Temporal vs. area-sum measurements of vowel nasality

Annual Meeting of the Linguistic Society of America

Michael Dow
Université de Montréal

January 9, 2016
Introduction
Objectives

2. Apply the DER on a (personally collected) nasometric corpus of French.
3. Offer phonetic explanation behind some effects.
Example: [i] in [in] sequence (French)

Nasal & oral energy

Differential energy

Temporal vs. area-sum measurements
Background questions

1. Q: What kind of data?
   A: Mainly “split-level” (separate but simultaneous measures of orality & nasality)

2. Q: What’s used now?
   A: Temporal formulae (proportion of nasal phase duration), using NAS as one example

3. Q: How does the DER compare in practice?
   A: Nasometric study of coarticulation in French: gives more nuanced scores, especially for vowels with rapid energy changes.

4. Q: Why?
   A: DER builds numbers directly into calculations.
Main points

1. DER & NAS correlate but disagree in certain key cases
2. DER inherently more precise than NAS: oral & nasal energies ($E_o, E_n$) not entirely interdependent – why?
   - **High vowels**: $E_n$ can rise either slowly or most rapidly of all heights (*underestimated* by NAS)
   - **Non-high vowels**: $E_o$ on average greater at start → sharp fall; $E_n$ can either rise at similar rates or barely rise (*overestimated* by NAS)
3. The DER is more appropriate at quantifying vowel nasality than the NAS (at least concerning coarticulation).

**Today**: Focus on high vowels, esp. where NAS < DER.
Outline

1. Introduction
2. Quantifying nasality
3. Methodology
4. Results
5. Discussion
6. Conclusion
7. Appendix

Dow

Temporal vs. area-sum measurements
Quantifying nasality
Phonetic correlates of nasality

Nasal vowels among most complex sounds of human language, several measurable phonetic correlates

- Articulatory: activation of *levator palatini* (e.g., Lubker 1968), lowering of velum (e.g., Henderson 1984)
- Aerodynamic: air pressure & area of velopharyngeal port opening (VPO; e.g., Warren et al. 1993)
- Acoustic: interaction between nasal cavity’s pole-zero pairs and oral formants (Maeda 1993), weakening of F1 (e.g., Delattre 1954), etc. (cf. Baken & Orlikoff (2000) for review)
Instruments & data

- Articulatory (imaging, mechanical, EMA, electromyography): size of VPO, velic height or muscle activation over time, positioning of oral articulators

- Acoustic (non-instrumental): formant tracking, amplitude differences (à la Chen 1997) or p0 prominence (Styler & Scarborough 2014)

- **Split-level**: separate but simultaneous oral & nasal channels (aerodynamic or instrumental acoustic)

  cf. Krakow & Huffman 1993, Delvaux 2012 for exhaustive surveys
Typical results

With example studies on French (so “e.g.,” all around).

- **Acoustic**: average or point-by-point difference in dB (cross-categorical), measurement of distance between oral & nasal “formants” (Chen 1997)
- **Formant tracking & split-level**: global score ("V = x% nasal")
  - Formant tracking: onset of nasal band (Spears 2006)
  - Aerodynamic: onset of (sufficient) nasal airflow (Delvaux et al. 2008)
  - Nasometric: onset of (sufficient) nasal energy (Montagu 2007)
Caveats

- Several discrepancies on studies of coarticulation in French (% nasality vs. height)
- Different methods = different correlates = different stages of pronunciation:
  activation → movement → aerodynamics → acoustics
- Simultaneous multiple instruments impossible for most combinations, no way of directly comparing results (yet)

So let’s use one data source – Dow (2014)
Methodology
Background & participants

- Nasometric (split-level acoustic) study of French (France)
- Objective: document nasal coarticulation patterns of French wrt vowel quality and duration
- Glottal Enterprises NAS-1 SEP Clinic hand-held nasometer: equally distant microphones (mouth, nose) separated by sound-attenuating plate
- 20 native speakers from Brittany, Picardy: 6 women, 14 men; average age = 57.4 (sd = 13.4); no significant differences between groups’ French data
Stimuli

- Noun + adjective combinations of **vowel targets** and **environments** (e.g., *la partis/an#/s/arcastique*)
  - $V =$ oral vowels in oral contexts, /a, e, ø, o, i, y, u/
  - $VN =$ oral vowels before noun-final nasal consonants
  - $V^n =$ contrastive nasal vowels, /a$^n$, e$^n$, ø$^n$, o$^n$/

- Each list read 3 times by each speaker (self-directed pace)
- Recorded in Praat in stereo (separate channel for oral, nasal)
- Total of **3,240 vowels**
Measurements & treatment

- 10 equally-spaced measurements of vowel’s energy in each channel (oral, nasal)
- Data re-centered around $sd$ of each channel’s readings within a speaker & repetition
- 2 measurements: nasalance-based (NAS) & Differential Energy Ratio (DER)
- Shared points:
  - $p$ = arbitrary threshold (both measurements); here, where nasal energy overtakes oral energy.
  - $C$ = end of vowel (orality $\rightarrow 0$).
- Vowel devoicing (occasional on word-final high vowels) caused some erroneous readings, but not enough to impact average 0% nasality on oral vowels
NAS calculations

Representative example of temporal measurements, along the lines of e.g., Rochet & Rochet 1991.

