suggests that speakers do not yet merge the vowels, it may serve as a good parallel for the merger process in NMF for lack of better studies regarding NMF specifically. I therefore suggest in this paper that the phenomenon observed in BF by Delvaux, Metens & Soquet (2002) is also occurring in NMF, for certain speakers to the extent that the vowels cannot be differentiated.

2.3 Hypothesis

I hypothesize that an F1 dominant interpretation of acoustic cues of nasal vowels by L1 learners caused nasal vowel lowering, which in turn triggered articulatory backing and loss of F2 distinctiveness, causing merger for some.

My first prediction is that it is indeed possible to make a functional BiPhon-OT model, using such a theoretical framework, to describe the merger of /ɛ̃/ and /œ̃/. If this prediction is correct, then one or more other predictions can be formulated on the basis of the model regarding experimental results from investigating the merger.

3 - Creating a Model

In this section, the model will be presented in two parts. First, the rules of the model will be presented. Specifically, the notation system for relative formant values - alongside the articulatory states and phonological markers they represent - will be outlined. Several sets of constraints will be presented, with their roles described, and a simple formula set will be added to represent person-external acoustics. After these rules are presented, the model will be used in order to demonstrate how it is capable of predicting the merger on the theoretical grounds outlined in section 2.

3.1 Rules of the model

3.1.1 Formant value representation

For the sake of the model, formant values will be given basic numerical values, as follows:

	F2: 3	F2: 2	F2: 1
F1: 1	/ĩ/	/ĩ̧/	/ỹ/
F1: 2	/ɪ̃/	/ɪ̧̃/	/ʏ̃/
F1: 3	/ẽ/	/ȩ̃/	/ø̃/
F1: 4	/ẽ̞/	/ȩ̞̃/	/ø̞̃/
F1: 5	/ɛ̃/	/ɛ̧̃/	/œ̃/
F1: 6	/ɛ̞̃/	/ɛ̧̞̃/	/œ̞̃/
F1: 7	/æ̃/	/æ̧̃/	/ɶ̃/

Figure 3: modeled formant values paired with (front) vowel types

This is to render the model more universal, and to be able to describe shifts in formant values as discrete steps. In reality, of course, formant values are not this discrete, and can alter based on a particular individual. For the sake of the model, it is better here to work with an abstraction, and which represents in a more easily quantifiable manner real life formant values and shifts in these values.

It should be mentioned that a particular simplification has been made here, concerning F2 values. Due to articulatory pressures, low front vowels cannot be as fronted as high front vowels. This means that, in front vowels, a formant gradient can be found from high to low regarding F2 as well as F1. This factor is possible to be represented in the model. For the time being, however, it is better to act as if F2 only shifts according to rounding, since this will make the model more easily manageable (with fewer and simpler constraints and possible candidates) and is the only factor relevant to the sound change under investigation.

3.1.2 Constraints of the model

These constraints will be presented in order of use from a production point of view. For each constraint set, there may be several variants, and several orderings. These account for the variation in the interpretation of acoustic data. It is important to note that the following constraints are calibrated to oral vowel values. That is to say, the model is made such that (front) oral vowels can be described accurately without predicted height change. This is to match the theoretical framework established in section 2, wherein it is proposed that nasal vowel height change is caused by a mismatch between 'default' oral mechanics and more marked nasal mechanics.

Cue-constraints (connecting surface and acoustic representations).

These constraints will serve as the basis, firstly, of how a new generation interprets particular acoustic signals as particular phones in surface form. They will also, in return, describe how a phonological form can lead to a particular target acoustic form. The set is divided into 3 sections. The constraints within each section can take any order without affecting the results, however the sections themselves must be ordered.

F1=Height interpretation

[[F1: 1]] /high/	[[F1: 2]] /high-mid/	[[F1: 3]] /mid-high/	[[F1: 4]] /mid/	[[F1: 5]] /mid-low/	[[F1: 6]] /low-mid/	[[F1: 7+]] /low/

Figure 4: relevant F1-targeting cue constraints

Having the constraints associate featural height with F1 value is what would be expected from someone who associates F1 (as defined in section 2) rather than LF with vowel height. Ranking this group first will cause such an interpretation in practice. It can also be ranked second, if the first section is that containing F2 constraints. If a speaker's interpretation of height is based on LF, these constraints must be ordered last.

LF=Height interpretation

[[LF: 1]] /high/	[[LF: 2]] /high-mid/	[[LF: 3]] /mid-high/	[[LF: 4]] /mid/	[[LF: 5]] /mid-low/	[[LF: 6]] /low-mid/	[[LF: 7+]] /low/

Figure 5: relevant LF-targeting cue constraints

Having the constraints associate featural height with LF value is what would be expected from someone who associates LF (as defined in section 2) rather than F1 with vowel height. Ranking this group first will cause such an interpretation in practice. It can also be ranked second, if the first section is that containing F2 constraints. If a speaker's interpretation of height is based on F1, these constraints must be ordered last.

Rounding selection

[[F2: 1]]: /round/	[[F2: 2]]: /neutral/	[[F2: 3]]: /spread/

Figure 6: relevant F2-targeting cue constraints

These constraints ensure the correct interpretation or creation of rounding specificity. This section must either be ranked first or second, but never 3rd, regardless of which of the two established acoustic interpretations a given speaker has. In the examples of section 3.2, it is always ranked second.

It must be said that, in theory, constraints exist which will associate any acoustic signal to any phonological feature. Such constraints, other than the ones shown above, must be considered to be ranked lower than those described. This can be expected to happen systematically, since they would bring about a complete mismatch between the real acoustic values of sounds and their real articulation, and would therefore disrupt any functional phonological system if they were ranked highly enough to participate in candidate decision.

Acoustic constraints (internal to acoustic representation).

This is an additional constraint set, not part of the usual layers of Boersma's model (2006, 2007, 2008, 2009, 2011). I add it here as a purely perceptual constraint set. It serves as a standardization layer, taking bare acoustic input and transforming it into values suitable for processing into a phonological form. Without this layer, the lowering effect would happen without fail for speakers with an F1=height interpretation of nasal vowel acoustics. Since such an automatic process of sound change is not something for which I can give evidence, it is more reasonable to expect that such a correcting layer exists, but that in certain individuals this layer prioritizes faithfulness, and that it is such individuals which cause a lowering of nasal vowels, which can then gradually (rather than near-instantly) spread throughout a language community.

Faith: [[F1]] [[FN]]	[[FN]] [[F1: F1 - 1]]	[[FN]] [[LF: LF + 1]]	[[FN]] [[LF: LF -1]]

Figure 7: relevant acoustic constraints

The layer is such that all non-optimal candidates will be eliminated by whichever constraint comes first. If the constraint focused on F1 is placed first, then this formant will be reanalyzed such that any interpretation of the sound as being lower than it is will be corrected. This corrects any lowering effect which may otherwise occur. The constraints focused on LF have similar effects, but for the other interpretation of nasal acoustics. There are multiple of them due to the more varied effects of such an interpretation. In any of these cases, it can be expected that an L1 learner of the language would rank first whichever constraint allows for a phonologically and articulatory correct analysis of the nasal vowel height of previous generations. It can be expected, however, that occasionally this fails to occur and that the faithfulness constraint remains the default first of the set. This will allow for the subsequent interpretation of acoustic factors into a phonological form (through cue constraints, see below) to be offset, such that a sound change takes place. It would therefore be individuals who rank 'faith' first in this layer who are responsible for the sound change under investigation in this thesis.

Sensorimotor constraints (connecting acoustic and articulatory representations)

This constraint set is purely productive. It connects an idealized (target) acoustic form and details how this target may be achieved articulatorilly. In perception, one begins with an acoustic form and uses acoustic constraints to refine it and cue constraints to map it phonologically. Sensorimotor constraints are therefore never needed in perception.

In practice, despite their differing roles, these constraints work very similarly to the cue constraints above. The details of their orderings are less relevant, however, since (as will be shown in section 3.2) the ranking of F1 specific and LF specific constraints is only relevant in perception. They are presented here according to acoustic cue specification.

