



ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

HỆ THỐNG CHUYỂN ĐỔI
TIẾNG NÓI TIẾNG VIỆT SANG NGÔN NGỮ KÝ HIỆU
DỰA TRÊN MÔ HÌNH NGÔN NGỮ LỚN
CHO NGƯỜI KHIẾM THÍNH



Thông tin chung



Họ và Tên: Hoàng Quy Quỳnh Chi

MSHV: 250201003



[Chi Hoang 's Research-Methods YouTube](#)



[Chi Hoang 's Research-Methods GitHub](#)



Tóm tắt

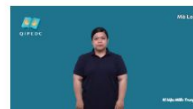


VSL thiếu dữ liệu căn chỉnh và
hệ thống dịch hiệu quả

Phương pháp hiện tại rời rạc,
thiếu nhịp điệu

Giới thiệu Vi-Sign:
không cần gloss, nhận biết ngữ điệu

Xây dựng bộ dữ liệu
Speech–Text–Sign (~6.000 câu)



Mã Lai (nước Ma-lai-xi-a)



số La Mã (Toán học)



Long Vương



đàng hoàng



phân vân



lau miệng



chó cắn



mèo cào



quan hệ từ



yêu nước Việt Nam



quan (chức vụ)



đi tuần



vỗ vè



hốt hoảng



chuyện nhỏ



chuyện lớn



cái gửi



hùng vĩ



bất khuất



nhưng

Giới thiệu



**Người khiếm thính
khó tiếp cận thông tin lời nói.**

**VSL chưa được triển khai rộng,
thiếu phiên dịch.**

**Thiếu dữ liệu căn chỉnh theo câu
và ngữ điệu cho VSL.**

**Cần hệ thống sinh ký hiệu liên tục,
tự nhiên, đồng bộ lời nói.**

Mục tiêu



Dữ liệu

Xây dựng
bộ dữ liệu ~6.000 câu,
căn chỉnh theo từ vựng

Hệ thống

Vi-Sign chuyển
speech/text → VSL,
không cần gloss



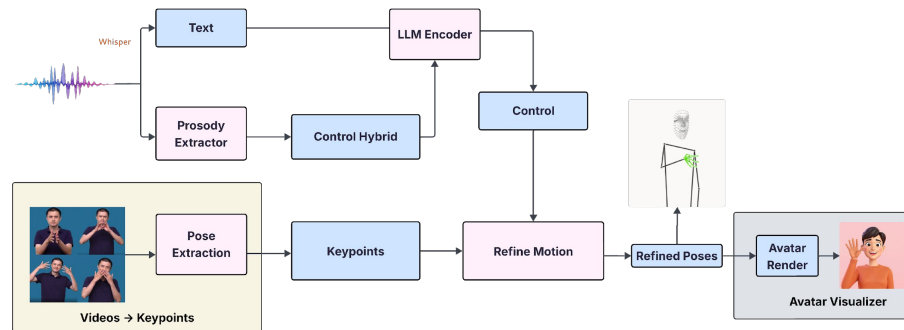
Đánh giá

Độ mượt,
độ ổn định,
đồng bộ ngữ điệu–cử chỉ





Nội dung và Phương pháp



Chuẩn hóa dữ liệu:
tiếng nói – văn bản – video

Trích xuất các điểm mốc
từ tay, cơ thể và khuôn mặt

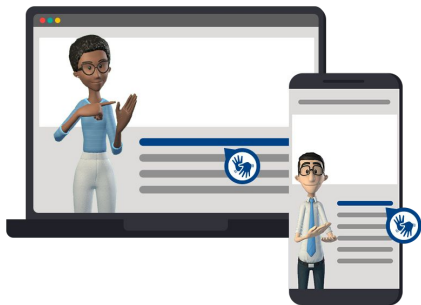
Trích xuất đặc trưng ngữ điệu
từ tín hiệu tiếng nói