1. **Nasalance** (nasal energy over total energy) at each point, expressed as percentage

2. Nasal phase defined wrt arbitrary threshold: all points whose nasalance $\geq 50\%$ (i.e., $E_n \geq E_o$)

3. $\text{NAS} = \#$ of points in nasal phase vs. total $\#$ of points

Simply put (specific to regressive nasalization): $\frac{C-p}{C}$
Hypothetical examples, NAS = 50%

Temporal vs. area-sum measurements
DER calculations

Generalizing to where $E_o = f(x)$ and $E_n = g(x)$...

1. **Differential energy curve ($\Delta$):** $f(x) - g(x)$

2. Phases separated around $p$, where $\Delta = 0$ (oral = positive values; nasal = negative)

3. Area-sum of each phase calculated

   $$A_o = \sum_{x=0}^{p} [f(x) - g(x)]$$

   $$A_n = \sum_{p}^{C} [|f(x) - g(x)|]$$

4. $$\text{DER} = \frac{A_n}{(A_n + A_o)}$$
DER crucial points, hypothetical $\Delta$
Results
Results by target, VN & V\textsuperscript{n}

Expressed in % nasal. V \approx 0\% ([V] aside)

<table>
<thead>
<tr>
<th>Target</th>
<th>Context</th>
<th>NAS</th>
<th>DER</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>VN</td>
<td>21.3</td>
<td>20.5</td>
<td>-0.8</td>
</tr>
<tr>
<td></td>
<td>V\textsuperscript{n}</td>
<td>86.2</td>
<td>89.0</td>
<td>2.8</td>
</tr>
<tr>
<td>/e/</td>
<td>VN</td>
<td>28.1</td>
<td>28.1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>V\textsuperscript{n}</td>
<td>86.2</td>
<td>89.4</td>
<td>3.2</td>
</tr>
<tr>
<td>/ø/</td>
<td>VN</td>
<td>22.3</td>
<td>20.5</td>
<td>-1.9</td>
</tr>
<tr>
<td></td>
<td>V\textsuperscript{n}</td>
<td>66.2</td>
<td>66.8</td>
<td>0.6</td>
</tr>
<tr>
<td>/o/</td>
<td>VN</td>
<td>20.4</td>
<td>16.1</td>
<td>-4.4</td>
</tr>
<tr>
<td></td>
<td>V\textsuperscript{n}</td>
<td>97.2</td>
<td>97.7</td>
<td>0.5</td>
</tr>
<tr>
<td>/i/</td>
<td>VN</td>
<td>61.2</td>
<td>69.6</td>
<td>8.3</td>
</tr>
<tr>
<td>/y/</td>
<td>VN</td>
<td>51.1</td>
<td>57.1</td>
<td>5.9</td>
</tr>
<tr>
<td>/u/</td>
<td>VN</td>
<td>34.8</td>
<td>39.2</td>
<td>4.4</td>
</tr>
</tbody>
</table>
Boxplot: difference by target (VN context)

**NB:** With a few exceptions, general trends hold for all speakers.
All heights (VN context) correlate strongly (between $r = 0.86$ and 0.93).

Paired t-test (VN context) shows difference extremely significant for mid (negative direction) & high (positive) vowels

<table>
<thead>
<tr>
<th></th>
<th>low</th>
<th>mid</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.8</td>
<td>-2.09</td>
<td>6.67</td>
</tr>
<tr>
<td>t</td>
<td>-0.93</td>
<td>-4.19</td>
<td>13.8</td>
</tr>
<tr>
<td>p</td>
<td>0.36</td>
<td>$&lt;0.001^{***}$</td>
<td>$&lt;0.001^{***}$</td>
</tr>
</tbody>
</table>

The two model & vary with nasality, but disagree somewhere.
Bland-Altman plot

AKA Tukey mean-difference plot
- $x$-axis: mean % nasal (NAS + DER)/2
- $y$-axis: difference (DER – NAS)
- Used to test agreement between two measurements of the same phenomena
High vowels (VN context)
Mid vowels (VN context)

Temporal vs. area-sum measurements
So?

- Simple averages mask the disagreement at rates of nasality where the difference is more evenly distributed.
- Disagreement skewed towards NAS > DER at lower levels, NAS < DER at higher levels.
- More pronounced (greater range of difference) on high vowels, esp. at higher end.
- Degree of precision aside, how can we decide which is more appropriate?
High vowels & nasalization

- **Aerodynamic**: High intraoral pressure on high vowels → greater nasal sound pressure levels on high vowels (Clark & Mackiewicz-Krassowska 1977)
- **Acoustic**: For VPO given, nasal coupling more likely to occur on high vowels (House & Stevens 1956) because of nasal pole-zero interactions (Maeda 1993)
- **Perception**: High vowels perceived as most nasal with least VPO size (e.g., Maeda 1982)
What about articulation?

