1. Technical Metric: Speech Analysis Accuracy (SAA)

1. **Metric name**: Speech Analysis Accuracy (SAA)
2. **Brief description**: SAA measures how accurately the system identifies and analyzes key speech elements (pace, tone, clarity, filler words) compared to expert human evaluations. This is essential because the core value proposition of VerbalVector is providing accurate feedback on communication patterns—if the system misidentifies these patterns, its feedback would be ineffective.
3. **How it is calculated**: SAA = (Number of correctly identified speech elements / Total number of speech elements evaluated) × 100% Implementation involves creating a validation dataset of speech samples pre-analyzed by communication experts, running these samples through your distilled model, and comparing the system's analysis with expert analysis.
4. **Monitoring strategy**: SAA should be automatically measured through regular validation testing with benchmark speech samples. The data science team should own this metric, reviewing it weekly during development and monthly after deployment, with results logged in a performance dashboard.
5. **Metric conflicts**: Optimizing for SAA could conflict with latency requirements, as higher accuracy might require more complex processing, increasing response time. This creates a trade-off between accuracy and user experience. Additionally, pursuing high accuracy on your validation dataset might lead to overfitting, reducing effectiveness with diverse speakers not represented in your training data.
6. **Metric lifecycle**: Initially, SAA should be measured broadly across all speech elements. As your model matures, it should evolve into more granular sub-metrics for specific elements (pace accuracy, tone accuracy, etc.) and eventually incorporate user-perceived accuracy measurements.

2. Business Metric (end-use): Communication Mastery Velocity (CMV)

1. **Metric name**: Communication Mastery Velocity (CMV)
2. **Brief description**: CMV measures how quickly users improve their communication skills across different contexts while using VerbalVector. This metric captures both depth (improvement rate) and breadth (context expansion) of value creation, directly correlating with user retention and potential monetization. It indicates whether VerbalVector creates accelerating value over time, essential for product stickiness.
3. **How it is calculated**: CMV = (Skill Improvement Rate) × (Context Expansion Factor) Where:

    - "Skill Improvement Rate" = % reduction in targeted communication issues per week
    - "Context Expansion Factor" = (Current # of communication contexts used) / (Initial # of contexts)
4. **Monitoring strategy**: CMV should be automatically calculated through the system's analysis of user data over time. The product and business development teams should jointly own this metric, reviewing it monthly. The system should generate visualizations showing improvement trajectories across different skills and contexts.
5. **Metric conflicts**: Optimizing for CMV might incentivize focusing on easy-to-improve issues rather than more complex but valuable communication challenges. This creates tension between showing quick wins and developing deeper skills. Additionally, pursuing high CMV might lead to feature bloat as the team adds more contexts to inflate the expansion factor, potentially sacrificing depth for breadth.
6. **Metric lifecycle**: In the MVP phase, CMV should focus on clearly measurable skills and common contexts. As the product matures, it should incorporate more nuanced skills and specialized contexts, eventually evolving into a "Communication ROI Velocity" that connects skill improvement directly to measurable outcomes in users' lives.

3. Ethical Metric: Demographic Feedback Parity (DFP)

1. **Metric name**: Demographic Feedback Parity (DFP)
2. **Brief description**: DFP measures whether the system provides equally accurate and helpful feedback across different demographic groups (gender, age, accent, dialect). This metric is crucial because speech analysis systems often perform differently for various groups, potentially reinforcing biases or providing less effective feedback to certain users.
3. **How it is calculated**: DFP = 1 - (Standard deviation of accuracy scores across demographic groups / Mean accuracy score) Where a score of 1 means perfect parity across groups and lower scores indicate greater disparities in system performance.
4. **Monitoring strategy**: DFP should be calculated monthly using a diverse validation dataset. A cross-functional team including data scientists and ethics specialists should own this metric, with regular testing using a demographically diverse benchmark dataset.
5. **Metric conflicts**: Optimizing for DFP could potentially reduce overall accuracy if the model is modified to perform more consistently across groups at the expense of maximum performance for any single group. Additionally, addressing DFP might require collecting more demographic data, conflicting with privacy goals.

There's also a risk that over-optimizing for demographic parity could lead to applying different standards to different groups rather than evaluating everyone by the same communication criteria.

6. **Metric lifecycle**: Initially, DFP should focus on the most significant demographic variations in your target market. As your system matures, the metric should incorporate more demographic dimensions. The goal should be reaching a state where demographic factors have negligible impact on feedback quality, with periodic audits continuing throughout the system's lifecycle.