'F1=Height' interpretation

[[F1: 1]] [high]	[[F1: 2]] [high-mid]	[[F1: 3]] [mid-high]	[[F1: 4]] [mid]	[[F1: 5]] [mid-low]	[[F1: 6]] [low-mid]	[[F1: 7+]] [low]

Figure 8: relevant F1-targeting sensorimotor constraints

'LF=Height' interpretation

[[LF: 1]] [high]	[[LF: 2]] [high-mid]	[[LF: 3]] [mid-high]	[[LF: 4]] [mid]	[[LF: 5]] [mid-low]	[[LF: 6]] [low-mid]	[[LF: 7+]] [low]

Figure 9: relevant LF-targeting sensorimotor constraints

Rounding selection

[[F2: 1]]: [round]	[[F2: 2]]: [neutral]	[[F2: 3]]: [spread]

Figure 10: relevant F2-targeting sensorimotor constraints

Articulatory constraints (internal to articulatory representation).

This constraint set represents articulatory specificities which impede on intended articulation. In this model, this layer is the one responsible for rounding neutralization.

[open] *[rounded]	[open] *[spread]	Rounding: Faith

Figure 11: relevant articulatory constraints

The faithfulness constraint governs the mechanism, such that any constraint before it causes neutralization of its specified rounding type, but any constraint after it fails to do so. Particularities of the different rankings of this constraint set are discussed in section 3.2.3.

3.1.3 Modeling nasal formant shifting

The system, as mentioned above, largely assumes oral vowel mechanisms, since these are the most basic which would be found in French, and the model assumes that oral vowels would also undergo the same processing as nasal vowels. BiPhon-OT works in such a way that, in production, the acoustic representation is a representation of the real acoustic properties of the vowel that is produced by the articulatory layer. For an oral vowel, the target acoustics and resulting real acoustics are (largely) the same. For a nasal vowel, as has been described extensively by now, a shift takes place. A formula must be employed, outside of either the representation layers or OT constraint sets, to describe how a given articulation results in the particular formants that another person would perceive. The expected relation between formants and both height and rounding (as articulatory processes as well as phonologically distinctive features) have been established in section 3.1.1, and this expected relation can be noticed again by looking at the cue constraints and sensorimotor constraints in section 3.1.2.

$$F2 = F2$$

$$FN = 4$$

$$F1' = F1 + 1$$

This formula set does not correspond to any BiPhon-OT layer. This is because layers indicate speakers' internal layers of speech processing. However, this formula set has to do with the real acoustic result of a given articulatory process. This happens for any sound when working with a BiPhon model, but this is usually roughly equivalent to what is aimed at by the AF. In the case of nasal vowels, however, there is a deviation between what is aimed at by the AF and what is actually produced after the ArtF, and this is accounted for by this layer external formula set.

3.2 Application of the model

There are several aspects of the model which must be demonstrated here. First, the model needs to predict that /ɛ̃/ and /œ̃/ both (over the course of some generations) shift to /æ̞̃/. Second, the model needs to be such that a (specific) different ordering of cue constraints and sensorimotor constraints will predict that these vowels raise to /ẽ̞/ and /ø̞̃/ instead. Finally, this model needs to allow for sound stability (regardless of LF-F1 interpretation), such that predicted sound changes (both in height and rounding) are sporadic rather than systematic. The following subsection will be divided in two: the treatment of /ɛ̃/, and the treatment of /œ̃/. This first part will also be used to demonstrate how different ranking of constraints can lead to different outcomes.

3.2.1 /ɛ̃/ to /æ̞̃/

As outlined in section 2, this process begins with a lowering effect, brought about by the interpretation of the nasal F1 (F1') as being the acoustic cue for height. Therefore, the cue constraints must be ordered in such a manner that constraints referring to F1 value are higher ranked than constraints referring to LF value. The chart below includes for each constraint section only the ones whose SF specification matches the SF input, as other constraints would not eliminate any candidates (and therefore have no effect).

Tableau 1: *Generation 1 - Cue Constraints (production)*

/ɛ̃/	[[F1: 5]] /mid-low/	[[F2: 3]] /spread/	[[LF: 5]] /mid-low/
[[F1: 1]] [[F2: 1]]	*!	*	*
[[F1: 2]] [[F2: 1]]	*!	*	*
[[F1: 3]] [[F2: 1]]	*!	*	*
[[F1: 4]] [[F2: 1]]	*!	*	*
[[F1: 5]] [[F2: 1]]		*!	
[[F1: 6]] [[F2: 1]]	*!	*	*
[[F1: 7]] [[F2: 1]]	*!	*	*
[[F1: 1]] [[F2: 2]]	*!	*	*
[[F1: 2]] [[F2: 2]]	*!	*	*
[[F1: 3]] [[F2: 2]]	*!	*	*
[[F1: 4]] [[F2: 2]]	*!	*	*
[[F1: 5]] [[F2: 2]]		*!	
[[F1: 6]] [[F2: 2]]	*!	*	*
[[F1: 7]] [[F2: 2]]	*!	*	*
[[F1: 1]] [[F2: 3]]	*!		*
[[F1: 2]] [[F2: 3]]	*!		*
[[F1: 3]] [[F2: 3]]	*!		*
[[F1: 4]] [[F2: 3]]	*!		*
👉 [[F1: 5]] [[F2: 3]]			
[[F1: 6]] [[F2: 3]]	*!		*
[[F1: 7]] [[F2: 3]]	*!		*

One thing to note about the formatting of the tableaus in this paper, is that they break OT convention. It is usual not to shade in cells with "*!", however for the sake of clarity in certain tableaus (e.g. tableau 3) such cells will be shaded. The optimal candidate of this chart is the acoustic representation of /ɛ̃/. In continuing with production, this representation will be used as an input which, when applied through sensorimotor constraints, will derive an articulatory representation of the sound. It is important to note that F1 and LF specified cue constraints both eliminated the same candidates. This is always the case in production, and therefore it would have been possible to order them differently and obtain the same optimal candidate. However, as we will see below, the order becomes decisive in perception, and so although it may not impact the acoustic signal which the following generation (*generation 2*) will receive, it will impact how that signal is interpreted. As mentioned, the next step for *generation 1*'s production is to use sensorimotor constraints to derive an articulatory form.

Tableau 2: *Generation 1 - Sensorimotor Constraints*

[[F1: 5]] [[F2: 3]]	[[F1: 5]] [Open-mid]	[[F2: 3]]: [spread]	[[LF: 5]] [Open-mid]
[Close] [spread]	*!		*
[Mid-close] [spread]	*!		*
[Close-mid] [spread]	*!		*
[Mid] [spread]	*!		*
👉 [Open-mid] [spread]			
[Mid-Open] [spread]	*!		*
[Open] [spread]	*!		*
[Close] [neutral]	*!	*	*
[Mid-close] [neutral]	*!	*	*
[Close-mid] [neutral]	*!	*	*
[Mid] [neutral]	*!	*	*
[Open-mid] [neutral]		*!	
[Mid-Open] [neutral]	*!	*	*
[Open] [neutral]	*!	*	*
[Close] [round]	*!	*	*
[Mid-close] [round]	*!	*	*
[Close-mid] [round]	*!	*	*
[Mid] [round]	*!	*	*
[Open-mid] [round]		*!	
[Mid-Open] [round]	*!	*	*
[Open] [round]	*!	*	*

As mentioned above, the ordering of these constraints is not crucial to the outcome. This particular constraint set is not very relevant to sound change, nor does it offer notable language specific outcomes (in this particular model, this is not necessarily true for all BiPhon modeling). Its role is instead to ensure the functioning of appropriate articulatory functions, by highly ranking the constraints which make an appropriate association between acoustic cues and articulatory details. The word appropriate, in this context, means corresponding to the real world physical correspondence between articulation and speech signal. As mentioned in section 3.1.2, for the purposes of this presentation of the model the (low ranked) constraints which do not make an appropriate association are ignored, since they will never have a functional effect on the selection of an optimal candidate.

Below is the final production constraint set, the articulatory constraint set, which does not connect different representations but rather modifies the articulatory representation on the basis of biological constraints. In this model, it represents the articulatory difficulty of creating meaningful rounding differences when articulating a low vowel. *In generation 1*, the vowel is not yet low, and therefore only the faithfulness constraint has an effect on choosing the optimal candidate. This means that, in this generation, the ranking of these constraints does not affect the vowel of interest.

Tableau 3: *Generation 1 - Articulatory Constraints*

[open-mid] [spread]	[open] *[spread]	[open] *[rounded]	Rounding: Faith
[open-mid] [rounded]			*!
👉 [open-mid] [spread]			
[open-mid] [neutral]			*!

The final (optimal) articulation of /ɛ̃/ is therefore with an open-mid mouth and spread lips. For an oral vowel, this would result in the following acoustic properties being born out:

F1: 5, F2: 3

However, since the vowel is nasal, these properties must be adjusted. The acoustic properties which would be born out are therefore as follows:

FN: 4, F1: 6, F2: 3

It is these values which the next generation 2 hears. The first step of the perception process for these values to be interpreted through the acoustic layer. For now, the F1-affecting constraint is ranked first. This ranking will lead to a stable system.