LLM hỗ trợ suy luận ngữ cảnh
và dự đoán các tín hiệu điều khiển

Sinh - tinh chỉnh
độ mượt và ổn định của chuyển động

Trực quan hóa
bằng mô hình nhân vật ảo

Kết quả dự kiến



Chuyển động ký hiệu mượt
và ổn định hơn

Căn chỉnh tốt hơn
giữa ngữ điệu và cử chỉ

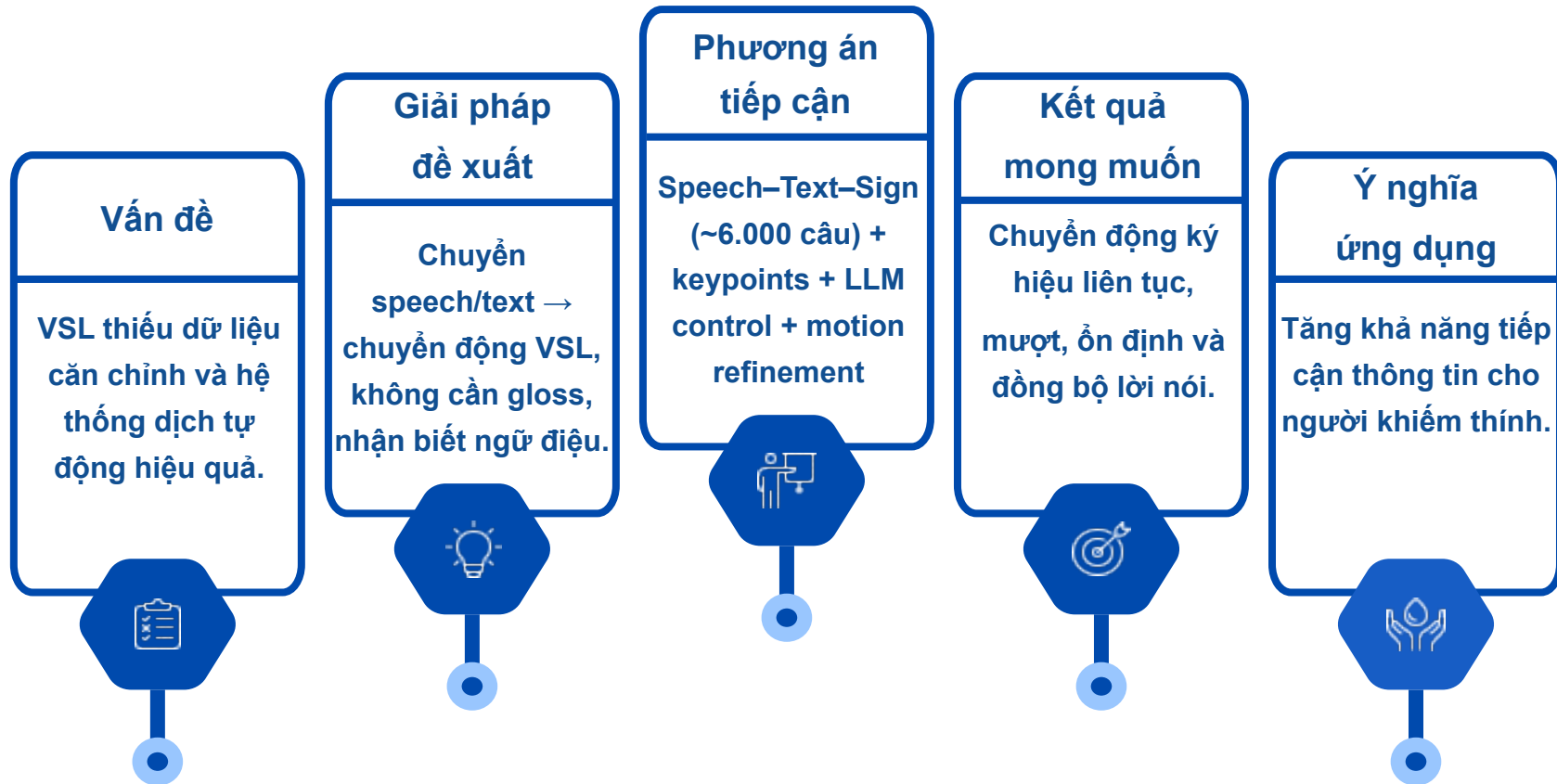
Cải thiện non-manual cues
(mặt, đầu, ánh nhìn)

Bộ dữ liệu và pipeline có thể
tái sử dụng





Tổng quan đề tài đề xuất





[1]. Danielle Bragg, Oscar Koller, Maryam Afsar, et al.:

Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. CHI 2019.

[2]. Dat Quoc Nguyen, Anh Tuan Nguyen:

PhoBERT: Pre-trained Language Models for Vietnamese. Findings of EMNLP 2020: 1037–1046.

[3]. Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, Richard Bowden:

Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation. CVPR 2020: 10023–10033.

[4]. Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, Richard Bowden:

Text2Sign: Towards Sign Language Production Using Neural Machine Translation and GANs. IJCV 129 (2021): 2757–2778.

[5]. MediaPipe Team: MediaPipe Holistic. URL: <https://developers.google.com/mediapipe> (accessed 2025-11-13).

[6]. Shaojie Bai, J. Zico Kolter, Vladlen Koltun:

An Empirical Evaluation of Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271 (2018).

[7]. QIPEDC: Vietnamese Sign Language Dictionary. URL: <https://qipcdc.moet.gov.vn/dictionary> (accessed 2025-11-13).