Spoilers: Unclear.

- Velic height and vowel height long thought proportionate: among oral vowels, velum lowest on low vowels, etc. (e.g., Bell-Berti 1976), but called into question more recent studies (e.g., Rossato et al. 2003)

- Originally thought to signify low vowels easier to nasalize (esp. physiological motivation for development of French nasal vowels, e.g., Straka 1955)

- However, low vowels in experiments...
  - In nasal contexts: produced with much larger VPO than high vowels (Chen & Wang 1975)
  - In oral contexts: occasionally open velum & trace amounts of nasality (e.g., Ohala 1975, Clumeck 1976)
  - ...but even these may not be universal (cf. Hajek & Maeda 2000)
Are high vowels harder or easier to nasalize, from articulatory point of view? Probably irrelevant.

- If VPO size solely responsible: perhaps harder
- However, inherent velic height inconclusive
- More importantly, both aerodynamic and acoustic factors take precedence & likely make this question irrelevant
Back to the data

- Nasal energy rises twice as fast on high vowels (vs. on any other vowel type)
- Higher nasality $\neq$ greater difference (cf. contrastive nasal vowels; also low vowel VN in Picard: 87% (NAS) vs. 91% (DER))
- $p$ being equal, a sharper change in one type of energy $\rightarrow$ greater difference between NAS & DER
Representative examples: 2 productions of [y] in _N#s by same speaker

Same crossover point but significantly different rates of nasal energy change

- NAS (V1, V2) = 24.1%
- DER (V1) = 39.3%
- DER (V2) = 12.2%

The DER reflects the difference between these 2 vowels, while the NAS conflates it, towards less nasal (V1) and more nasal (V2).
Temporal vs. area-sum measurements
Conclusion
Summary

- DER more reflective of nasality because of... 
  - Its ability to differentiate rates of change 
  - Its direct incorporation of energy readings 
  - Its inherent nuance (not solely a function of crossover point)

- The difference between the NAS and DER is crucial on vowels where energy changes rapidly

- In French, this applies most strongly to high vowels’ nasal energy – typically an underestimation by NAS, but not exclusively

- Possible explanation: Even if high vowels are harder to nasalize from an articulatory point of view, they are easier from an aerodynamic, acoustic and perceptual point of view
(A few) remaining questions

- Are these effects language-specific?
- Can this level of nuance be perceived? Can languages encode it in the grammar?
- What about mid vowels? Or more generally those cases where the NAS overestimates % nasal?
- How are we to account for speaker variation?
- Can a similar formula be applied to aerodynamic and/or non-instrumental acoustic data?
Potential implications

- In experimental phonological descriptions, vowel nasality may either be underreported or overreported.
- The DER may be a more accurate gauge of vowel nasality in clinical applications.
Acknowledgements

Funding: NSF Doctoral Dissertation Research Grant 1360758; internal SSHRC grant (Université de Montréal).

Many thanks to: Karthik Durvasula, Ken de Jong, Veronique Delvaux and Angélique Amelot for their thoughts on this topic, as well as to my participants. Also thanks to Anne-Michelle Tessier, Michael Becker, Chris Green, Eric Beuerlein, and any others who helped with earlier versions of this presentation or its abstract.


Appendix
(A few) remaining questions, I

- **Q:** Are these effects language-specific?  
  **A:** We can start by looking at Picard. Then...

- **Q:** Can this level of nuance be perceived? Can languages encode it in the grammar?  
  **A:** Quite likely (cf. Maeda 1982), but link with DER needs to be established. Might be encodable in a more unidimensional interface/module.

- **Q:** What about mid vowels? Or those cases where the NAS overestimates % nasal?  
  **A:** In all contexts (even V), $E_o$ starts out much higher on non-high than on high vowels $\rightarrow$ sharper fall towards N. If $E_n$ fails to rise at similar rates, this means $p \approx 70\%$ but negligible $A_n$.  

Temporal vs. area-sum measurements 37/40
Q: How are we to account for speaker variation?
A:

- Anecdotes from the data: speakers seem to fall into groups according to rates of nasal energy rise, high vowel VN (small, middling and large rise)
- Consistency within speakers may point to different status of phonologization of high vowel nasalization (needs to be teased apart with duration, as well)
- More random variation may point to sloppy articulation (no contrastive high nasal vowels in French, ease of nasalization)

Q: Can a similar formula be applied to aerodynamic and/or non-instrumental acoustic data?
A: Quite likely again. Although difference between aerodynamic instruments and acoustic need to be worked out.
NAS: 30%, DER: 7.8%
Low vowel, Bland-Altman plot