Tableau 4: *Generation 2 (stable) - Acoustic Constraints*

[[F1: 6]] [[FN: 4]] [[F2: 3]]	[[FN]] [[F1: F1 - 1]]	Faith: [[F1]] [[FN]]	[[FN]] [[LF: LF + 1]]	[[FN]] [[LF: LF -1]]
[[F1: 6]] [[FN: 4]] [[F2: 3]]	*!		*	*
👉 [[F1: 5]] [[FN: 4]] [[F2: 3]]		*	*	*
[[F1: 6]] [[FN: 5]] [[F2: 3]]	*!	*		*
[[F1: 6]] [[FN: 3]] [[F2: 3]]	*!	*	*	

The result of this layer is therefore to anticipate that the F1 is of a higher value due to the vowel's nasality, and correct this for appropriate phonological interpretation. The formant values being interpreted are now better aligned with the expected values of equivalent oral vowels, making it possible for the model to correctly interpret the significance of these values. Tableau 1 shows how cue constraints derive [[F1: 5]] [[F2:3]] from /ɛ̃/, and the inverse is also true. We can therefore understand that Tableau 4 allows *generation 2* to perceive (and therefore produce) the vowel as did *generation 1*. Let us now see what happens if this constraint set instead has Faith ranked first.

Tableau 5: *Generation 2 (unstable) - Acoustic Constraints*

[[F1': 6]] [[FN: 4]] [[F2: 3]]	Faith: [[F1]] [[FN]]	[[FN]] [[F1: F1 - 1]]	[[FN]] [[LF: LF + 1]]	[[FN]] [[LF: LF -1]]
👉 [[F1': 6]] [[FN: 4]] [[F2: 3]]		*	*	*
[[F1': 5]] [[FN: 4]] [[F2: 3]]	*!		*	*
[[F1': 6]] [[FN: 5]] [[F2: 3]]	*!	*		*
[[F1': 6]] [[FN: 3]] [[F2: 3]]	*!	*	*	

We see that this ranking leads to an unaltered version of the input. Such an acoustic representation does not account for the shifting of F1 in nasal vowels, which means that the vowel will be perceived by *generation 2* as different from the manner in which *generation 1* represents the vowel. We see this in Tableau 6.

Tableau 6: *Generation 2 (lowering) - Cue Constraints (perception)*

[[F1: 6]] [[FN: 4]] [[F2: 3]]	[[F1: 6]] /low-mid/	[[F2: 3]] /spread/	[[LF: 4]] /mid/
/ĩ/	*!		*
/ɪ̃/	*!		*
/ẽ/	*!		*
/ẽ̞/	*!		
/ɛ̃/	*!		*
👉 /ɛ̞̃/			*
/æ̃/	*!		*
/ĩ̹/	*!	*	*
/ɪ̹̃/	*!	*	*
/ẽ̹/	*!	*	*
/ẽ̞̹/	*!	*	
/ɛ̹̃/	*!	*	*
/ɛ̞̹̃/		*!	*
/æ̹̃/	*!	*	*
/ỹ/	*!	*	*
/ʏ̃/	*!	*	*
/ø̃/	*!	*	*
/ø̞̃/	*!	*	
/œ̃/	*!	*	*
/œ̞̃/		*!	*
/ɶ̃/	*!	*	*

We can see that, without correction from the acoustic constraints, the model anticipates that *generation 1*'s /ɛ̃/ will have become lowered /ɛ̞̃/ in *generation 2*. The model, therefore, has accurately predicted the nasal vowel lowering effect, as expected in NMF. If this same process is repeated from generation 2 to generation 3, the model will predict another shift, from /ɛ̞̃/ to low /æ̃/. We will see below how by generation 4 /æ̃/ will have neutralized its rounding to become /æ̹̃/. Before this, however, let us examine how a different ranking of cue constraints leads to a different interpretation of acoustic input, such that /ɛ̃/ centralizes rather than lowering.

Tableau 7: *Generation 2 - Cue Constraints (perception)*

[[F1: 6]] [[FN: 4]] [[F2: 3]]	[[LF: 4]] /mid/	[[F2: 3]] /spread/	[[F1: 6]] /low-mid/
/ĩ/	*!		*
/ɪ̃/	*!		*
/ẽ/	*!		*
👉 /ẽ̞/			*
/ɛ̃/	*!		*
/ɛ̞̃/	*!		
/æ̃/	*!		*
/ĩ̜/	*!	*	*
/ɪ̜̃/	*!	*	*
/ẽ̜/	*!	*	*
/ẽ̞̜/		*!	*
/ɛ̜̃/	*!	*	*
/ɛ̞̜̃/	*!	*	
/æ̜̃/	*!	*	*
/ỹ/	*!	*	*
/ʏ̃/	*!	*	*
/ø̃/	*!	*	*
/ø̞̃/		*!	*
/œ̃/	*!	*	*
/œ̞̃/	*!	*	
/ɶ̃/	*!	*	*

A reranking of cue constraints, in perception, causes the same acoustic input to be interpreted rather differently. While in Tableau 6 we observe the lowering effect, in Tableau 7 we observe a centralizing effect from /ɛ̃/ to /ẽ̞/. This is the effect observed in QF, and demonstrates this model's ability to predict the different types of nasal vowel height shifts.

Coming back to the lowering effect, as found in NMF, we have seen that this model can predict generational nasal vowel lowering. We have observed /ɛ̃/ shift to /ɛ̞̃/ in one generation, and can assume that in one generation /ɛ̰̃/ will shift to /æ̃/. We can then verify that the model can predict a rounding neutralization, from /æ̃/ to /æ̹̃/.

The cue constraints and sensorimotor constraints for generation 3's production function as in Tableau 1 and 2, identifying /æ̃/ as a sound which must be articulated as [open] and [spread]. The articulatory constraints in this generation have a different effect, as seen in tableau 8.

Tableau 8: *Generation 3 - Articulatory constraints*

[Open] [spread]	[open] *[spread]	[open] *[rounded]	Rounding: Faith
[open] [rounded]		*!	*
[open] [spread]	*!		
👉 [open] [neutral]			*

Unlike in tableau 3, all constraints of this set have a function due to [open] articulation of the sound. Because Faith is ranked lowest, both rounded and spread variants are eliminated as candidates by the markedness constraints, which seek to prevent low vowels from rounded or spread. Of course, given that our initial vowel was spread, it is not relevant where the '[open] *[rounded]' constraint is placed, so long as '[open] *[spread]' is placed before Faith. We will see in section 3.2.1 that the ranking of Tableau 8 is necessary for a full merger, however.

This articulation, accounting for the nasalization caused formant shifts, results in the following acoustic properties:

F1: 8, FN: 4 , F2: 2

This, when interpreted by the next generation (*generation 4*), looks as follows:

Tableau 9: *Generation 4 - Acoustic Constraints*

[[F1': 6]] [[FN: 4]] [[F2: 3]]	Faith: [[F1]] [[FN]]	[[FN]] [[F1: F1 - 1]]	[[FN]] [[LF: LF + 1]]	[[FN]] [[LF: LF -1]]
👉 [[F1: 8]] [[FN: 4]] [[F2: 2]]		*	*	*
[[F1': 5]] [[FN: 4]] [[F2: 3]]	*!		*	*
[[F1': 6]] [[FN: 5]] [[F2: 3]]	*!	*		*
[[F1': 6]] [[FN: 3]] [[F2: 3]]	*!	*	*	

Tableau 10: *Generation 4 - Cue Constraints (production)*

[[F1: 8]] [[FN: 4]] [[F2: 2]]	[[F1: 7+]] /low/	[[F2: 2]] /neutral/	[[LF: 7+]] /low/
/ĩ/	*!	*	*
/ɪ̃/	*!	*	*
/ẽ/	*!	*	*
/ẽ̞/	*!	*	*
/ɛ̃/	*!	*	*
/ɛ̞̃/	*!	*	*
/æ̃/		*!	
/ĩ̹/	*!		*
/ɪ̹̃/	*!		*
/ẽ̹/	*!		*
/ẽ̞̹/	*!		*
/ɛ̹̃/	*!		*
/ɛ̞̹̃/	*!		*
👉 /æ̹̃/			
/ỹ/	*!	*	*
/ʏ̃/	*!	*	*
/ø̃/	*!	*	*
/ø̞̃/	*!	*	*
/œ̃/	*!	*	*
/œ̞̃/	*!	*	*
/ɶ̃/		*!	

The F1 constraints are such that it is accounted for that articulation can only go so low (physiologically speaking). Attempting to reach the F1 value here referred to as '8' causes the same result as '7', since the physical limits of articulation would have necessarily been discovered by L1 learners. More importantly, the shift in rounding means that *generation 4* interprets as /æ̹̃/ the phone which *generation 3* interpreted as /æ̃/. The model will have no further effect on the sound, less constraints be reranked, thus predicting increased stability in future generations. The model now has to explain the same process, but beginning with /œ̃/.

3.2.2 /œ̃/ to /æ̹̃/

The process of lowering of /œ̃/ is nearly identical to that of /ɛ̃/. The constraints associating F2 of value 3 with spread lips will not have an effect, and instead the constraints associating F2 of value 1 with rounded lips will have an effect; this effect will be in practice perfectly equivalent to what was seen above. The model predicts, by *generation 3*, that initial /œ̃/ will now be interpreted in its surface form as /ɶ̃/, just as initial /ɛ̃/ became /æ̃/. Starting from generation 3, therefore, let us examine how /ɶ̃/ has its rounding neutralized to /æ̹̃/. If, when using the same primary constraint rankings as explored in section 3.2.1, /ɶ̃/ shifts to /æ̹̃/, then the model will have predicted the sound merger observed in NMF.

Tableau 11 *Generation 3 - Cue Constraints (production)*

/œ̃/	[[F1: 7]] /low/	[[F2: 1]] /round/	[[F1: 7]] /low/
[[F1: 1]] [[F2: 1]]	*!		*
[[F1: 2]] [[F2: 1]]	*!		*
[[F1: 3]] [[F2: 1]]	*!		*
[[F1: 4]] [[F2: 1]]	*!		*
[[F1: 5]] [[F2: 1]]	*!		*
[[F1: 6]] [[F2: 1]]	*!		*
👉 [[F1: 7]] [[F2: 1]]			
[[F1: 1]] [[F2: 2]]	*!	*	*
[[F1: 2]] [[F2: 2]]	*!	*	*
[[F1: 3]] [[F2: 2]]	*!	*	*
[[F1: 4]] [[F2: 2]]	*!	*	*
[[F1: 5]] [[F2: 2]]	*!	*	*
[[F1: 6]] [[F2: 2]]	*!	*	*
[[F1: 7]] [[F2: 2]]		*!	
[[F1: 1]] [[F2: 3]]	*!	*	*
[[F1: 2]] [[F2: 3]]	*!	*	*
[[F1: 3]] [[F2: 3]]	*!	*	*
[[F1: 4]] [[F2: 3]]	*!	*	*
[[F1: 5]] [[F2: 3]]	*!	*	*
[[F1: 6]] [[F2: 3]]	*!	*	*
[[F1: 7]] [[F2: 3]]		*!	

Tableau 12: *Generation 3 - Sensorimotor Constraints*

[[F1: 7]] [[F2: 1]]	[[F1: 7]] [open]	[[F2: 1]] [round]	[[LF: 7]] [open]
[Close] [spread]	*!		
[Mid-close] [spread]	*!		
[Close-mid] [spread]	*!		
[Mid] [spread]	*!		
[Open-mid] [spread]	*!		
[Mid-Open] [spread]	*!		
[Open] [spread]		*!	
[Close] [neutral]	*!	*	
[Mid-close] [neutral]	*!	*	
[Close-mid] [neutral]	*!	*	
[Mid] [neutral]	*!	*	
[Open-mid] [neutral]	*!	*	
[Mid-Open] [neutral]	*!	*	
[Open] [neutral]		*!	
[Close] [round]	*!	*	
[Mid-close] [round]	*!	*	
[Close-mid] [round]	*!	*	
[Mid] [round]	*!	*	
[Open-mid] [round]	*!	*	
[Mid-Open] [round]	*!	*	
👉 [Open] [round]			

Tableau 11 and 12 offer nothing new in terms of modeling. They are included for the sake of completeness. Tableau 13 functions much like Tableau 8. The constraints are ranked such that any form of rounding (assuming an [open] articulation) is neutralized. A deeper demonstration of the articulatory constraint set and how it relates to sound merger will be presented in section 3.2.3.

Tableau 13: *Generation 3 - Articulatory Constraints*

[Open] [round]	[open] *[spread]	[open] *[rounded]	Rounding: Faith
[open] [rounded]		*!	
[open] [spread]	*!		*
👉 [open] [neutral]			*

This articulation, accounting for the nasalization-caused formant shifts, creates the following acoustic properties: F1: 8, FN: 4, F2: 2.

As above, the articulatory constraints are arranged here such that rounded and spread vowels both neutralize, if the vowel is low. It is important to note that the acoustic properties of this sound /œ̃/ are identical to that of /æ̃/ in this generation (3). This means that *generation 3*, while differentiating the sounds in perception, and while considering them phonologically different, produces them identically. This point is a principle prediction of this model. This, again, when interpreted by the next generation (*generation 4*), looks as follows:

Tableau 14: *Generation 4 - Acoustic constraints*

[[F1: 8]] [[FN: 4]] [[F2: 2]]	Faith: [[F1]] [[FN]]	[[FN]] [[F1: F1 - 1]]	[[FN]] [[LF: LF + 1]]	[[FN]] [[LF: LF -1]]
👉 [[F1: 8]] [[FN: 4]] [[F2: 2]]		*	*	*
[[F1: 7]] [[FN: 4]] [[F2: 2]]	*!		*	*
[[F1: 8]] [[FN: 5]] [[F2: 2]]	*!	*		*
[[F1: 8]] [[FN: 3]] [[F2: 2]]	*!	*	*	

Tableau 15: *Generation 4 - Cue constraints (perception)*

[[F1: 8]] [[FN: 4]] [[F2: 2]]	[[F1: 7+]] /low/	[[F2: 2]] /neutral/	[[LF: 4]] /mid/
/ĩ/	*!	*	*
/ɪ̃/	*!	*	*
/ẽ/	*!	*	*
/ẽ̞/	*!	*	
/ɛ̃/	*!	*	*
/ɛ̞̃/	*!	*	*
/æ̃/		*!	*
/ĩ̧/	*!		*
/ɪ̧̃/	*!		*
/ȩ̃/	*!		*
/ȩ̞̃/	*!		
/ɛ̧̃/	*!		*
/ɛ̧̞̃/	*!		*
👉 /æ̧̃/			*
/ỹ/	*!	*	*
/ʏ̃/	*!	*	*
/ø̃/	*!	*	*
/ø̞̃/	*!	*	
/œ̃/	*!	*	*
/œ̞̃/	*!	*	*
/ɶ̃/		*!	*

As with the initially spread variant, *generation 4* treats this rounded variant as phonologically /æ̧̃/. As with this initial spread variant, this sound has also found stability according to the model's predictions. Since both initial /ɛ̃/ and initial /œ̃/ have through generations become /æ̧̃/, and since both are now predicted to stay as such, we can say that the sounds have fully merged. In perception as well as in production, *generation 4* is unable to differentiate these two nasal sounds, having neutralized both of them in rounding, just as exemplified by Malécot & Lindsay (1976).

3.2.3 Details of the rounding merger

Before moving on to experimental testing of the model, let us investigate how the merger can happen either at once, or in stages. We have seen how the constraint ranking in figure 12 can lead both /æ̃/ and /œ̃/ to neutralize to /æ̹̃/ within the same generation.

[open] *[spread]	[open] *[rounded]	Rounding: Faith

Figure 12: relevant articulatory constraints, 1st ordering

The ranking in figure 13 will have the same effect.

[open] *[rounded]	[open] *[spread]	Rounding: Faith

Figure 13: relevant articulatory constraints, 2nd ordering

The rankings in figures 14 and 15 will prevent any neutralization and keep both vowels in their more conservative states. Placing the faithfulness constraint first prevents the lip articulation of any vowel from being altered in this layer. Such rankings mean that neutralization of rounding may not be systematic for speakers who produce the nasal vowels of interest as low.

Rounding: Faith	[open] *[spread]	[open] *[rounded]

Figure 14: relevant articulatory constraints, 3rd ordering

Rounding: Faith	[open] *[rounded]	[open] *[spread]

Figure 15: relevant articulatory constraints, 4th ordering

The two other possible constraint combinations lead to a more nuanced outcome. The first combination is used in Tableau 18 and 19. It leads to rounded vowels being neutralized, but allows spread vowels to remain spread. The second combination is used in Tableau 20 and 21. It leads to spread vowels being neutralized, but allows rounded vowels to remain spread.

Tableau 16: *Articulatory Constraints - 1st combination (rounded vowel)*

[open] [rounded]	[open] *[rounded]	Rounding: Faith	[open] *[spread]
[open] [rounded]	*!		
[open] [spread]		*	*!
👉 [open] [neutral]		*	

Tableau 17: *Articulatory Constraints - 1st combination (spread vowel)*

[open] [spread]	[open] *[rounded]	Rounding: Faith	[open] *[spread]
[open] [rounded]	*!	*	
👉 [open] [spread]			*
[open] [neutral]		*!	

Tableau 18: *Articulatory Constraints - 2nd combination (rounded vowel)*

[open] [rounded]	[open] *[spread]	Rounding: Faith	[open] *[rounded]
👉 [open] [rounded]			*
[open] [spread]	*!	*	
[open] [neutral]		*!	

Tableau 19: *Articulatory Constraints - 2nd combination (spread vowel)*

[open] [spread]	[open] *[spread]	Rounding: Faith	[open] *[rounded]
[open] [rounded]		*	*!
[open] [spread]	*!		
👉 [open] [neutral]		*	

One of these rankings can lead to reduction but not elimination of a speakers rounding contrast in production. The loss of rounding should therefore be understood as being both unsystematic and prone to graduality.

4 - Experimental Methodology

As per this study’s first prediction, it was possible to create a BiPhon-OT model of the vowel merger in NMF, taking into account theoretical explanations for this merger. A comprehensive discussion regarding the model will be found in section 6. For now, experimental methods can be used to investigate whether this model is accurate (as opposed to simply functional, as it has been shown to be). The model has produced two important predictions regarding the nature of the nasal vowel merger in NMF. The first is that there may be a generation of speakers who perceive the sounds as distinct, but produce them identically. In section 3.2, this would be *generation 3*. This same prediction entails that no speakers produce them distinctly but perceive them identically. The second hypothesis is that, for speakers who do not merge the vowels in production, the degree of differentiation between them will vary significantly from speaker to speaker. To investigate these hypotheses, two tests are conducted which determine general properties of production and perception of these vowels.

4.1 Participants

Participants are 15 (8F, 7M) native NMF speakers of 18 years or more. All grew up in metropolitan France, in one of the following regions: Hauts-de-France, Normandie, Ile-de-France, Grand-Est, Bretagne. No participant had any known speech, hearing, or otherwise language impairment.

4.2 Production

This basic production experiment was inspired by Delvaux, Metens & Soquet (2002), who used a similar method in order to examine the acoustic properties of Belgian French nasal vowels. The goal of this experiment is to, for each participant, find out whether the vowels are merged or unmerged in production. That is to ask: do participants say these vowels in a manner (statistically) different enough that they can be systematically differentiated? And if so, what is the size of that effect?

4.2.1 Stimuli

Stimuli consist of individual written words. 10 words contain the sound /ɛ̃/ (e.g. ‘main’) in their standard pronunciation, and 10 contain the sound /œ̃/ (e.g. ‘commun’). These words are chosen to, as much as possible, be in pairs such that the syllables containing the target vowels are minimally paired (e.g. ‘main’ and ‘commun’ have a minimally paired final syllable). This, however, could not be perfectly done without lowering the stimuli count to lower than is acceptable for this experiment. Due to the low frequency of words containing standard /œ̃/ in the language, the words with that sound were chosen first and the words with /ɛ̃/ were chosen as a match. 20 filler words - containing neither of these target phonemes in the standard - were added in, as to minimize the chance that participants identify what is being investigated while undergoing the test (e.g. ‘tour’). See appendix A for this experiment’s full stimuli list.

4.2.2 Procedure

Stimuli are shown on a screen that the participant is reading. They are presented one at a time, in a randomized order. Participants see a word, speak it, and then the next word is shown, and so on until all words have been spoken. This is entirely auditorily recorded (with the participant's consent). The filler stimuli are then deleted. With the remaining stimuli, the target vowel (the nasal vowel of interest) is isolated, and the rest of the word is deleted.

4.2.3 Analysis

This analysis will be done on a participant to participant basis. This means that a separate analysis will be conducted for the data of each participant, rather than one analysis including all participants’ data. Praat (2023) software is used for sound analysis. For each target word, I isolate the target vowel. For each participant, I split the vowels into two groups based on what the (historical) phoneme was. Given the unusual acoustic properties of nasal vowels, relying on software for accurate formant identification, particularly in the case of the lowest formants (F1’ and FN) can lead to issues. Malécot & Lindsay (1976) suggest that the F2 measure is less likely to be contaminated by nasal-oral track interaction, and therefore serves as a more accurate feature of analysis than other formants. As they point out, this is practical given that the vowels of interest in this analysis are differentiated acoustically by F2. Using a linear statistical test, the two groups are compared on the basis of their F2 values. If there is a significant effect between the two, then the given participant can be said not to merge these vowels in production. If there is no significant effect between the two groups, then the given participant can be said not to differentiate them, and therefore to merge them.

4.3 Perception

This minimal pair discrimination experiment was inspired by Nicholas, Fagyal & Carigan (2019), who used a similar method to investigate both cross-dialectal and inter-dialectal perception of 3 nasal vowels in NMF and QF. The goal of this experiment is to, for each participant, find out whether the vowels are merged or unmerged in perception. That is to ask: can participants consistently differentiate these vowels when listening to them? Or do they fail at identifying them correctly? Or do they fail at perceiving them as different altogether?

4.3.1 Stimuli

Stimuli consist of pairs of written fake-words and associated spoken versions. Specifically, for each written pair, two stimuli exist: one with the first word as the associated spoken word, and the other with the second word as the associated spoken word. Each word is monosyllabic. The words in each pair are a minimal pair, differentiated by having either /ɛ̃/ (e.g. 'tin' /tɛ̃/) or /œ̃/ (e.g. 'tun' /tœ̃/) as its vowel. There are 10 pairs, for a total of 20 stimuli. 1 pair consists of the sole vowels, 5 consist of a consonant followed by the vowel, and 4 consist of the vowel followed by a consonant. This choice is made to account for Gottfried (1984)'s finding that the lack or presence of a consonant after a vowel in metropolitan French can affect accuracy in its perception. For every word in each pair, there is a given written form and a target spoken form, to be recorded by the researcher. They are all recorded in an 'unmerger' manner, such that there is a systematic difference in rounding between /ɛ̃/ and /œ̃/ in these stimuli. There are also an additional 18 filler pairs (based on 3 minimally paired French phoneme pairs), for an additional 36 stimuli (e.g. 'sa' /sa/ & 'cha' /ʃa/). Given both the non-real nature of the words involved, as well as their monosyllabism, a greater proportion of filler to target stimuli was employed here as compared to the production task. See appendix B for this experiment's full stimuli list.

4.3.2 Procedure

Participants are shown both fake words in a given pair, and after 3 seconds one of the fake words recordings is played. Participants are instructed to choose which of the two they believe to have heard. They are told to take their time in answering, and if they ask, the word can be played again. Once they answer, another pair is shown. Every pair is shown twice, each time with a different recording played, such that every recorded stimulus is eventually played. The same pair cannot be shown again until at least two other pairs have been shown, as to minimize influence. Apart from this caveat, the ordering of the stimuli is randomized. For every target pair shown, it is recorded whether or not the participant gave the 'correct' (target) answer or not. Filler stimuli are ignored.

4.3.3 Analysis

For each participant, a comparison is drawn between their test (with a certain % of 'correct' answers) and a perfect test (with 100% 'correct' answers). If there is a significant relation, then the participant can be said to distinguish /ɛ̃/ and /œ̃/ in perception. If there is not, then the answers can be said to either be random or motivated by another factor, and the participant cannot be said to distinguish the phonemes.

5 - Experimental results

This section is divided into 4 parts. First, two examples are given of data processing for both the production (5.1) and perception tasks (5.2). In 5.3, the statistical effects in non-merging speakers' production data are compared. In 5.4, there is an overview of the existing patterns connecting the tasks, looking at if either of the possible partial merger patterns occur.

5.1 Production data analysis

In this section, the production data analysis of two participants is presented in depth. It is important that preliminary data analysis be done on a participant to participant basis, since merger (or lack thereof) exists on an individual basis. First, an analysis is shown for a speaker who does not merge the vowels. Second, an analysis is shown for a speaker who does merge the vowels. Since participants either do or do not merge them, the general details of both of these analyses apply to all participants. The specific numerical values are all unique to each participant, but the methodology is uniform.

5.1.1 A non-merging participant

/ɛ̃/	*F2*	/œ̃/	*F2*
Ain	1494 Hz	**Bungalow**	1441 Hz
Linge	1426 Hz	**Un**	1225 Hz
Quinze	1469 Hz	**Jungle**	1334 Hz
Simple	1359 Hz	**Aucun**	1576 Hz
Mannequin	1727 Hz	**Parfum**	1306 Hz
Bain	1544 Hz	**Humble**	1240 Hz
Main	1337 Hz	**Chacun**	1500 Hz
Brin	1472 Hz	**Brun**	1160 Hz
Cinglé	1606 Hz	**Commun**	1335 Hz
Faim	1437 Hz	**Lundi**	1405 Hz

Figure 16: F2 values of target vowels for a single participant

Figure 16 shows the values in Hertz of the second formants of the target nasal vowels for each target word in the production task. Applying a linear regression test allows us to compare the values of the /ɛ̃/ and /œ̃/ groups, and establish whether or not there is a systematic difference between these. It establishes the average difference in value between each group and estimates how well this average can predict the difference between any value from one group and any value from the other. The test is run on RStudio (2020), which is instructed to calculate the change in F2 value from intercept /ɛ̃/ to target /œ̃/. When applied to the data in Figure 16, a difference is found such that sounds from the /ɛ̃/ group are on average 134.9Hz (intercept estimate: 1487.10; intercept std. error: 38.86; target estimate: -134.90; target std. error: 54.95) higher than sounds from the /œ̃/ group. The statistical test also reveals a p-value of 0.0245, indicating that the difference between the two groups is statistically significant and can be an indication of systematic differentiation between the two groups. Therefore, we can conclude that the participant at the origin of this data seemingly produces the sounds /ɛ̃/ and /œ̃/ distinctly.

5.1.2 A merging participant

/ɛ̃/	*F2*	/œ̃/	*F2*
Ain	1329 Hz	**Bungalow**	1554 Hz
Linge	1238 Hz	**Un**	1377 Hz
Quinze	1781 Hz	**Jungle**	1504 Hz
Simple	1245 Hz	**Aucun**	1743 Hz
Mannequin	1552 Hz	**Parfum**	1220 Hz
Bain	1235 Hz	**Humble**	1144 Hz
Main	1554 Hz	**Chacun**	1796 Hz
Brin	1253 Hz	**Brun**	1299 Hz
Cinglé	1687 Hz	**Commun**	1453 Hz
Faim	1310 Hz	**Lundi**	1272 Hz

Figure 17: F2 values of target vowels for a single participant

Figure 17 shows the values in Hertz of the second formants of the target nasal vowels for each target word in the production task. Applying a linear regression test, as above, reveals that the /ɛ̃/ group has formant values on average 17.8Hz lower than the /œ̃/ group (intercept estimate:

1418.40; intercept std. error: 66.96; target estimate: 17.80; target std. error: 94.69). The test also indicates a p-value of 0.853, which means that the difference between the groups is not statistically significant. There is no systematic differentiation of the sounds and therefore, we can conclude that the participant at the origin of this data does not produce /ɛ̃/ and /œ̃/ as distinct sounds.

5.2 Perception data analysis

In this section, the perception data analysis of two participants is presented in depth. As with production, it is important that preliminary data analysis be done on a participant to participant basis, since merger (or lack thereof) exists on an individual basis. First, an analysis is shown for a speaker who does not merge the vowels. Second, an analysis is shown for a speaker who does merge the vowels. Since participants either do or do not merge them, the general details of both of these analyses apply to all participants. The specific numerical values are all unique to each participant, but the methodology is uniform.

5.2.1 A non-merging participant

Sound	Response	Sound	Response
/ɛ̃/	in	/œ̃/	un
/tɛ̃/	tin	/tœ̃/	tun
/sɛ̃/	sin	/sœ̃/	sun
/nɛ̃/	nin	/nœ̃/	nun
/lɛ̃/	lin	/lœ̃/	lun
/jɛ̃/	yun	/jœ̃/	yun
/ɛ̃t/	inte	/œ̃t/	unte
/ɛ̃s/	ince	/œ̃s/	unce
/ɛ̃l/	inle	/œ̃l/	unle
/ɛ̃k/	inque	/œ̃k/	inque

Figure 18: Chosen word in response to target sound stimuli for a single participant; green indicates correct sound-spelling matching, red indicates incorrect matching

Figure 18 separates the stimuli into two groups, those with /ɛ̃/ and those with /œ̃/. Each stimulus is paired with the participant's response. A logistic regression test is applied to compare the groups on the basis of their response types ('in' versus 'un'), such that it can be determined how likely the relation between stimuli and response is to be coincidental. The test is run on RStudio (2020), which is instructed to calculate the change in the odds of getting a 'in' response from intercept sound /ɛ̃/ to target sound /œ̃/. The statistical test reveals a p-value of 0.0032 (intercept estimate: -2.197; intercept std. error: 1.054; target estimate: 4.394; target std. error: 1.491), indicating that the difference between the two groups is statistically significant and is an indication of systematic differentiation between the two groups. We can therefore conclude that the participant at the origin of this data perceives the sounds /ɛ̃/ and /œ̃/ distinctly.

5.2.2 A merging participant

Sound	Response	Sound	Response
/ɛ̃/	in	/œ̃/	un
/tɛ̃/	tin	/tœ̃/	tun
/sɛ̃/	sin	/sœ̃/	sin
/nɛ̃/	nin	/nœ̃/	nin
/lɛ̃/	lun	/lœ̃/	lin
/jɛ̃/	yun	/jœ̃/	yun
/ɛ̃t/	inte	/œ̃t/	unte
/ɛ̃s/	unce	/œ̃s/	unce
/ɛ̃l/	unle	/œ̃l/	unle
/ɛ̃k/	inque	/œ̃k/	inque

Figure 19: Chosen word in response to target sound stimuli for a single participant; green indicates correct sound-spelling matching, red indicates incorrect matching

Figure 18 separates the stimuli into two groups, those with /ɛ̃/ and those with /œ̃/. Applying a logistic regression test, as above, reveals a p-value of 0.374 (intercept estimate: -0.4055; intercept std. error: 0.6455; target estimate: 0.8109; target std. error: 0.9129), indicating that the difference between the two groups is not statistically significant. There is no systematic differentiation of the sounds and therefore, we can conclude that the participant at the origin of this data does not perceive /ɛ̃/ and /œ̃/ as distinct sounds.

5.3 F2 Distance in production

The analyses in 5.1 and 5.2 can be applied to each participant. This could be used to test the prediction that participants will merge the vowels in production to differing degrees. However, doing so would require comparing the data of participants who do not merge the vowels in production, and there is only one such participant (see figure 20 below), which is not enough for statistical significance of data. The prediction can therefore not be satisfactorily verified nor disproven.

5.4 Relation between production and perception

Applying the analyses also gives us participant profiles regarding merger in production and perception (see figure 20 below), which can test the other prediction of this study. If there are participants who merge the vowels in production only, then the model will have made a correct prediction. If there are participants who merge the vowels in perception only, then the data will be such that the model is unable to predict it.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
Production	-	-	-	-	-	.	-	-	-	-	*	-	-	-	-
Perception	*	-	-	-	-	-	*	-	-	-	-	**	-	-	.

Figure 20: indication of merger (or lack thereof) in both production and perception for all participants; green indicates a significant effect (no merger), red indicates no significant effect (merger), orange indicates potential significance (we interpret this showing merger, but it is less certain than with red); the symbols represent the p-values for each test (- > 0.1 > . > 0.05 > * > 0.01 > ** > 0.001)

Green represents a statistically significant difference between the /ɛ̃/ and /œ̃/ groups (no merger), and red represents no statistically significant difference (the sounds are merged). Most (9/15) participants completely merge the vowels, and indicate no specifics regarding the merger (apart from being an additional source of evidence for the merger's existence). No participants displayed fully unmerged vowels. 4 participants could perceptually differentiate the vowels to some statistically significant degree, but not produce them in a significantly differentiated manner. These 4 participants provide evidence for the model's accuracy. 2 participants, however, could not perceptually differentiate the vowels to any statistically significant degree, but could produce them in a significantly differentiated manner. These participants therefore provide data that cannot be described by the model, meaning that it is inadequate in describing the merger of the vowels.

6 - Discussion

The specific failure of the model in predicting results is unusual. This is because the model did not predict anything that did not take place, but rather failed to predict something which also occurred. Making a model which predicts in some speakers' perception only merger would also be wrong. In fact more of the evidence would go against such a model than does against my model. A functional model would have to find a way to predict simultaneous yet independent factors triggering in some speakers production only merger and in some speakers perception only merger. Which (as of yet overlooked) factors could cause participant 11's results, without preventing the model from predicting the results of the other participants?

6.1 Possible errors in the model & method

A number of factors relating to the theoretical phonetics and phonology underlying the details of the model, as well as possible issues in data analysis, appear to offer potential explanations for the data. However, as will be shown below, short of suggesting that participants 6 and 11 do in fact merge the sounds in production, none of them can adequately be used as a basis on which to correct the model. One potential factor relates to the understanding of the process of rounding loss in low vowels. Stevens (1998) shows that the distinctive acoustics features of rounding are less pronounced in low vowels. In creating the model, I assumed this was due to rounding loss in articulation (as shown by Linker 1982), but is it possible that it is independent? While this could potentially be the case, it wouldn't actually explain the strange data, since the production test focused on acoustics and not articulation. I specifically found that the sounds are acoustically distinct for some people who also cannot distinguish them when listening. It could be that the problem is not in the conflation of articulation and acoustics but in the conflation of a perception effect with a production effect. Terbeek (1977) suggests that the perception of rounding is lower in low vowels. I, again, assumed this was due to the lessening of articulatory and acoustic differences, but could it be its own factor which makes people have a harder time distinguishing rounding in low sounds even when acoustically distinctive? Such an idea seems unlikely, simply because there is no evidence that I am aware of showing that perception is modulated distinctly from acoustic input, such that one acoustic feature could cause another independently present acoustic feature to be perceived differently than it otherwise would have been. Good evidence is required in order to believe that an F2 of value ‘x’ will be heard either as ‘x’ or ‘y’ depending on the F1's value.

There are also potential issues in data collection and analysis, specifically in terms of understanding whether or not the results actually indicate a lack of merger although a statistical test can differentiate the two data groups when these are explicitly compared, would it be able to differentiate them on their own? Testing this would require many tokens of each word rather than the one per word from this study. This is something to think about in future studies. For now, it is possible that humans cannot actually differentiate these sounds based on the slight differences of the two non-merging speakers. However, this does not really resolve the issue. Unless we suppose that the statistical difference between the sounds in production for participant 11 is a mere coincidence not reflective of any real difference in production, these speakers would still be producing them with a slight difference, which they must have learned somewhere. However they somehow learned these sounds as different but are not able to tell them apart. This causes an inherent issue, which cannot be resolved within a system like BiPhon. BiPhon works by first taking in the acoustic signals produced by others and interpreting them in a meaningful manner, and then attempting to recreate those signals in production.

In order to explain the results of participant 11, BiPhon would have to first take in the signals from /ɛ̃/ and /œ̃/ (as produced by others), and analyze them such that they are treated as a single phoneme in the surface and underlying forms. This single phoneme would then have to undergo transformations such that 2 distinct phones are being produced by the articulatory layer. This is only possible if there is a particular environment which can systematically predict which vowel quality will be produced, which would mean that the sounds are allophones of the same phoneme and not distinct phonemes. However, 16/20 stimuli contain their target vowel in a syllable which is minimally paired with that of a stimulus in the other group. 4 of the stimuli as a whole, 'brin' - 'brun' and 'Ain' 'un', are in true minimal pairs. Allophony could not predict a significant difference in F2 value between the particular groups given, since there is no systematic differentiating environment. Without such an environment, the only way to derive two distinct phones is to start with distinct phonemes; the only way for there to be two distinct phonemes is for the phones to be analyzed as distinct phonemes when perceived. For this reason, BiPhon alone cannot describe these particular results, unless we are dealing with a false significance result.

This issue extends beyond BiPhon. Instead, this is the nature of the interaction between linguistic production and perception. This is because, in L1 learning, production must attempt to imitate the data received through perception. In order to learn how to articulate ‘brun’ (/bʁœ̃/) and ‘brin’ (/bʁɛ̃/) as two distinct words, a native speaker must first listen to these words, analyze them as containing different (vowel) phonemes, and then try to articulate them in such a way that other speakers (who themselves already know both words) recognize them as two distinct words. If L1 learners do not perceive any phonemic difference between two words uttered by an adult L1 speaker, they will not have any basis on which to attempt to articulate them with a phonemically meaningful difference. For this reason, it cannot be the details of BiPhon's bidirectionality which have led to P11's data, since the nature of first language acquisition is such that perception must be the first step in phoneme categorization.

6.2 A potential explanation

One factor which could help explain the results, and which BiPhon is not equipped to account for (therefore accounting BiPhon's inability to describe the data), is the interference of a prescribed standard language. Such interference in French is supported by Morin & Kaye (1982). They made French participants read aloud sentences, one of which contained a plural noun. In colloquial speech, the plural marker would not be pronounced and therefore the noun's number would be left ambiguous. In standard speech, however, it would be incorrect to have a plural marker there at all. Participants consistently pronounced the marker, creating speech which was neither authentic colloquial French nor correct standard French. They use this to show how great the instinct to make yourself sound more 'proper' in (at least) experimental reading conditions is, and that researchers shouldn't expect purely naturalistic speech if there is reading involved.

It is therefore possible, then, that these two participants instinctively adopted a more formal/standard pronunciation when confronted with having to read words out loud in front of a microphone. This could explain why their results showed somewhat, although not very, significant differentiation between the vowels. It could also explain why they could not distinguish the vowels in perception. If these participants actually merge the sounds in their daily speech production, then there is nothing unexpected about their inability to tell the sounds apart. It could be that they are aware (through the public education system in France) enough of the differences to lightly mimic them in particular settings, but have not in their lives heard the sounds as produced distinctly enough for them to have learned how to systematically differentiate them. All of this is supported by one participant's comments to me after the study was done: P6 noted feeling as though an antiquated accentuation (pronunciation) that is normally not authentic to his speech had resurfaced while reading some words. P6's production data was such that the difference between the vowels was not very significant, but more significant than other participants (P11 is the exception).

Of course, this solution remains unproven. Morin & Kaye show that it is possible, P6's comments suggest that this is the phenomenon at play, and it offers an explanation as to why BiPhon failed to predict this, since the model is not designed to handle interference from distinct language varieties stored within single speakers. For these reasons, although this solution is the one most apt at describing experimental data given current knowledge, more research would be required in order to vindicate its application. For example, a study could be conducted in which a participant is made to have to figure out words (stimuli) from clues given by researchers. This would remove reading as a factor of interference and make participants focus on word guessing rather than pronunciation. If the solution proposed above is correct, production data could then be more naturalistic. If it is not, doing this will not prevent these unanticipated results. Doing a study such as this can therefore reveal whether standard language interference is truly at play or not.

6.3 General analysis of BiPhon

BiPhon in its bare form allows for two patterns of diachronic sound merger: production triggered and perception triggered merger. Production triggered merger causes one generation to merge sounds in production only, and only following generations fully merge sounds. Perception triggered merger, however, causes full merger of the sounds from one generation to the next. When feeding to the model the phonetic and phonological factors for the merger of /ɛ̃/ and /œ̃/, BiPhon-OT predicts a 2 stepped production triggered merger. 4/15 participants originated data characteristic of this type of merger, and which could not be explained if the model had predicted a single-stepped perception triggered merger. The model had partial success in predicting the results.

The biggest failure of the model in section 3 has now been well discussed. Although I have offered a possible reason for why this has occurred, other factors may be at play. What is apparent, however, is that a BiPhon model is not capable of describing a merger in perception only, since its bi-directional learning nature entails that production is derived from the results of a perceptual analysis of external input. If this is in fact the result of standard language interference, then a limitation of BiPhon is highlighted here. BiPhon models a singular language processing network, and is only capable of describing one particular variety of a single language. Modeling interference of any kind requires a model capable of describing several simultaneous language systems within an individual and how these interact and influence one-another. Beyond this, in the case of the particular form of interference which I am suggesting is at play, one needs to be able to describe the external contexts which trigger interference, which BiPhon in its current state is not apt to do. Even if my explanation of the results is incorrect, it is still clear that some process is at play which the model cannot describe. If this is the case, discovering what exactly did happen can be insightful as to the current limitations of BiPhon and BiPhon-OT. What seems certain is that modeling a singular perception to production pathway does not allow for more accurate predictions.

It must be said that, even if both my predictions had been confirmed experimentally, and the model developed in this paper had therefore been entirely successful, the original BiPhon format has been altered in order to create it. An extra layer had to be added in order to slow down the predicted rate of change. Without it, change would become faster than what I can justify it being. With it, a range of different rates of sound change are possible, which makes the model as a whole more realistic. Of course, without it the merger is still predicted and the predictions under investigation in this paper's study would be the same with the same outcomes. Therefore, to an extent it could be argued that the model functions to the same extent without the need for my alteration. However, the inability of BiPhon to predict slower sound change without an extra layer (with this case study at least) can be seen as a limitation of the model. On the other hand, the flexibility of the model, such that I could add this layer and increase accuracy without damaging any other aspect of its functionality, should be seen as a positive.

7 Conclusion

The purpose of this research was to find out whether it was possible to create a functional and accurate BiPhon-OT model which uses generational cue reinterpretation in order to predict the merger of phonemes /ɛ̃/ and /œ̃/ in north metropolitan French. It was possible to make this model, although a slight alteration from the basic design was made in order to achieve this. The model is able to predict that the phonemes, depending on cue interpretation, can either raise to a more central height (as in Quebecois French (Carignan, 2011)) or lower (as observed in Belgian French (Delvaux Metens & Soquet 2002)). It also predicts that lowered vowels can lose their rounding distinction (as in NMF (Malécot & Lindsay 1976)), although this can be progressive (accounting for BF (Delvaux Metens & Soquet 2002)). Experimentation showed that the model also predicted accurately that vowel merger was triggered by speakers who perceived two phonemes but articulated them identically. The model was not perfectly accurate, being unable to predict a participant who could not distinguish the two vowels perceptually but articulated them distinctly. This highlights BiPhon's inability to accommodate several linguistic codes and their interaction within a single person. Overall, BiPhon-OT was mostly successful in allowing for a functional and accurate model of the diachronic evolution of /ɛ̃/ and /œ̃/, but showed itself to be limited in its ability to account for certain variables.

References

Académie Normande. Le système phonologique du Français. Inspection académique de la Manche - Circonscription de Mortain.

Anderson, R. H. (1968). The French sound system: its phonology and orthography. Thesis: Masters of Arts, Fresno State college.

Beddor, P., Krakow, R., & Goldstein, L. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology, 3*, 197-217. doi:10.1017/S0952675700000646

Boersma, P., Benders, T., & Seinhorst, K. (2020): Neural networks for phonology and phonetics. *Journal of Language Modelling* 8: 103-177.

Boersma, P., Chládková, K., & Benders, T. (2022).Phonological features emerge substance-freely from the phonetics and the morphology. *Canadian Journal of Linguistics, 67*, 611–669.

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology, 25*, 217–270.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.10, retrieved 3 May 2023 from http://www.praat.org/

Boersma, P. (2006). Prototypicality judgments as inverted perception. Fanselow, G., Féry, C., Schlesewsky, M., & Vogel, R. (eds.): *Gradience in Grammar*, 167–184. Oxford: Oxford University Press.

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

Boersma, P. (2009). Cue constraints and their interactions in phonological perception and production. Boersma, P., & Hamann, S. (eds.): *Phonology in perception*, 55–110. Berlin: Mouton de Gruyter.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. Benz, A. & Mattausch, J. (eds.): *Bidirectional Optimality Theory*, 33–72. Amsterdam: John Benjamin.

Boersma, P. (2019). Simulated distributional learning in deep Boltzmann machines leads to the emergence of discrete categories. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, 5–9 August 2019. 1520–1524.

Carignan, C. (2011). Oral articulation of nasal vowels in French. University of Illinois at Urbana-Champaign, USA.

Delvaux, V., Metens, T., & Soquet, A. (2002). Propriétés acoustiques et articulatoires des voyelles nasales du français. *Thierry Metens (2014)*.

De Mareüil, P. B., Adda-Decker, M., & Woehrling, C. (2007). Analysis of oral and nasal vowel realisation in northern and southern French varieties. LIMSI-CNRS, Orsay, France.

Fant, G. (1960). Acoustic theory of speech production. *The Hague: Mouton*.

Fujimura, O. & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. *JASA 49*, 541–558.

Gottfried, T. L. (1983). Effect of consonant context on the perception of French vowels. *Journal of phonetics (1984), 12*, 91-114. State University of New York at Purchase, Purchase, New York 10577, U.S.A.

Hamann, S. (2009). The learner of a perception grammar as a source of sound change. Boersma, P., & Hamann, S. (eds.): *Phonology in perception*, 55–110. Berlin: Mouton de Gruyter.

Hayes, B., Kirchner, R., Steriade, D (2004). The Typology of Rounding Harmony in Phonetically Based Phonology. Kaun, A. R. (ed.): Phonetically Based Phonology, Chapter 4.

Hirsch, F., Sock, R., Connan, P-Y., Brock, G (2003). Auditory effects of anticipatory rounding in relation with vowel height in French. Institut de Phonétique de Strasbourg.

Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. PhD dissertation, UCLA. *UCLA Working Papers in Phonetics 56*.

Malécot, A., & Lindsay, P. (1976). The Neutralization of /ɛ̃/ - /œ̃/ in French. *Phonetica, 33*, 45–61.

Morin, Y., & Kaye, D. K. (1982). The Syntactic Bases for French Liaison. *Journal of Linguistics, 18(2)*, 291-330. Retrieved June 22, 2021, from http://www.jstor.org/stable/4175640

Nicholas, J. A., Fagyal, Z., & Carignan, C. (2019). What's with your nasals? Perception of nasal vowel contrasts in two dialects of French.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Stevens, K. N. (1998). Acoustic phonetics. *MIT Press*.

Terbeek, Dale. (1977). A cross-language multidimensional scaling study of vowel perception. PhD dissertation, UCLA. *UCLA Working Papers in Phonetics 37*.

van Leussen, J-W. (2020). The emergence of French phonology. Thesis: Amsterdam Center for Language and Communication, University of Amsterdam.

Wright, J. (1980). The behavior of nasalized vowels in the perceptual vowel space. *Report of the Phonology Laboratory, University of California, Berkeley 5*, 127–163.

Appendices

Appendix A

Production stimuli list

/ɛ̃/

- brin
- ain
- main
- simple
- linge
- faim
- quinze
- cinglé
- mannequin
- bain

/œ̃/

- brun
- un
- commun
- humble
- lundi
- parfum
- aucun
- jungle
- chacun
- bungalow

filler

- terre
- déjeuner
- bouteille
- lunettes
- voiture
- maison
- samedi
- coude
- nez
- huit
- brique
- écran
- tour
- bâtiment
- nuage
- pluie
- soleil
- peur
- sommeil

Appendix B

Perception stimuli list

Target pairs

- in - un
  - /ɛ̃/ - /œ̃/
- tin - tun
  - /tɛ̃/ - /tœ̃/
- sin - sun
  - /sɛ̃/ - /sœ̃/
- nin - nun
  - /nɛ̃/ - /nœ̃/
- lin - lun
  - /lɛ̃/ - /lœ̃/
- yin - yun
  - /jɛ̃/ - /jœ̃/
- inte - unte
  - /ɛ̃t/ - /œ̃t/
- ince - unce
  - /ɛ̃s/ - /œ̃s/
- inle - unle
  - /ɛ̃l/ - /œ̃l/
- inque - unque
  - /ɛ̃k/ - /œ̃k/

Filler pairs (type 1)

- o - ou
  - /o/ - /u/
- to - tou
  - /to/ - /tu/
- so - sou
  - /so/ - /su/
- no - nou
  - /no/ - /nu/
- aute - oute
  - /ot/ - /ut/
- ausse - ousse
  - /os/ - /us/

Filler pairs (type 2)

- sa - cha
  - /sa/ - /ʃa/
- sé - ché
  - /se/ - /ʃe/
- sof - chof
  - /sɔf/ - /ʃɔf/
- si - chi
  - /si/ - /ʃi/
- sur - chur

- /syr/ - /ʃyr/
- son - chon
  - /sõ/ - /ʃõ/

Filler pairs (type 3)

- ga - ka
  - /ga/ - /ka/
- gou - kou
  - /gu/ - /ku/
- gué - ké
  - /ge/ - /ke/
- gu - ku
  - /gy/ - /ky/
- gan - kan
  - /gɑ̃/ - /kɑ̃/
- git - kit
  - /git/ - /kit/